frontiers | Research Topics

Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery

Edited by
Mohammad Mofatteh, Mohamad Bydon and Mohammed Ali Alvi

Coordinated by
Mohammad S. Mashayekhi

Published in
Frontiers in Surgery
Frontiers in Human Neuroscience

ISSN 1664-8714
ISBN 978-2-8325-4645-1
DOI 10.3389/978-2-8325-4645-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery

Topic editors
Mohammad Mofatteh — Queen's University Belfast, United Kingdom
Mohamad Bydon — Mayo Clinic, United States
Mohammed Ali Alvi — University Health Network (UHN), Canada

Topic coordinator
Mohammad S. Mashayekhi — University of Ottawa, Canada

Citation
Mofatteh, M., Bydon, M., Alvi, M. A., Mashayekhi, M. S., eds. (2024). *Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4645-1

Table of contents

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 04 April 2023
DOI 10.3389/fnhum.2023.1154038

OPEN ACCESS

EDITED BY
Andrea Pigorini,
University of Milan, Italy

REVIEWED BY
Simone Russo,
University of Milan, Italy
Vassiliy Tsytsarev,
University of Maryland, United States

*CORRESPONDENCE
Andrei Barborica
✉ andrei.barborica@fizica.unibuc.ro
Christian G. Bénar
✉ christian.benar@univ-amu.fr

SPECIALTY SECTION
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

RECEIVED 30 January 2023
ACCEPTED 06 March 2023
PUBLISHED 04 April 2023

CITATION
Barborica A, Mindruta I, López-Madrona VJ, Alario F-X, Trébuchon A, Donos C, Oane I, Pistol C, Mihai F and Bénar CG (2023) Studying memory processes at different levels with simultaneous depth and surface EEG recordings. *Front. Hum. Neurosci.* 17:1154038. doi: 10.3389/fnhum.2023.1154038

Studying memory processes at different levels with simultaneous depth and surface EEG recordings

Andrei Barborica[1]*, Ioana Mindruta[2,3], Víctor J. López-Madrona[4], F-Xavier Alario[5], Agnès Trébuchon[6,7], Cristian Donos[1], Irina Oane[2], Constantin Pistol[1], Felicia Mihai[1] and Christian G. Bénar[4]*

[1]Department of Physics, University of Bucharest, Bucharest, Romania, [2]Epilepsy Monitoring Unit, Department of Neurology, Emergency University Hospital Bucharest, Bucharest, Romania, [3]Department of Neurology, Medical Faculty, Carol Davila University of Medicine and Pharmacy Bucharest, Bucharest, Romania, [4]Aix Marseille University, INSERM, INS, Institute of Neuroscience System, Marseille, France, [5]Aix Marseille University, CNRS, LPC, Marseille, France, [6]APHM, Timone Hospital, Epileptology and Cerebral Rhythmology, Marseille, France, [7]APHM, Timone Hospital, Functional and Stereotactic Neurosurgery, Marseille, France

Investigating cognitive brain functions using non-invasive electrophysiology can be challenging due to the particularities of the task-related EEG activity, the depth of the activated brain areas, and the extent of the networks involved. Stereoelectroencephalographic (SEEG) investigations in patients with drug-resistant epilepsy offer an extraordinary opportunity to validate information derived from non-invasive recordings at macro-scales. The SEEG approach can provide brain activity with high spatial specificity during tasks that target specific cognitive processes (e.g., memory). Full validation is possible only when performing simultaneous scalp SEEG recordings, which allows recording signals in the exact same brain state. This is the approach we have taken in 12 subjects performing a visual memory task that requires the recognition of previously viewed objects. The intracranial signals on 965 contact pairs have been compared to 391 simultaneously recorded scalp signals at a regional and whole-brain level, using multivariate pattern analysis. The results show that the task conditions are best captured by intracranial sensors, despite the limited spatial coverage of SEEG electrodes, compared to the whole-brain non-invasive recordings. Applying beamformer source reconstruction or independent component analysis does not result in an improvement of the multivariate task decoding performance using surface sensor data. By analyzing a joint scalp and SEEG dataset, we investigated whether the two types of signals carry complementary information that might improve the machine-learning classifier performance. This joint analysis revealed that the results are driven by the modality exhibiting best individual performance, namely SEEG.

KEYWORDS

EEG, stereo-EEG, simultaneous recordings, multivariate pattern analysis, recognition memory

1. Introduction

Electroencephalography (EEG) is routinely used to understand cognitive processes (Kappenman and Luck, 2011). The ability of these non-invasive recordings to capture cognitive processes accurately and entirely is the subject of ongoing investigations. A primary challenge is the well-known ill-posed problem of source reconstruction (Grech et al., 2008).

Knowing the actual sources and their time course in detail would provide invaluable information to disentangle brain activities. Clinical uses of EEG face a similar challenge, e.g., concerning the surface visibility of epileptiform activity, either ictal or inter-ictal. The challenge has been addressed through the simultaneous recording of intracranial and surface, both with EEG (Tao et al., 2005; Ray et al., 2007; Koessler et al., 2015; Antony et al., 2019; Barborica et al., 2021) and MEG (Pizzo et al., 2019). An asset of the clinical context is that many forms of epileptiform activity, sometimes paroxysmal, involve relatively large patches of cortical tissue that present synchronized activity, evoking potentials on the scalp having reasonable signal-to-noise ratio (SNR). By contrast, cognitive processes evoke more subtle activities and variations, involving deep brain structures, and high-frequency activity. These factors may cumulatively contribute to poor scalp visibility of the corresponding EEG activity.

Recognition memory (Yonelinas, 2002) provides an ideal test case to explore how neural activities evoked by cognitive tasks are captured at the scalp by EEG. Recognition memory is a complex cognitive function generally broken down into encoding, storage, and retrieval processes (Mandler, 1980; Besson et al., 2012). These are known to involve lateral and deep structures such as the hippocampus (Rutishauser et al., 2006; Merkow et al., 2015). Recognition memory has been extensively studied with EEG, using recordings made either on the scalp (Ratcliff et al., 2016) or in the brain (Merkow et al., 2015), but not simultaneously. Here, we assess to what extent the postulated processes are visible on scalp EEG by validating the source localization results with the simultaneous scalp intracranial recordings. The data are collected from patients undergoing stereoelectroencephalographic (SEEG) pre-surgical evaluation for drug-resistant epilepsy; they performed a standard task requiring them to encode and later recognize pictures of objects (Besson et al., 2012; Despouy et al., 2020).

In animal studies, memory processes have been widely studied using electrophysiological recordings. The reduced size of the brain limits the number of electrodes that can be implanted in the behaving animal. Therefore, most studies include recordings from only one or two regions, mainly in the hippocampal formation (Mizuseki et al., 2009; López-Madrona et al., 2020). However, their spatial resolution is much higher than human SEEG, with a single electrode composed of up to seven shanks in the array (Csicsvari et al., 2003) or containing hundreds of recording channels (Steinmetz et al., 2021).

For our data collected simultaneously at different scales (mesoscale – SEEG, macroscale – scalp) in human subjects, providing wide (SEEG) or whole-brain (surface EEG) spatial coverage, we have performed a high-sensitivity multivariate pattern analysis (MVPA) (Haxby et al., 2001; Grootswagers et al., 2017), not only on individual sets of signals of different modalities (intracranial, scalp, or reconstructions) but also on combined sets having higher dimensionality, to evidence possible synergies between signals recorded at different scales.

2. Methods

2.1. Subjects

We selected 12 patients diagnosed with focal drug-resistant epilepsy that underwent long-term simultaneous EEG and SEEG recordings in the Emergency University Hospital Bucharest between 2020 and 2022 (Table 1). Patients were considered surgical candidates and underwent pre-surgical non-invasive evaluation using extended patient history, video-electroencephalography, brain structural and functional imaging (inter-ictal FDG-PET CT), and neuropsychological profile. Consequently, in these patients, invasive recordings were considered necessary to delineate the epileptogenic zone and to map the functional cortex for tailoring the surgical resection (Munari et al., 1994; Kahane et al., 2003; Jayakar et al., 2016; Isnard et al., 2018). The details regarding the patients' gender, age, type of epilepsy, and lateralization are provided in Table 1. In addition, as part of this research protocol, scalp electrodes were attached, allowing for simultaneous surface and intracranial long-term recordings. This study included only patients with unmodified anatomy, no previous major resection, and no major cognitive deficit.

The study has been performed under Bucharest University ethical committee approval CEC 23/20.04.2019. All patients, or their legal guardian/next of kin, signed a written informed consent, in accordance with the Declaration of Helsinki, for the simultaneous recordings and data sharing procedures.

2.2. Experimental paradigm

We used the same experimental visual memory paradigm as in López-Madrona et al. (2022). In summary, we used 168 images from the database of Duñabeitia et al. (2018) that were organized in blocks of 12 or 24 images, presented on a computer screen. There were two block types: encoding ("ENC"), where a set of 12 images were presented to the patient, followed by a recognition block where the same 12 familiar images ("OLD") were randomly interleaved with other 12 novel images ("NEW"). The patient was required to indicate by pressing two buttons on the keyboard, using two fingers of the right hand, whether the images were familiar or not, within 1500 ms. A distracting video of 1 min was presented in between encoding and recognition blocks. The sequence of 36 image presentations was repeated seven times using different images from the 168-image set and pseudo-random distribution of the OLD and NEW items, with the constraint that there were never more than three "old" or "new" items in a row. Stimuli presentation and response logging were controlled by the software E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA).

2.3. Simultaneous scalp and intracranial recordings

Stereoelectroencephalographic exploration was performed using depth electrodes (Dixi Medical, Chaudefontaine, France)

TABLE 1 Patients included in this study.

Patient	ID	Age	Epilepsy	Lateralization	Language organization	SEEG electrodes	SEEG contacts	Scalp electrodes	SEEG electrode location
1	89	37	Insular	R	Left typical	14	172	30	Left, posterior
2	90	17	Insular-opercular	L	Left typical	9	86	30	Left, central
3	92	27	Insular	L	Left typical	10	145	30	Left, posterior
4	96	26	Temporal	R	Left typical	11	152	38	Right, anterior
5	97	26	Rolandic Operculum	L	Atypical bilateral	10	135	35	Left, central
6	98	39	Temporal	R	Left typical	9	129	38	Right, posterior
7	99	24	Insular	L	Left typical	13	189	38	Left, anterior
8	101	31	Temporal	B	Left typical	14	187	40	Bilateral, central
9	102	31	Temporo-insular	B	Left typical	16	229	40	Bilateral, posterior
10	104	20	Insular	L	Left typical	10	124	40	Left, central
11	105	26	Frontal	R	Left typical	12	161	37	Right, anterior
12	107	26	Frontal	L	Left typical	8	176	40	Left, anterior

with 8 to 18 contacts per electrode, 2 mm contact length, 3.5 mm center-to-center contact spacing, and 0.8 mm diameter. Multiple electrodes (9 – 16) were placed following a patient-specific hypothesis regarding the localization of the seizure onset zone and the pathways of ictal spread (Kahane et al., 2003; Jayakar et al., 2016), allowing for up to 229 contacts to be available in each patient. Electrodes were placed intracranially using the microTargeting™ Multi-Oblique Epilepsy STarFix Platform (FHC, Bowdoin, ME USA) (Dewan et al., 2018; Yu et al., 2018; Pistol et al., 2021) or the Leksell stereotactic frame (Elekta AB, Stockholm, Sweden). To determine the exact location of each electrode and contact, the post-implantation CT scan was loaded into surgical planning software (Waypoint Planner, FHC, Bowdoin, ME USA), co-registered with the pre-implantation MRI, and adjustments to the initially planned trajectories were made to match the postop location of the electrodes. Manual labeling of the SEEG contacts has been performed using the abbreviations listed in Supplementary Table S1.

In view of the group analysis, the pre-surgical MRI of each patient was also used for running an analysis pipeline implemented in FreeSurfer (Fischl, 2012) that allowed us to obtain the patient's cortical surface reconstruction, used for visualization purposes, but also—more importantly—, for performing a non-rigid registration of the patient's MRI to the "cvs_avg35_inMNI152" FreeSurfer template (Postelnicu et al., 2009), providing us with the coordinates of each intracranial contact in a common MNI space.

One up to three days after the SEEG implantation, between 30 and 40 scalp electrodes were placed according to the 10-20 system. A few electrodes were repositioned on adjacent 10-10 grid location, due to interference with the SEEG electrodes, and up to 10 electrodes could not be placed at all. The exact number of scalp electrodes in each patient is provided in Table 1.

Signals were collected using a setup as described by Barborica et al. (2021). In summary, two identical Natus Quantum 128-channel amplifiers (Natus Neuro, Middleton, WI) were used, one for each modality (scalp/intracranial) and having separate signal references. The reference for the SEEG recordings was chosen on one contact located in white matter exhibiting minimal activity, whereas the reference for the scalp system was Fpz. Raw data were acquired at a sample rate of 4096 Hz. The hardware was synchronized using digital triggers to both systems and a 50 Hz sine reference signal, recorded simultaneously using DC inputs of the two systems. Patients 9–12 were recorded with a single Quantum 256-channel amplifier that did not require external synchronization hardware. The data were combined and saved in a single file in AnyWave ADES format (Colombet et al., 2015), containing both types of signals. The analysis workflow is shown in Figure 1.

The synchronization between stimuli presentation and (S)EEG recordings has been performed using a photodiode part of the Chronos response box (Psychology Software Tools, Pittsburgh, PA) attached to a corner of the screen where trial start synchronization flashes were presented. The response time and correctness were merged into the AnyWave event file by reading the E-Prime log files using MATLAB (MathWorks, Natick, MA) custom scripts.

Only intracranial sensors located outside the seizure onset zone and gray matter were included in the analysis. Additional artifacted

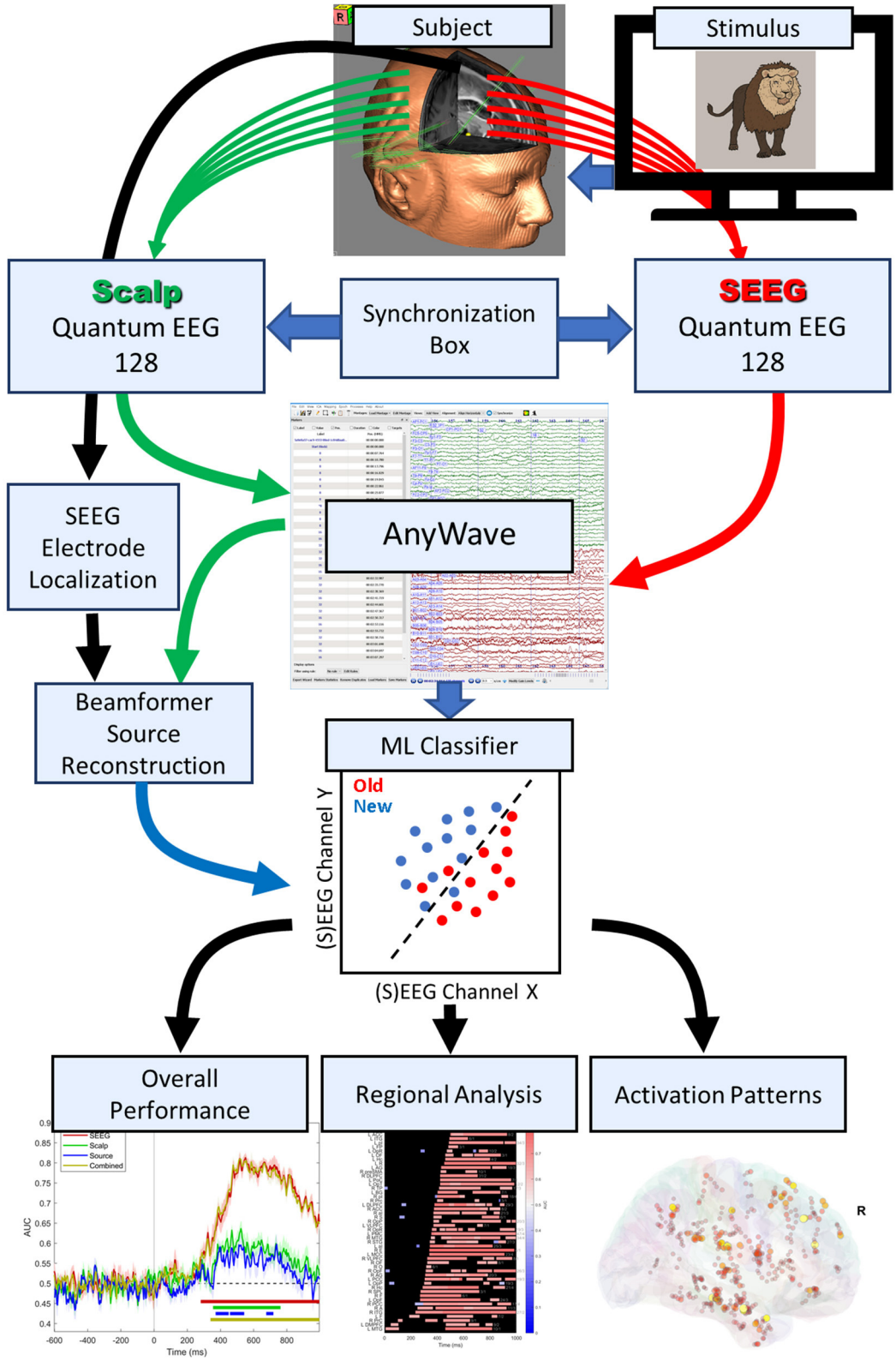


FIGURE 1
Signal collection and analysis workflow.

trial removal, as well as bad channel removal, was performed manually by visually inspecting the recordings.

2.4. ERP processing

Signals were loaded into EEGLab (Delorme and Makeig, 2004) software, resampled at 256 Hz, and filtered in the 0.5–45 Hz interval. Scalp EEG was re-referenced to the common average, and artifacts were removed using independent component analysis (ICA). Only correct trials have been retained for further analysis.

2.5. Source localization

To test the inverse solution of scalp EEG for finding brain areas that are involved in task decoding, we have calculated source signals at the location of the intracranial electrodes. To achieve that, we have performed a beamformer analysis on the standard FreeSurfer's *fsaverage* template, brain electrical model, and 10–20 electrode positions available in MNE-Python (Gramfort et al., 2014). The beamformer spatial filters calculated using linearly constrained minimum variance (LCMV) (Van Veen et al., 1997) were used to calculate source time courses on a 5-mm grid covering the brain. The source time course on the grid point nearest to the midpoint between a pair of SEEG contacts that were part of a bipolar-recorded signal was considered to approximate the source signal at each intracranial site. We, therefore, obtained a set of signals with the same dimensionality as the SEEG, which we analyzed using the common MVPA pipeline. Calculating the source signals at the locations near intracranial recording sites allowed the SEEG recordings to be used as ground truth for source reconstructions (Mikulan et al., 2020) and provided the ability to compare task-related activations at the same locations for both types of signals.

2.6. Independent component analysis

To test whether a method that is known to separate temporally correlated neuronal sources can enhance MVPA decoding results, we have performed an independent component analysis (ICA) of scalp signals using second-order blind identification (SOBI) blind source separation (Belouchrani et al., 1993, 1997; Tang et al., 2005), using EEGLab software (Delorme and Makeig, 2004).

2.7. Multivariate pattern analysis

For multivariate pattern analysis, we have generally followed the workflow described in Grootswagers et al. (2017). The processing has been performed using the MNE-Python toolbox (Gramfort et al., 2013, 2014) and custom Python and MATLAB (MathWorks, Natick, MA) scripts. A logistic regression linear classifier was trained to discriminate between responses for the OLD and NEW conditions using the L-BFGS-B – (Large-scale Bound-constrained Optimization) solver. The model was fitted to the standardized data, and its performance was scored using the receiver operating characteristic (ROC) area under the curve (AUC). The scores were evaluated using 20-fold cross-validation, and time intervals where they were statistically different from chance were evaluated using a one-sample permutation cluster test applied to the set of scores calculated for each fold (Maris and Oostenveld, 2007).

The processing pipeline was applied to SEEG bipolar signals, to the EEG signals, to the entire set of independent components of scalp EEG, or to the scalp source signals at the SEEG sensor location obtained using a beamformer. Specific to our study, the simultaneous collection of the scalp and SEEG data allowed the pooling of the signals for the two modalities to investigate whether combined data provide a classifier performance significantly different from analyzing individual sets.

We have calculated the contribution of signals at each intracranial sensor location (recorded or reconstructed) to the recognition process by calculating the activation patterns associated with fitting the data with a linear model (Haufe et al., 2014) using the MNE-Python toolbox, which in turn resorts extensively to scikit-learn Python toolbox (Pedregosa et al., 2011; Abraham et al., 2014). In contrast to the classifier weights associated with each sensor, which do not have a direct interpretation, the reconstructed activation patterns are interpretable as neural sources encoding the studied processes that can be projected onto the sensors (Haufe et al., 2014; Fahrenfort et al., 2018).

To assess the contribution of various brain structures to decoding task conditions, we repeated the MVPA analysis on a subset of signals recorded or reconstructed within the same brain area or structure (Despouy et al., 2020), according to the labeling we have described earlier in this section. We will further refer to this analysis restricted to a region of interest (ROI) as "regional analysis" (Ebrahiminia et al., 2022). In contrast to activation patterns (Haufe et al., 2014), which have no statistical significance associated with their time courses, regional analysis allows inferring, in a probabilistic way, the time intervals where decoding performance is different from chance, evidencing the sequential/hierarchical processing of stimulus novelty within the brain.

3. Results

A total of 136 intracranial electrodes having 1,885 contacts were implanted in 12 patients. Additional 436 surface electrodes were attached to the scalp. After data curation and application of inclusion criteria, signals recorded at 965 intracranial sites and 391 scalp locations were further included in the analysis. The subjects correctly identified stimulus novelty in 89.53% of the trials. The MVPA analysis was applied to 1,729 correct recognition trials (OLD: 822, NEW: 907) having a mean ± SD response time of 719.1 ± 162.4 ms (OLD) and 765.0 ± 191.2 ms (NEW).

3.1. Responses on the single scalp and SEEG electrodes

The ERPs for the scalp sensor and SEEG sensor having the highest magnitude multivariate activation patterns among all scalp

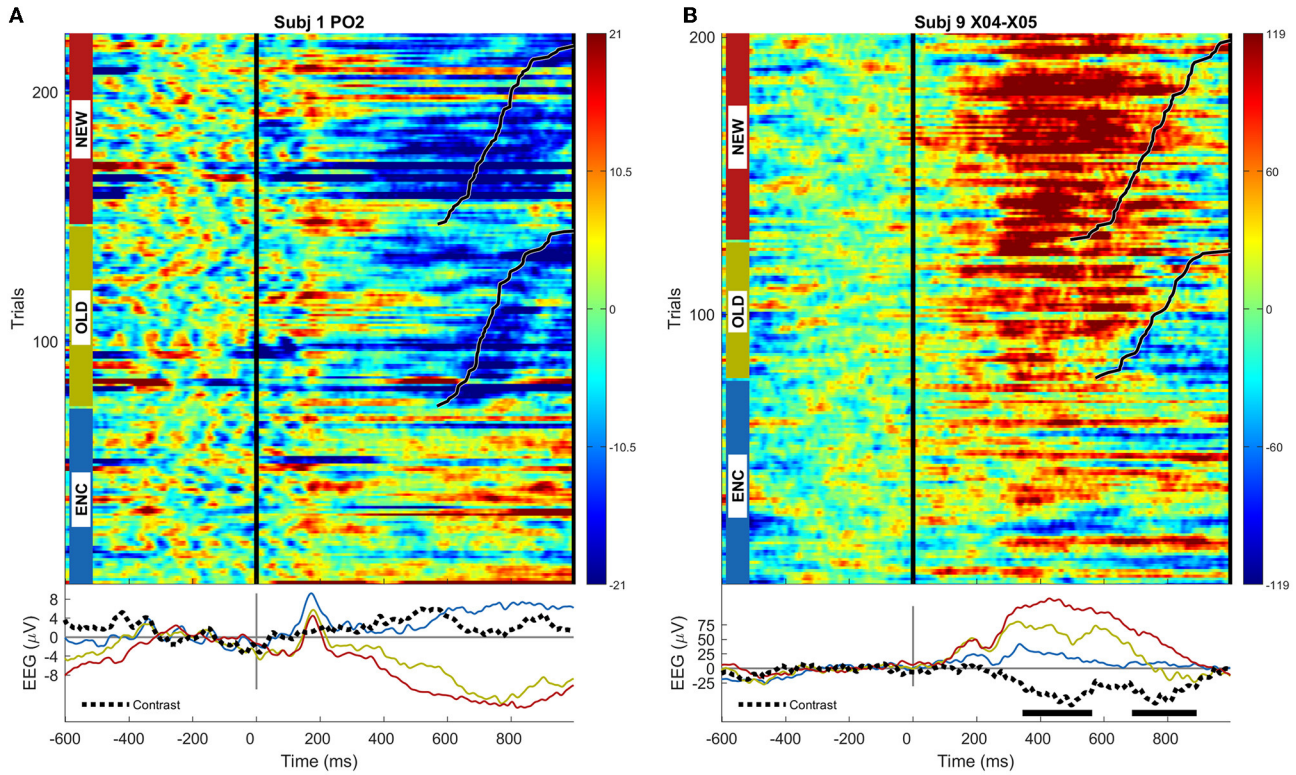


FIGURE 2
(A) ERP image for the scalp sensor PO2 in subject 1, exhibiting the highest multivariate activation pattern; trials are grouped by condition and sorted by the response time, marked using black lines; average ERPs for each condition, as well as the contrast between OLD and NEW conditions, are shown; the statistical significance of the univariate (permutation cluster test) difference, if present, between OLD and NEW conditions at a significance level $p < 0.05$ is shown using thick horizontal lines; **(B)** same as **(A)**, but for the intracranial sensor X04-X05 located in right anterior insula in subject 9.

and $m = 965$ SEEG signals recorded in all $n = 12$ patients are shown in Figure 2.

While a typical high-amplitude ERP presents prominent peaks either following the stimulus presentation (∼200 ms) or around response time, depending on sensor location, these examples rather capture situations where the novelty of the stimulus is best captured, between 400 ms and 600 ms and around the response time (∼800 ms).

3.2. Single-subject multivariate analysis

The results of the MVPA analysis of responses at the SEEG, scalp, source level, and combined scalp SEEG in patient 3 are shown in Figure 3. The classifier performance for the SEEG signals is consistently above-chance through the interval ∼450 ms through ∼900 ms (permutation cluster test, $p < 0.05$). By contrast, the scalp signals provide a statistically significant classification performance only during the memory retrieval and stimulus recognition processes between ∼500 ms and ∼600 ms. Computing source signals at SEEG sensor locations provide classification results that are similar in magnitude to the scalp sensor signals, with eventually better results in terms of the extent of the clusters reflecting the scores significantly different from chance ($p < 0.05$).

A regional MVPA analysis presented in Figures 3C, D highlights the regions that contribute most to the overall decoding performance, namely the anterior cingulate cortex and hippocampus. The ACC, as sampled by SEEG, exhibits sustained better-than-chance scores in the late interval ∼500 ms through ∼900 ms, whereas Hc presents early (∼500 ms), but limited duration (∼100 ms) activations. The scalp, source, and combined signals provide similar results in Hc, but rather different ones in ACC.

3.3. Group analysis

At the population level ($n = 12$ subjects), the classifier performance based on intracranial signals was much higher than the one based on scalp or source signals, as shown in Figure 4.

The use of source signals calculated at SEEG sensor locations provides slightly lower classifier performance than the one based on signals from which it was derived, i.e., scalp signal (Figure 4). The MVPA analysis applied to the independent components of the scalp signal provides results that are virtually identical to the scalp ones. Combined scalp and SEEG scores follow closely the time course of the SEEG scores.

The time course of classification performance using SEEG signals is consistent across subjects, as shown in Figure 5, where

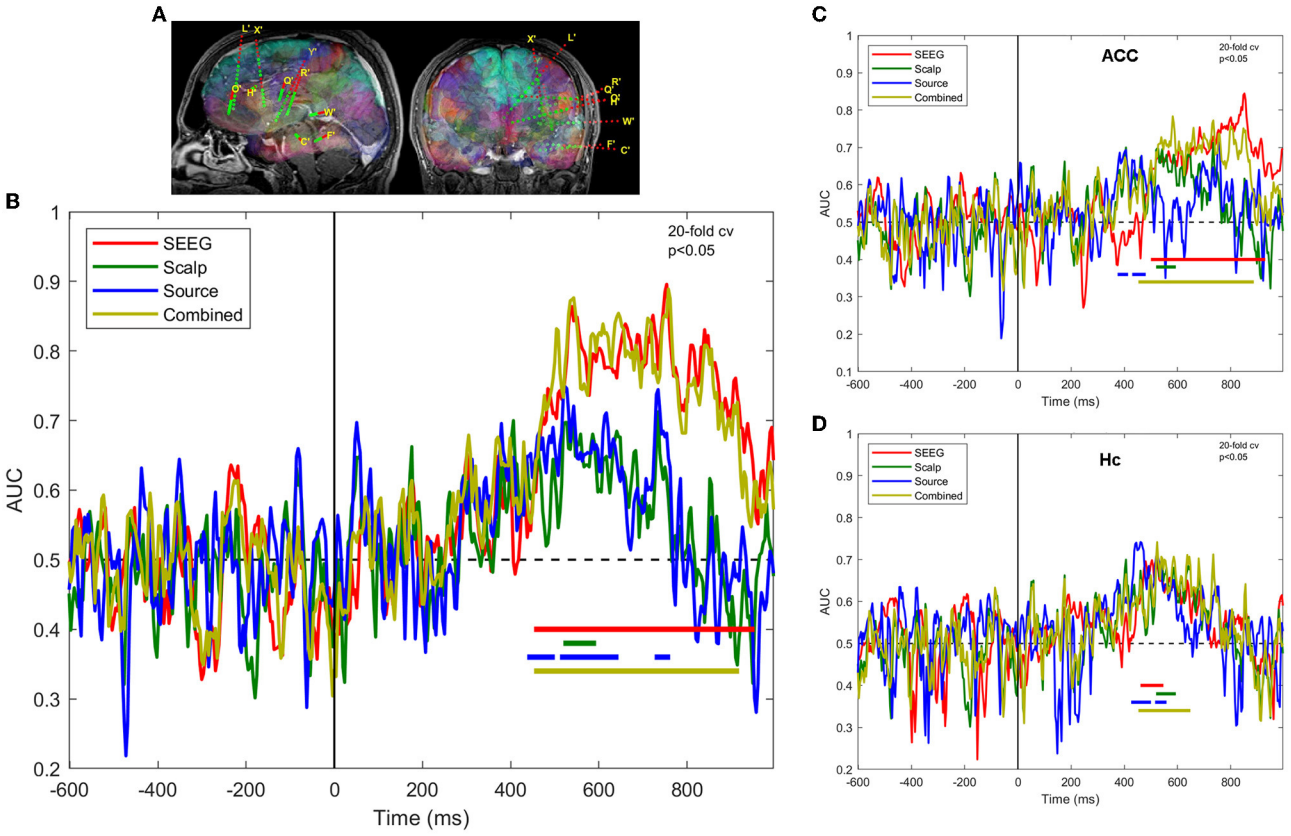


FIGURE 3
Task decoding performance is expressed as the area under curve of the receiver operating characteristic of the classifier for SEEG, scalp, and source signals in patient 3. **(A)** SEEG electrode locations in the left hemisphere; **(B)** ROC-AUC for sensors of different types, as well as for combined scalp and SEEG; **(C)** same as **(B)**, but for the contacts located in the anterior cingulate cortex; **(D)** same as **(B)**, but for contacts located in the hippocampus.

we have plotted the scores for all subjects as well as the grand average. This is somehow unexpected, as the areas implanted with depth electrodes can be quite different. In Figures 5A, C, we show two implantation schemes providing similar scores, which are highlighted in Figure 5E with green and blue colors.

When we perform a regional analysis of the performance in decoding task conditions, we see that the findings at the level of all $m = 965$ sites in $n = 12$ subjects, shown in Figure 4, are confirmed at a regional scale (Figure 6), with scores significantly different from chance associated with SEEG signals being higher and more sustained over time, compared to source signals reconstructed at the same locations. Among the areas exhibiting the highest and earliest SEEG scores, we can count F, ITG, and Hc, as well as the insular-opercular complex. One has to keep in mind that all these findings are strongly influenced by the coverage of each ROI with SEEG electrodes.

The 3D representation of multivariate activation patterns (Haufe et al., 2014) of SEEG and source-space data is shown in Figure 7. One has to keep in mind that these activation patterns do not reflect the magnitude of the ERPs, but rather represent a virtual signal corresponding to how well a site encodes the stimulus novelty, in our case (Wardle et al., 2016; Grootswagers et al., 2017). A wide-area brain activation (Figure 7A) over the course of the recognition process is visible for the intracranial signals; whereas at a comparable amplitude scale, the source data show much less activations. The activation patterns of various brain areas are sequential, following a posterior-to-anterior flow, as illustrated in Figure 7 and in the Supplementary Data Movie. The activations associated with EEG source signals show a roughly similar spatiotemporal pattern. At a closer visual inspection of Figure 7, we can find evidence of known leakage-related effects (Schoffelen and Gross, 2009), as multiple contacts in several electrodes exhibit similar activation values.

In SEEG recordings (Figure 7A), we can divide the activation into four clinically relevant time intervals. The significant activation starts at ∼20 0ms and between 200 and 400 ms, we can observe the recognition process that activates the network of structures that mainly involves the temporal-basal and hippocampus on the right side. Then, between 400 and 600 ms, we can see the activations related to the decision-making process that significantly involves bilaterally the perisylvian, prefrontal, and mesial temporal lobe structures. The sensorimotor activation overlaps the 400–600 ms and continues in the next interval of 600–800 ms and represents the response phase of the task. The last time interval (600–800 ms) highlights the activation of the prefrontal cortex possibly related to self-evaluation or memory storage. The EEG source

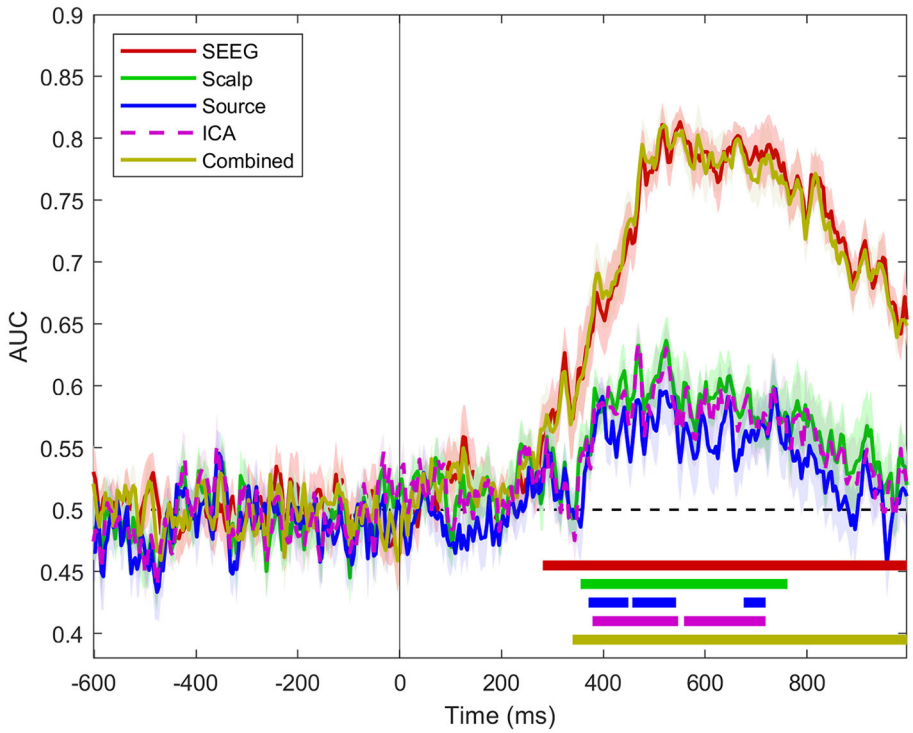


FIGURE 4
Classifier performance for SEEG, scalp, source, ICA, and combined scalp SEEG signals for all $n = 12$ subjects. The dashed areas show standard error intervals for the set of classifier scores for all patients. The horizontal bars indicate the intervals where the scores are statistically different from chance (one-sample permutation cluster test, $p < 0.05$).

(Figure 7B) displays a similar timeline of the activation pattern in the 200–600 ms. However, the late phase between 600 and 800 ms is not informative.

4. Discussion

While other studies comparing intracranial to scalp data used a sequential recording of the two modalities (Ebrahiminia et al., 2022) or even different sets of participants (Haufe et al., 2018), we have simultaneously acquired data in the two modalities, allowing us to validate the results of the EEG source reconstruction using SEEG recordings in decoding task conditions, as well as investigate the possible synergy between invasive and non-invasive recording in decoding stimulus novelty.

Our results show that our task requiring the subjects to categorize visual stimuli based on novelty, involving memory encoding and retrieval, activates large areas of the brain. This finding is supported by the widespread activation visible in Figure 7, as well as by the fact that SEEG implantations at totally different locations result in decoding performances over time that are close to each other and to the group average (Figure 5).

The decoding performance of the ML classifier is maximal when using intracranial signals, although the SEEG implantation has limited spatial coverage of the brain, compared with the scalp EEG which is supposed to provide full-brain coverage, as visible in Figures 3, 4. The relatively poor decoding performance of the classifier that uses scalp signals can be attributed, in our opinion, to the significantly lower signal-to-noise ratio (SNR) of the scalp EEG compared to SEEG. It is also possible that scalp EEG provides poor visibility of activity in deep structures of the brain, whereas SEEG samples with undegraded SNR all implanted locations, regardless of depth. A previous study by Ebrahiminia et al. (2022) performing sequential scalp and electrocorticographic (ECoG) recordings has shown that scalp EEG provides slightly better classification performance of passively viewing visual stimuli of different categories (Liu et al., 2009). Without counting the differences in the tasks, one reason for this discrepancy may relate, once again, to the fact that ECoG does not record activity in deep brain structures, therefore both modalities provide information from the outer cortex, with scalp EEG providing slightly better spatial coverage. Another factor that may favor EEG in other studies is that in our simultaneous protocol, the EEG electrodes were glued to the scalp 1 day or more before running the memory task (part of a wider set of investigations), resulting in a degradation of the quality of the contact within this interval, uncorrectable due to the requirement of maintaining sterility at the scalp level. Furthermore, due to spatial constraints related to pre-existing SEEG electrode anchors, the coverage with scalp electrodes was non-uniform.

Interestingly, using SEEG electrodes, the classifiers were always able to decode the task conditions using task-evoked intracranial EEG recorded 300 to 1000 ms post-stimuli presentation. This was true not only at the group, but also at the individual subject level, even when the spatial sampling of the SEEG

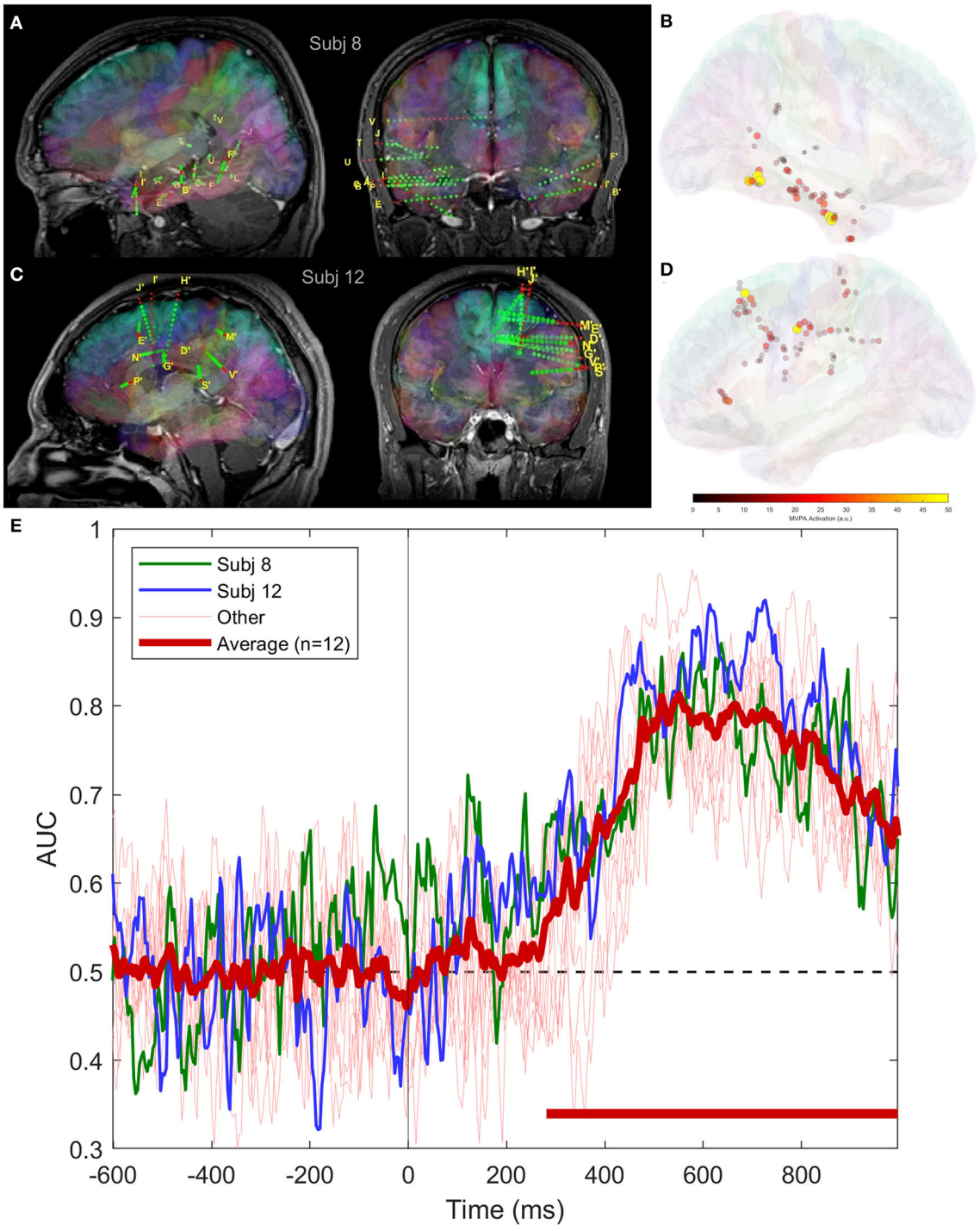


FIGURE 5
Classifier performance using intracranial signals for two patients having SEEG implantation covering different areas of the brain; **(A)** bilateral implantation in subject 8, covering temporal lobe, including mesial structures; **(B)** mean magnitude of activation patterns in subject 8 across the entire trial duration; **(C)** electrode locations in subject 12, frontal, parietal and cingulate areas; **(D)** same as **(B)**, but for subject 12; and **(E)** average and individual classifier scores.

electrodes was completely different (Figure 5). Recent studies have shown the "traveling wave" behavior of brain activity (Lubenov and Siapas, 2009; Muller et al., 2014; Liang et al., 2021; Bhattacharya et al., 2022), and it is possible that we have observed such effects in our analysis. Under the assumption that the task-evoked intracranial EEG activity is recorded on a critical number of electrodes, sufficient for the classifier to learn the propagation patterns of the traveling wave, we may decode the task conditions from various brain regions, without a loss in decoding performance. Similar effects have been observed by

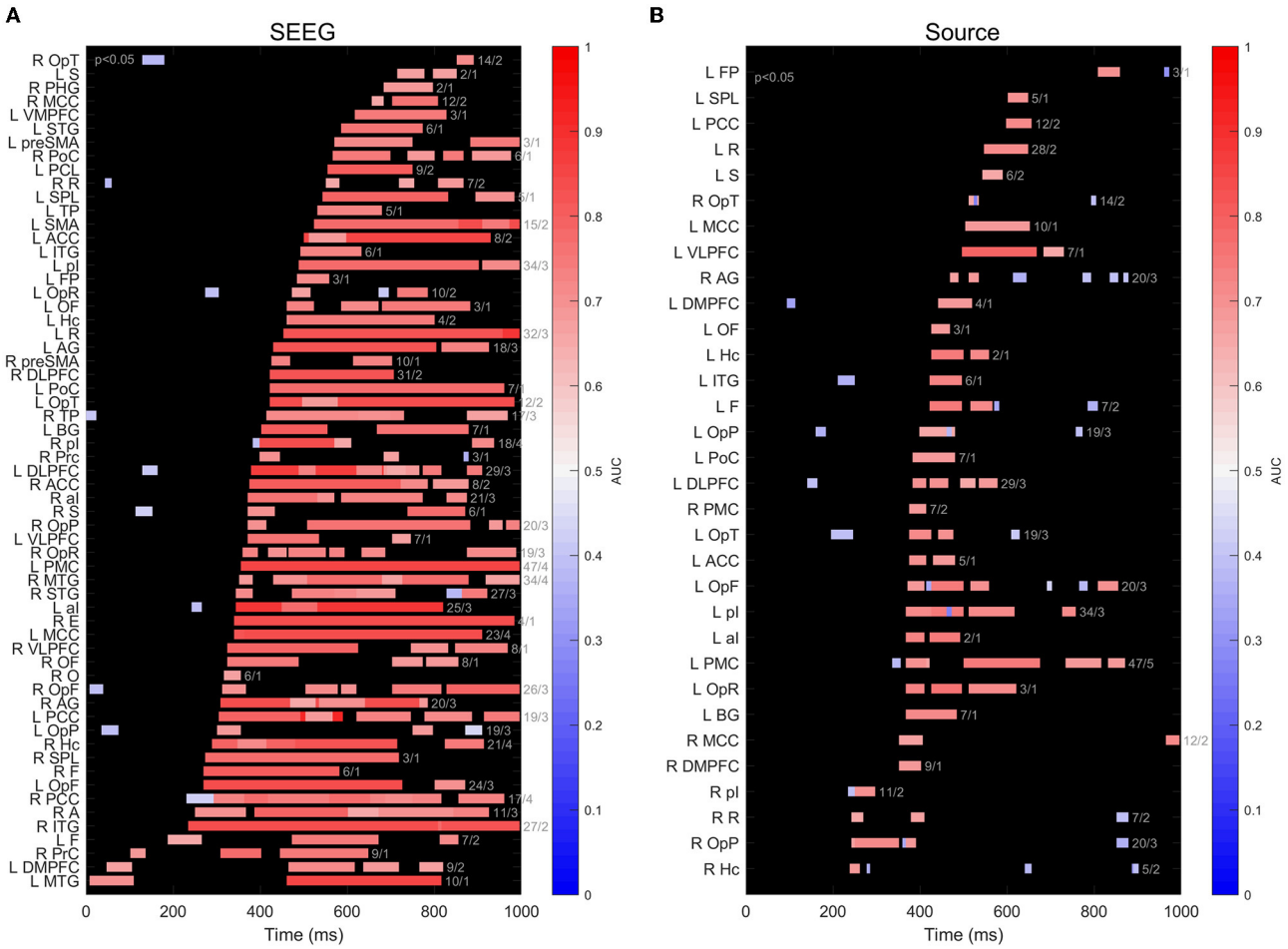


FIGURE 6
(A) Timeline of decoding performance significantly different from chance ($p < 0.05$) for signals recorded on subsets of intracranial contacts implanted in different brain structures. The color of the bars indicates the maximum value of the AUC score within a cluster. The numbers at the right of each bar indicate the number of sites and number of patients for clusters in each ROI; **(B)** same as **(A)**, but for scalp sources calculated at the location of intracranial contacts using beamformer.

groups that studied the representation and processing of emotion in the brain with machine learning methods, concluding that emotion representation is encoded as patterns of activations over widely distributed brain networks (Wager et al., 2015; Donos et al., 2022).

The process of reconstructing the EEG source signals using beamforming does not result in a significant improvement at the population level of the classifier's performance, yielding results comparable to signals on scalp sensors, as shown in Figure 4. There are exceptions to that general finding in some individual patients, as illustrated in Figure 3B, where the decoding performance of a classifier operating on source signals show earlier and longer statistically significant above-chance scores than sensor-based analysis, at a significance level of $p < 0.05$. However, such results have to be treated with caution, given the probabilistic nature of the statistical tests applied (Sassenhagen and Draschkow, 2019). The regional analysis of the classification performance shown in Figure 6 is in agreement with the overall results in Figure 4, where source signals result in more sparse and limited-duration significant scores than the intracranial signals.

The beamformer source reconstruction is based on linear matrix operations on the responses (Westner et al., 2022). A linear transformation based on a square non-singular matrix is equivalent to an affine transformation in the n-dimensional response space, which is the space in which the MVPA operates (Grootswagers et al., 2017).

Another approach that also uses linear matrix transformations for separating statistically independent components in a set of signals is the independent component analysis (ICA). We have tested whether applying ICA to the scalp EEG responses, results in a set of independent components that provide better decoding of the task conditions. The results, presented in Figure 4, show that classifier performance operating on the full set of independent components, without dimensionality reduction, is virtually identical to the one for the original signals on the scalp sensors.

These data-driven findings related to deriving virtual signals/components based on linear transformations suggest more general theoretical considerations: a non-singular

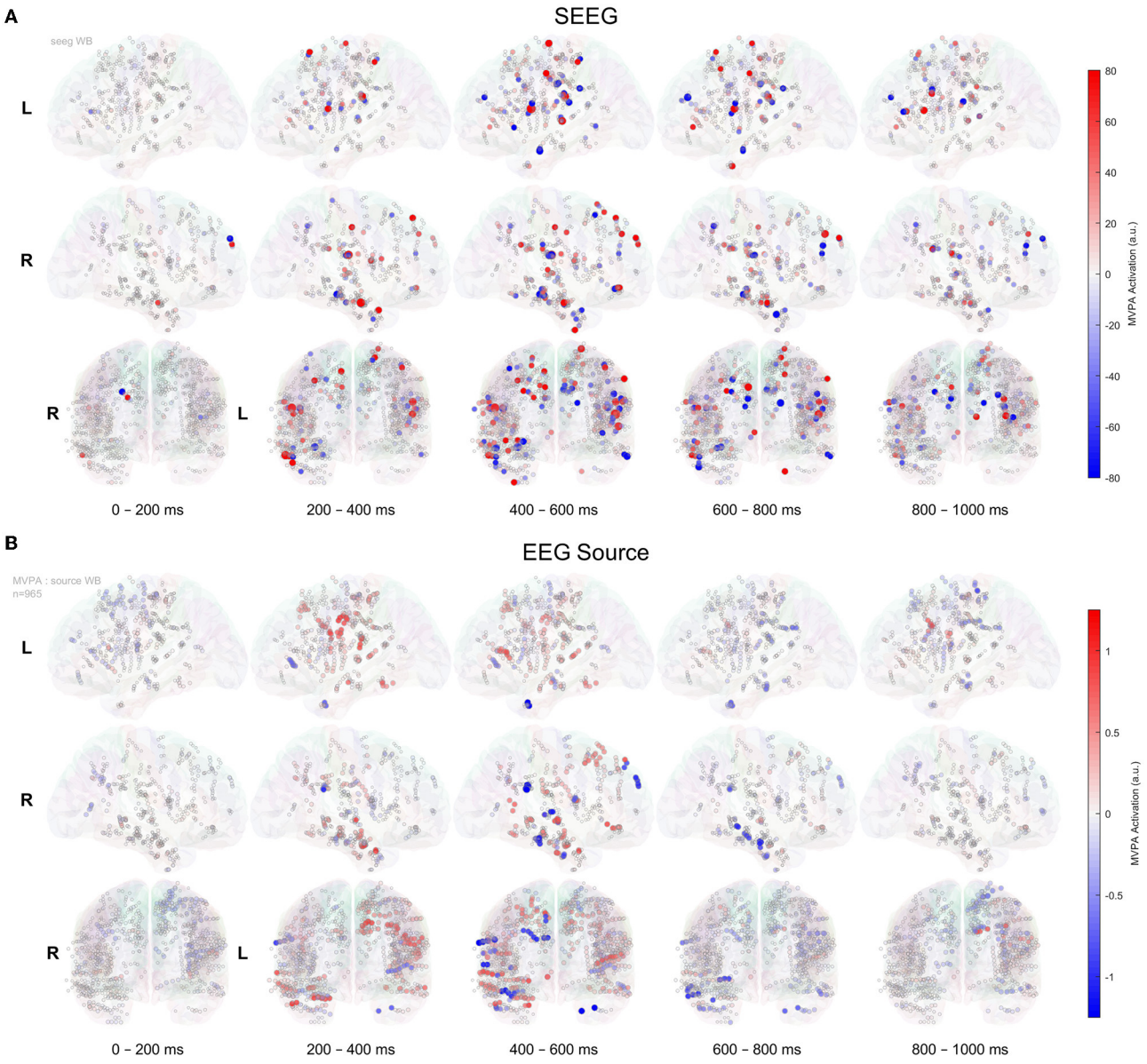


FIGURE 7
MVPA timeline of activation patterns, as introduced by Haufe et al. (2014), for SEEG signals **(A)** and in EEG source space **(B)** for all 965 contacts implanted in 12 patients, shown on the glass brain. The mean of activation patterns' values within 200 ms bins is represented.

affine transformation, equivalent to a series of elementary transformations such as rotation, scaling, and shear (Goldman, 1992), that do not change the relationships between points representing the set of *n* responses at a particular point in time; therefore, it is not expected to significantly affect the performance of an ML classifier operating on the transformed set of points. It should be noted that only linear transformations that do not perform dimensionality reduction (i.e., the transformation matrix is square), such as the ones we have considered, are equivalent to affine transformations and thus unlikely to affect MVPA classifier accuracy, based on geometrical considerations. By contrast, other approaches based on non-square linear transformation matrices, such as principal component analysis with dimensionality reduction, have been shown to have the potential for improving ML classifiers' accuracy (Grootswagers et al., 2017; Hatamimajoumerd et al., 2020).

In investigating whether scalp and intracranial signals contain complementary information that might contribute to a classifier performance, we did find that the modality providing the best performance (i.e., SEEG) is determining the combined performance (Figures 3, 4).

A limitation of the study is the partial and non-uniform spatial sampling of both scalp and intracranial sensors, due to objective reasons. The source reconstructions are based on the *fsaverage* template and associated volume conduction model, without taking into account the individual patient anatomy, which might lead to different results (Ramon et al., 2006; Céspedes-Villar et al., 2020). The scalp electrode locations used for calculating inverse solution

were the standard 10–20 ones, aligned with the *fsaverage* head model, without taking into account the displacements related to the inaccuracy of the electrode positioning and the avoidance of the SEEG electrodes. Only one source-reconstruction-algorithm with fixed parametrization, among many different possible ones, has been considered (Grech et al., 2008; Mikulan et al., 2020). Another limitation is that our analysis pipeline is the most conservative one, being based on wide-band single-trial data. Creating "super-trials" or "pseudo-trials" by averaging several trials (Despouy et al., 2020; Ashton et al., 2022) might improve the SNR of EEG and correspondingly of the source reconstruction signals. Further measures for improving SNR can be possibly implemented (Grootswagers et al., 2017), alleviating some of the apparent limitations of non-invasive recordings.

5. Conclusion

Analysis of invasive EEG provides the highest amount of information related to stimulus novelty, compared with scalp recordings, despite the limited spatial sampling of the brain with depth electrodes. This may be related to the limited scalp visibility of the activity related to memory processes in deep brain structures, particularly if containing higher frequency components. The synergy between the two modalities—enabled by pooling data recorded simultaneously—is limited, with the SEEG sensors providing the best decoding performance driving the combined, overall, performance.

Data availability statement

The datasets presented in this study can be found in online repositories. The URL of the repository is: http://epi.fizica.unibuc.ro/scalesoldnew/.

Ethics statement

The studies involving human participants were reviewed and approved by the University of Bucharest Ethics Committee for Research, approval number CEC 23/20.04.2019. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

AB: conceptualization, methodology, software, formal analysis, resources, data curation, writing, visualization, supervision, and funding acquisition. IM: methodology, investigation, writing, and supervision. VL-M and AT: methodology. F-XA: methodology and writing—review and editing. CD: methodology, software, formal analysis, and writing—initial draft. IO: investigation. CP: investigation, formal analysis, and data curation. FM: investigation and data curation. CB: conceptualization, methodology, writing—review and editing, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Romanian UEFISCDI COFUND-FLAGERA II-SCALES, PN-III-P4-ID-PCE-2020-0935, and Agence Nationale de la Recherche ANR-17-HBPR-0005 SCALES.

Acknowledgments

The authors would like to thank Cornel Tudor, Aurelia Dabu, Jean Ciurea, for their contributions to performing the SEEG implantations and surgical procedures, as well as Flavius Bratu and Camelia Lentoiu for their contribution to collecting the clinical data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2023.1154038/full#supplementary-material

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* 8, 14. doi: 10.3389/fninf.2014.00014

Antony, A. R., Abramovici, S., Krafty, R. T., Pan, J., Richardson, M., Bagic, A., et al. (2019). Simultaneous scalp EEG improves seizure lateralization during unilateral intracranial EEG evaluation in temporal lobe epilepsy. *Seizure* 64, 8–15. doi: 10.1016/j.seizure.2018.11.015

Ashton, K., Zinszer, B. D., Cichy, R. M., Nelson, C. A., Aslin, R. N., and Bayet, L. (2022). Time-resolved multivariate pattern analysis of infant EEG data: A practical tutorial. *Dev. Cogn. Neurosci.* 54, 101094. doi: 10.1016/j.dcn.2022.101094

Barborica, A., Mindruta, I., Sheybani, L., Spinelli, L., Oane, I., Pistol, C., et al. (2021). Extracting seizure onset from surface EEG with independent component analysis: Insights from simultaneous scalp and intracerebral EEG. *NeuroImage. Clin.* 32, 102838. doi: 10.1016/j.nicl.2021.102838

Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., and Moulines, É. (1993).Second Order Blind Separation of Temporally Correlated Sources. *Proc. Int. Conf. Digit. Signal Process.*, 346–351.

Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., and Moulines, É.(1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* 45, 434–444. doi: 10.1109/78.554307

Besson, G., Ceccaldi, M., Didic, M., and Barbeau, E. J. (2012). The speed of visual recognition memory. *Vis. cogn.* 20, 1131–1152. doi: 10.1080/13506285.2012.724034

Bhattacharya, S., Brincat, S. L., Lundqvist, M., and Miller, E. K. (2022). Traveling waves in the prefrontal cortex during working memory. *PLOS Comput. Biol.* 18, e1009827. doi: 10.1371/journal.pcbi.1009827

Céspedes-Villar, Y., Martinez-Vargas, J. D., and Castellanos-Dominguez, G. (2020). Influence of patient-specific head modeling on EEG source imaging. *Comput. Math. Methods Med.* 2020, 5076865. doi: 10.1155/2020/5076865

Colombet, B., Woodman, M., Badier, J. M., and Benar, C. G. (2015). AnyWave: a cross-platform and modular software for visualizing and processing electrophysiological signals. *J. Neurosci. Methods* 242, 118–126. doi: 10.1016/j.jneumeth.2015.01.017

Csicsvari, J., Henze, D. A., Jamieson, B., Harris, K. D., Sirota, A., Barth,ó, P., et al. (2003). Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J. Neurophysiol.* 90, 1314–1323. doi: 10.1152/jn.00116.2003

Delorme, A., and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Despouy, E., Curot, J., Deudon, M., Gardy, L., Denuelle, M., Sol, J. C., et al. (2020). A fast visual recognition memory system in humans identified using intracerebral ERP. *Cereb. Cortex* 30, 2961–2971. doi: 10.1093/cercor/bhz287

Dewan, M. C., Shults, R., Hale, A. T., Sukul, V., Englot, D. J., Konrad, P., et al. (2018). Stereotactic EEG via multiple single-path omnidirectional trajectories within a single platform: Institutional experience with a novel technique. *J. Neurosurg.* 129, 1173–1181. doi: 10.3171/2017.6.JNS17881

Donos, C., Blidarescu, B., Pistol, C., Oane, I., Mindruta, I., Barborica, A., et al. (2022). A comparison of uni- and multi-variate methods for identifying brain networks activated by cognitive tasks using intracranial EEG. *Front. Neurosci.* 16, 946240. doi: 10.3389/fnins.2022.946240

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., et al. (2018). MultiPic: a standardized set of 750 drawings with norms for six european languages. *Q. J. Exp. Psychol.* 71, 808–816. doi: 10.1080/17470218.2017.1310261

Ebrahiminia, F., Cichy, R. M., and Khaligh-Razavi, S., M. (2022). A multivariate comparison of electroencephalogram and functional magnetic resonance imaging to electrocorticogram using visual object representations in humans. *Front. Neurosci.* 16, 983602. doi: 10.3389/fnins.2022.983602

Fahrenfort, J. J., van Driel, J., van Gaal, S., and Olivers, C. N. L. (2018). From ERPs to MVPA using the amsterdam decoding and modeling toolbox (ADAM). *Front. Neurosci.* 12, 368. doi: 10.3389/fnins.2018.00368

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Goldman, R. N. (1992). "Decomposing Linear and Affine Transformations," in *Graphics Gems III (IBM Version)*, ed. D. B. T. Kirk (San Francisco: Morgan Kaufmann), 108–116.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7, 267. doi: 10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., et al. (2008). Review on solving the inverse problem in EEG source analysis. *J. Neuroeng. Rehabil.* 5, 25. doi: 10.1186/1743-0003-5-25

Grootswagers, T., Wardle, S. G., and Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cogn. Neurosci.* 29, 677–697. doi: 10.1162/jocn_a_01068

Hatamimajoumerd, E., Talebpour, A., and Mohsenzadeh, Y. (2020). Enhancing multivariate pattern analysis for magnetoencephalography through relevant sensor selection. *Int. J. Imaging Syst. Technol.* 30, 473–494. doi: 10.1002/ima.22398

Haufe, S., DeGuzman, P., Henin, S., Arcaro, M., Honey, C. J., Hasson, U., et al. (2018). Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. *Neuroimage* 179, 79–91. doi: 10.1016/j.neuroimage.2018.06.016

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., Pietrini, P., et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430. doi: 10.1126/science.1063736

Isnard, J., Taussig, D., Bartolomei, F., Bourdillon, P., Catenoix, H., Colnat-Coulbois, S., et al. (2018). French guidelines on stereoelectroencephalography (SEEG). *Neurophysiol. Clin.* 48, 5–13. doi: 10.1016/j.neucli.2017.11.005

Jayakar, P., Gotman, J., Harvey, A. S., Palmini, A., Tassi, L., Schomer, D., et al. (2016). Diagnostic utility of invasive EEG for epilepsy surgery: Indications, modalities, and techniques. *Epilepsia* 57, 1735–1747. doi: 10.1111/epi.13515

Kahane, P., Minotti, L., Hoffmann, D., Lachaux, J. P., and Ryvlin, P. (2003). *Invasive EEG in the definition of the seizure onset zone: depth electrodes Handbook of Clinical Neurophysiology* (Amsterdam: Elsevier), 109–133. doi: 10.1016/S1567-4231(03)03009-0

Kappenman, E. S., and Luck, S. J. (2011). *The Oxford Handbook of Event-Related Potential Components*. Oxford: Oxford University Press

Koessler, L., Cecchin, T., Colnat-Coulbois, S., Vignal, J. P., Jonas, J., Vespignani, H., et al. (2015). Catching the invisible: mesial temporal source contribution to simultaneous EEG and SEEG recordings. *Brain Topogr.* 28, 5–20. doi: 10.1007/s10548-014-0417-z

Liang, Y., Song, C., Liu, M., Gong, P., Zhou, C., Knöpfel, T., et al. (2021). Cortex-wide dynamics of intrinsic electrical activities: propagating waves and their interactions. *J. Neurosci.* 41, 3665–3678. doi: 10.1523/JNEUROSCI.0623-20.2021

Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290. doi: 10.1016/j.neuron.2009.02.025

López-Madrona, V. J., Medina Villalon, S., Badier, J. M. Trébuchon, A., Jayabal, V., Bartolomei, F., et al. (2022). Magnetoencephalography can reveal deep brain network activities linked to memory processes. *Hum. Brain Mapp.* 43, 4733–4749. doi: 10.1002/hbm.25987

López-Madrona, V. J., Pérez-Montoyo, E., Álvarez-Salvado, E., Moratal, D., Herreras, O., Pereda, E., et al. (2020). Different theta frameworks coexist in the rat hippocampus and are coordinated during memory-guided and novelty tasks. *Elife* 9. doi: 10.7554/eLife.57313.sa2

Lubenov, E. V., and Siapas, A. G. (2009). Hippocampal theta oscillations are travelling waves. *Nature* 459, 534–539. doi: 10.1038/nature08010

Mandler, G. (1980). Recognizing: the judgment of previous occurrence. *Psychol. Rev.* 87, 252–271. doi: 10.1037/0033-295X.87.3.252

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Merkow, M. B., Burke, J. F., and Kahana, M. J. (2015). The human hippocampus contributes to both the recollection and familiarity components of recognition memory. *Proc. Natl. Acad. Sci. U. S. A.* 112, 14378–14383. doi: 10.1073/pnas.1513145112

Mikulan, E., Russo, S., Parmigiani, S., Sarasso, S., Zauli, F. M., Rubino, A., et al. (2020). Simultaneous human intracerebral stimulation and HD-EEG, ground-truth for source localization methods. *Sci. Data* 7, 127. doi: 10.1038/s41597-020-0467-x

Mizuseki, K., Sirota, A., Pastalkova, E., and Buzsáki, G. (2009). Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron* 64, 267–280. doi: 10.1016/j.neuron.2009.08.037

Muller, L., Reynaud, A., Chavane, F., and Destexhe, A. (2014). The stimulus-evoked population response in visual cortex of awake monkey is a propagating wave. *Nat. Commun. 2014 51* 5, 1–14. doi: 10.1038/ncomms4675

Munari, C., Hoffmann, D., Francione, S., Kahane, P., Tassi, L., Lo Russo, G., et al. (1994). Stereo-electroencephalography methodology: advantages and limits. *Acta Neurol. Scand.* 152, 56–67. doi: 10.1111/j.1600-0404.1994.tb05188.x

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

Pistol, C., Daneasa, A., Ciurea, J., Rasina, A., Barborica, A., Oane, I., et al. (2021). Accuracy and safety of customized stereotactic fixtures for stereoelectroencephalography in pediatric patients. *Stereotact. Funct. Neurosurg.* 99, 17–24. doi: 10.1159/000510063

Pizzo, F., Roehri, N., Medina Villalon, S., Trébuchon, A., Chen, S., Lagarde, S., et al. (2019). Deep brain activities can be detected with magnetoencephalography. *Nat. Commun.* 10, 971. doi: 10.1038/s41467-019-08665-5

Postelnicu, G., Zollei, L., and Fischl, B. (2009). Combined volumetric and surface registration. *IEEE Trans. Med. Imaging* 28, 508–522. doi: 10.1109/TMI.2008.2004426

Ramon, C., Schimpf, P. H., and Haueisen, J. (2006). Influence of head models on EEG simulations and inverse source localizations. *Biomed. Eng. Online* 5, 10. doi: 10.1186/1475-925X-5-10

Ratcliff, R., Sederberg, P. B., Smith, T. A., and Childers, R. (2016). A single trial analysis of EEG in recognition memory: tracking the neural correlates of memory strength. *Neuropsychologia* 93, 128–141. doi: 10.1016/j.neuropsychologia.2016.09.026

Ray, A., Tao, J. X., Hawes-Ebersole, S. M., and Ebersole, J. S. (2007). Localizing value of scalp EEG spikes: a simultaneous scalp and intracranial study. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* 118, 69–79. doi: 10.1016/j.clinph.2006.09.010

Rutishauser, U., Mamelak, A. N., and Schuman, E. M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron* 49, 805–813. doi: 10.1016/j.neuron.2006.02.015

Sassenhagen, J., and Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* 56, e13335. doi: 10.1111/psyp.13335

Schoffelen, J. M., and Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Hum. Brain Mapp.* 30, 1857–1865. doi: 10.1002/hbm.20745

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., et al. (2021). Neuropixels 2, 0. A miniaturized high-density probe for stable, long-term brain recordings. *Science* 372, eabf4588. doi: 10.1126/science.abf4588

Tang, A. C., Liu, J. Y., and Sutherland, M. T. (2005). Recovery of correlated neuronal sources from EEG: the good and bad ways of using SOBI. *Neuroimage* 28, 507–519. doi: 10.1016/j.neuroimage.2005.06.062

Tao, J. X., Ray, A., Hawes-Ebersole, S., and Ebersole, J. S. (2005). Intracranial EEG substrates of scalp EEG interictal spikes. *Epilepsia* 46, 669–676. doi: 10.1111/j.1528-1167.2005.11404.x

Van Veen, B. D., van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44, 867–880. doi: 10.1109/10.623056

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., Barrett, L. F., et al. (2015). A bayesian model of category-specific emotional brain responses. *PLOS Comput. Biol.* 11, e1004066. doi: 10.1371/journal.pcbi.1004066

Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S. M., and Carlson, T. A. (2016). Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *Neuroimage* 132, 59–70. doi: 10.1016/j.neuroimage.2016.02.019

Westner, B. U., Dalal, S. S., Gramfort, A., Litvak, V., Mosher, J. C., Oostenveld, R., et al. (2022). A unified view on beamformers for M/EEG source reconstruction. *Neuroimage* 246, 118789. doi: 10.1016/j.neuroimage.2021.118789

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517. doi: 10.1006/jmla.2002.2864

Yu, H., Pistol, C., Franklin, R., and Barborica, A. (2018). Clinical accuracy of customized stereotactic fixtures for stereoelectroencephalography. *World Neurosurg.* 109, 82–88. doi: 10.1016/j.wneu.2017.09.089

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 04 April 2023
DOI 10.3389/fnhum.2023.1128548

OPEN ACCESS

EDITED BY
Harish Chander,
Mississippi State University, United States

REVIEWED BY
Tiphanie Raffegeau,
George Mason University, United States
Jan M. Hondzinski,
Louisiana State University, United States

*CORRESPONDENCE
Ryan M. Peters
✉ ryan.peters1@ucalgary.ca

SPECIALTY SECTION
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

RECEIVED 20 December 2022
ACCEPTED 14 March 2023
PUBLISHED 04 April 2023

CITATION
Hodgson DD, King JA, Darici O, Dalton BH, Cleworth TW, Cluff T and Peters RM (2023) Visual feedback-dependent modulation of arousal, postural control, and muscle stretch reflexes assessed in real and virtual environments.
Front. Hum. Neurosci. 17:1128548.
doi: 10.3389/fnhum.2023.1128548

Visual feedback-dependent modulation of arousal, postural control, and muscle stretch reflexes assessed in real and virtual environments

Daniel D. Hodgson[1], Jordan A. King[2], Osman Darici[1], Brian H. Dalton[3], Taylor W. Cleworth[4], Tyler Cluff[1,5] and Ryan M. Peters[1,2,5]*

[1]Faculty of Kinesiology, University of Calgary, Calgary, AB, Canada, [2]Biomedical Engineering, University of Calgary, Calgary, AB, Canada, [3]School of Health and Exercise Sciences, University of British Columbia Okanagan, Kelowna, BC, Canada, [4]Faculty of Health, York University, Toronto, ON, Canada, [5]Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada

Introduction: The mechanisms regulating neuromuscular control of standing balance can be influenced by visual sensory feedback and arousal. Virtual reality (VR) is a cutting-edge tool for probing the neural control of balance and its dependence on visual feedback, but whether VR induces neuromodulation akin to that seen in real environments (eyes open vs. closed or ground level vs. height platform) remains unclear.

Methods: Here we monitored 20 healthy young adults (mean age 23.3 ± 3.2 years; 10 females) during four conditions of quiet standing. Two real world conditions (eyes open and eyes closed; REO and REC) preceded two eyes-open virtual '*low*' (ground level; VRL) and '*high*' (14 m height platform; VRH) conditions. We measured arousal via electrodermal activity and psychosocial questionnaires rating perceived fear and anxiety. We recorded surface electromyography over the right soleus, medial gastrocnemius, and tibialis anterior, and performed force plate posturography. As a proxy for modulations in neural control, we assessed lower limb reflexive muscle responses evoked by tendon vibration and electrical stimulation.

Results: Physiological and perceptual indicators of fear and anxiety increased in the VRH condition. Background soleus muscle activation was not different across conditions; however, significant increases in muscle activity were observed for medial gastrocnemius and tibialis anterior in VRH relative to REO. The mean power frequency of postural sway also increased in the VRH condition relative to REO. Finally, with a fixed stimulus level across conditions, mechanically evoked reflexes remained constant, while H-reflex amplitudes decreased in strength within virtual reality.

Discussion: Notably, H-reflexes were lower in the VRL condition than REO, suggesting that these ostensibly similar visual environments produce different states of reflexive balance control. In summary, we provide novel evidence that VR can be used to modulate upright postural control, but caution that standing balance in analogous real and virtual environments may involve different neural control states.

KEYWORDS

muscle stretch reflexes, virtual reality, electromyography, electrodermal activity, H-reflexes, tendon vibration, postural sway

1. Introduction

Human standing balance is maintained through a dynamic integration of multiple sensory and motor systems. Alpha-motor neurons (αMNs) are the final common pathway for motor control, with their output being the culmination of complex spinal and supra-spinal integration processes (Burke et al., 1983; Pierrot-Deseilligny and Burke, 2005). Interestingly, the neural control mechanism that support standing balance do not appear to be fixed; rather, they readily adapt to changes in static posture, like standing vs. sitting (Hayashi et al., 1992; Cattagni et al., 2014), as well as different motor behaviors, like standing vs. walking (Capaday and Stein, 1986). The neural control of standing appears to be further modulated in the presence of different task demands and postural threat, such as when there is a chance of being perturbed (Horslen et al., 2013; Lim et al., 2017), when the eyes are closed (Roman-Liu, 2018), or when standing at the edge of a height platform in real (Carpenter et al., 1999, 2004; Sibley et al., 2007; Horslen et al., 2013, 2014, 2017, 2018; Naranjo et al., 2015, 2016) and virtual environments (Cleworth et al., 2012, 2016; Nielsen et al., 2022).

Here we follow the general approach of using muscle stretch reflex excitability to characterize the neural mechanisms supporting standing balance (e.g., Davis et al., 2011; Horslen et al., 2013, 2017, 2018). Our focus is on lower-limb muscle stretch reflexes as a proxy for modulations in the neural control of standing. In response to changes in muscle length, spindle afferents generate reflexive activation through the myotatic stretch reflex arc (Liddell and Sherrington, 1924, 1925; Matthews, 1991). Muscle spindles are also unique in that they have their own dedicated motor innervation ('*fusimotor neurons*' or γMNs), enabling active control over receptor sensitivity by the central nervous system (Hulliger, 1984; Burke, 2021; Dimitriou, 2021). Context-dependent adaptations of the stretch reflex may also result from altered descending commands that modulate the excitability of spinal interneurons and Renshaw cells, and/or changes in muscle spindle sensitivity via γMN activation (Burke et al., 1983; Hulliger, 1984; Dimitriou and Edin, 2010; Burke, 2021; Dimitriou, 2021, 2022). Furthermore, muscle stretch reflexes are traditionally assessed in two different ways, which offer complementary bits of information: mechanical stimulation (imposed joint movements, tendon taps, and muscle/tendon vibration) and electrical nerve stimulation (Hoffman's reflex). The major distinction between mechanical and electrical stimulation is that the former relies on mechanotransduction by the muscle spindle, whereas the latter bypasses these mechanoreceptors and directly activates the peripheral nerve along its length. This is important because changes in muscle spindle sensitivity via γMN input may only be inferred using mechanical stimuli, whereas electrical stimuli give an indirect measure of spinal excitability and presynaptic inhibition (Pierrot-Deseilligny and Burke, 2005; Horslen et al., 2013, 2018).

In real-world environments, muscle stretch reflex excitability appears to be modulated in the presence of increased postural demands, altered visual feedback, and postural threat. For instance, Hoffman (H) reflexes in the soleus were reduced during standing compared to sitting, even when background αMN activity was matched in the seated position (Cattagni et al., 2014). Postural control is challenged when the eyes are closed (Roman-Liu, 2018), however, H-reflexes appear to not be significantly affected by closing the eyes (Sibley et al., 2007). Furthermore, Sibley et al. (2007) demonstrated that H-reflexes are reduced in amplitude when standing on the edge of a real height platform, which was taken as evidence of altered pre-synaptic inhibition or fusimotor drive. By comparing T- and H-reflexes during sympathetic arousal in the same participants, researchers have suggested the presence of enhanced muscle spindle sensitivity under threatening conditions (Davis et al., 2011). Horslen et al. (2013), found that muscle stretch reflexes evoked with tendon taps become sensitized in response to postural threat from a height platform, while the excitability of the spinal circuitry – assessed with H-reflexes – remained largely unaffected. Additional studies have suggested that H-reflexes reduce in amplitude under enhanced postural threat (Llewellyn et al., 1990; McIlroy et al., 2003; Nafati et al., 2004; Sibley et al., 2007), rather than remain unchanged (Horslen et al., 2013). To our knowledge, no studies have investigated such effects within virtual environments and how this compares to simply removing vision by closing the eyes, which is a main objective of the present study. Horlings et al. (2009) induced changes in postural control during quiet standing in both eyes closed and virtual reality (VR) low height conditions compared to eyes open, however, muscle activity and reflex responses were not monitored.

Virtual reality has emerged as an increasingly popular tool for studying human motor control. As a validation of VR, balance control and physiological and/or psychological stress appear to occur similarly in real and VR environments (Meehan et al., 2002; Cleworth et al., 2012, 2016; Nielsen et al., 2022). For example, Cleworth et al. (2012) compared responses to standing on an elevated platform vs. at virtual height in an environment matched in scale and visual detail. The authors reported similar changes

in the amplitude and frequency of centre of pressure (COP) excursions, electrodermal activity (EDA), anxiety, fear, and balance confidence for real and VR height conditions, providing evidence for the potential utility of VR as a tool for manipulating balance control and arousal. The safety and practical challenges (cost and space requirements) of having participants at the edge of a real height platform, building, or cliff is a barrier to studies of this sort. VR may offer a means of circumventing these challenges by simulating tasks and environments that create threat and challenge the balance system. Commercial VR systems (e.g., Oculus Rift, HTC Vive, etc.) have decreased in cost while providing better quality visual feedback in the last 5-10 years. Although VR may offer a fruitful tool to explore how the balance system adapts control for a wide range of visual environments, it is still unclear how exposure to VR modulates the neural control of balance, such as muscle activity and reflex excitability.

The primary objective of this study was to investigate visual feedback driven neural modulations in balance control across real (eyes open/closed; REO and REC) and virtual (low/high; VRL and VRH) conditions. In the real environment, standing with eyes closed served as a proxy to manipulate both arousal and visual feedback. In VR, this was accomplished by modifying platform height. We used mechanical (noisy tendon vibration – NTV) and electrical (soleus H-reflex) stimulation to discern global (spinal excitability) and fusimotor (muscle spindle-dependent) changes in reflex responses under these different visual feedback conditions. A secondary objective was to compare physiological (electrodermal activity) and psychological (fear, anxiety, and confidence questionnaires) indicators of stress when balancing at virtual height. Based on previous findings, it was predicted that: (i) physiological and psychosocial indicators of stress would be greatest in the VRH condition (Cleworth et al., 2012), (ii) similarly to real height platform studies, mechanically evoked stretch reflex amplitudes would increase (Horslen et al., 2013) under conditions of elevated postural threat from closing the eyes, and a virtual height platform, (iii) electrically evoked H-reflex amplitudes would not be altered by closing the eyes (Sibley et al., 2007), and (iv) H-reflex amplitudes will also be reduced (Llewellyn et al., 1990; McIlroy et al., 2003; Nafati et al., 2004; Sibley et al., 2007) or remain unchanged (Horslen et al., 2013), when standing on the edge of a virtual height platform.

2. Materials and methods

2.1. Experimental design

2.1.1. Participants

A total of 20 healthy young adults (mean age: 23.3 ± 3.2 years; 10 females) participated in this study. Participants were free from musculoskeletal injury and any diagnosed neurological conditions. All participants had limited-to-no previous VR experience, and normal to corrected-normal vision with contact lenses, assessed via self-report. Participants requiring eyeglasses were excluded, as the eyeglasses could not fit underneath the VR headset. All participants provided written informed consent prior to taking part in this study. The University of Calgary's Conjoint Health Research Ethics Board approved all experimental protocols.

2.1.2. Procedure

Once written consent was obtained, participants were outfitted with EMG, EDA, and stimulating electrodes, as well as a custom-built wearable tendon vibrator (**Figure 1**). Prior to beginning the testing protocol, a one-time electrical stimulation recruitment curve protocol was performed to determine the maximum peak-to-peak amplitude and corresponding stimulus intensity values for the H-reflex (H-max), M-wave (M-max), and ½ H-max in the soleus muscle. The electrically evoked values vary between individuals; thus, the purpose for ascertaining this information was to determine an appropriate stimulation current for the testing protocol specific to each participant. To accomplish this, participants stood on a force plate with a fixed forward-facing gaze and arms relaxed at their sides while 1 ms square-wave pulses of electrical stimulation were delivered over the tibial nerve within the popliteal fossa. Stimulations were delivered at least 10 seconds apart to minimize risk of interference from presynaptic inhibition (Zehr, 2002; Palmieri et al., 2004). Stimulation current began at 5 mA and was subsequently increased in 2-5 mA increments until the desired physiological response was observed in the soleus. Once the H-wave amplitude began to descend with increasing current, the voltage supply was slowly dialed back during subsequent deliveries until the H-max was determined. To locate the M-max, 2-5 mA current increments were resumed until a plateau was observed in the rising M-wave. The H- and M-max were defined as the greatest peak-to-peak amplitude produced by the smallest amount of current. Finally, to locate ½ H-max, the stimulation current was reduced incrementally from H-max until the H-wave amplitude was equal to half of H-max. The current required to elicit ½ H-max peak-to-peak amplitude was used for the H-reflex assessment throughout the study. The ½ H-max targets a point near the middle of the ascending H-wave and is used to clearly differentiate between an increased/decreased response without ceiling or floor effects (Zehr and Stein, 1999; Sibley et al., 2007; Horslen et al., 2013). This recruitment curve protocol took approximately 5 min to complete, with 15-30 total electrical pulses delivered, and was followed by two minutes of seated rest prior to commencing the testing protocol.

During the subsequent testing protocol, participants experienced five test conditions while standing comfortably on the centre of the force plate with a fixed straight-forward gaze and arms at their sides. Each participant's feet were separated by approximately 20 cm at the medial heel. The placement of the feet was kept consistent using tape markings on the force plate. Participants were instructed to stand relaxed throughout each of the five trials while NTV was used to probe T-reflexes (NTV-reflex) and tibial nerve stimulation was used to probe the H-reflex. During each condition, two minutes of NTV was applied to the right Achilles tendon either prior to or immediately following a series of five H-reflex pulses (separated by 10 s). The stimulus presentation order was interleaved across participants, such that half received NTV first and the other half received H-reflex pulses first, across all trials. In the real environment, participants completed one baseline trial with eyes open (REO), followed by another trial with eyes closed (REC). Participants were then introduced to the virtual environment. An Oculus Rift headset (Meta, CA, USA) displayed the VR environment coded through Vizard Enterprise 64-bit (WorldViz, CA, USA). The VR environment was a concrete

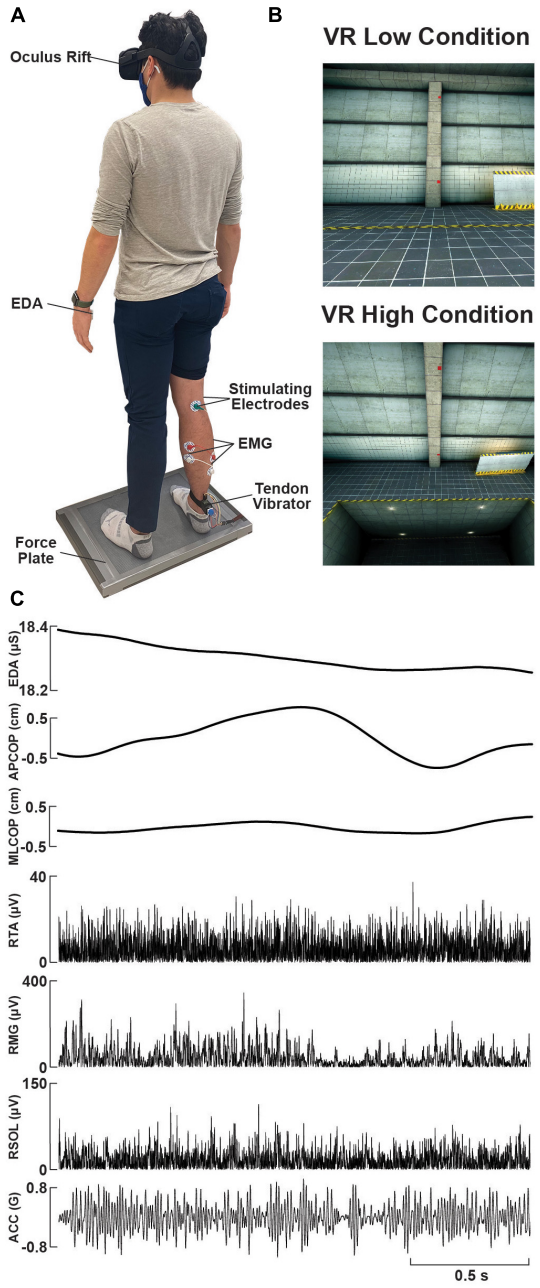


FIGURE 1
Experimental setup and raw data. **(A)** Photo of a participant standing on the force plate, instrumented with a VR headset, lower-limb surface electromyography (EMG) electrodes, tibial nerve electrical stimulation electrodes over the popliteal fossa, a custom-made wearable tendon vibrator over the right Achilles tendon, and electrodermal activity (EDA) electrodes on the left palm. **(B)** Sample visual environment displayed to participants within the 'Low' and 'High' (looking down) VR conditions. **(C)** Raw data traces of EDA, force plate (centre of pressure, COP, in the anterolateral, AP, or mediolateral, ML, planes), rectified EMG (right soleus, RSOL; right medial gastrocnemius, RMG; right tibialis anterior, RTA), and vibrator acceleration (ACC), during an exemplary trial.

warehouse with a small square platform flush with the floor, on which participants stood (**Figure 1**). A red square was positioned straight ahead on a concrete pillar and was used for a reference for participants to focus their gaze during trials. During an initial VR familiarization protocol, while standing on the force plate, participants were asked to: 1) describe their surroundings, 2) find three red cubes placed randomly around the room, and 3) turn their head and read the message on a screen located behind them ("Virtual Reality at the University of Calgary"). The familiarization protocol took 1-2 minutes to complete. Participants then completed one practice trial at ground level to mitigate potential first trial effects (Zaback et al., 2021). Two VR test trials proceeded with either the low (VRL) or high (VRH) condition performed first/second in randomized order. Half the participants received 'VRL1, VRH, VRL2' and the other half received 'VRL1, VRL2, VRH'. We observed no significant effect of VR condition ordering on NTV ($F(1,75) = 0.025$; $p = 0.875$) and H-reflex ($F(1,71) = 0.111$; $p = 0.739$) responses, and therefore, we collapsed the data across the different orders. Prior to the VRH condition, the floor surrounding the platform was lowered seven meters and the platform was raised seven meters (net height change = 14 m) where another red square located at eye height was used for visual reference during the trial. **Figure 1** depicts a visual representation of the scene in VRL and VRH conditions. Prior to each VR trial, participants completed a confidence questionnaire, and after each VR trial, participants completed a 16-item questionnaire addressing their perceived stability, fear, and anxiety during the trial (Adkin et al., 2002). Questions appeared on screen within the VR environment and participants were handed a controller to select their response on a numerical scale ranging from 0 to 100 (confidence, stability, and fear) or 1-10 (anxiety) scale. Each of the five condition trials lasted approximately 3 min including NTV and H-reflex assessment. Questionnaires following VR trials were completed in 3-5 min, and 60-s seated breaks were provided between all conditions to mitigate fatiguing effects.

2.2. Data collection

2.2.1. Electrodermal activity

The galvanic skin conductance (i.e., sweat secretion) in response to each condition was monitored throughout the study by measuring electrodermal activity (EDA), which indicates the degree of sympathetic arousal (Boucsein et al., 2012; Horslen et al., 2013, 2017, 2018). For each condition, EDA during NTV and electrical stimulation was assessed separately rather than pooled together to account for any influence of stimulation method on arousal. Two surface electrodes (1-cm Ag/AgCl; MediTrace 133, Kendall, Technical products, Canada) were placed on the thenar eminence of the left hand (Cleworth et al., 2012). A Skin Conduction Unit (model 2502; CED Limited, England) recorded EDA with a sampling rate of 1 KHz.

2.2.2. Psychosocial assessment

A series of questionnaires were delivered for conditions performed in VR to assess balance confidence, fear, stability, and anxiety (Smith et al., 1990; Adkin et al., 2002; Cleworth et al., 2012; Horslen et al., 2013, 2014, 2017, 2018). Individual questions appeared in the VR field of view, and participants were handed a controller to navigate to and select their appropriate response. Prior to each VR trial, participants rated their confidence in their ability to balance throughout the upcoming trial on a scale from

0% (no confidence) to 100% (complete confidence). Post-trial, participants rated their stability and fear on a scale of 0% (no fear; no instability) to 100% (completely fearful; complete stability), and a 16-item questionnaire assessed perceived anxiety using a nine-point scale ranging from 1 (I did not feel this at all) to 9 (I feel this extremely). These questionnaires have been demonstrated to have moderate to high reliability under height-induced postural threat in young adults (Hauck, 2011) and have been used in VR conditions (Nielsen et al., 2022). Scores from the anxiety questions were concatenated to achieve an overall measurement of anxiety.

2.2.3. Electromyography

EMG was monitored to assess muscle activity throughout the study. Changes in muscle activity may be indirectly associated with changes in posture, fatigue, or arousal. Furthermore, changes in background EMG amplitude may influence other variables such as reflex amplitude and centre of pressure dependent variables. Prior to surface electrode placement, the skin overlaying the muscles of interest was shaved and cleaned with alcohol swabs. Two surface electrodes (1-cm Ag/AgCl; MediTrace 133, Kendall, Technical products, Canada) were applied to the skin, 2 cm centre-to-centre spacing, over the right soleus (SOL), medial gastrocnemius (MG), and tibialis anterior (TA). Electrodes were placed in line with muscle fiber orientation over the thickest region of the muscle belly (for MG and TA) or just distal and lateral to the gastrocnemius-soleus intersection for SOL. A reference electrode was placed over the lateral malleolus. EMG was sampled at 5 kHz and amplified with a gain of 2000 (NeuroLog NL844 pre-amplifier and NL820A isolated amplifier; Digitimer Ltd., England).

2.2.4. Force plate

During testing, participants stood in the centre of a six-axis force plate (OR6-7; AMTI, USA). Ground reaction forces and moments, sampled at 1 kHz, were collected to monitor the center of pressure (COP) in the anterior-posterior and medial-lateral directions during all trials. See **Figure 1** for a depiction of the experimental set up.

2.2.5. Noisy tendon vibration

Mechanical stimulation of the Achilles tendon produced spindle-generated reflexes in the triceps surae muscles. A novel wearable device, secured to the posterior ankle using an elastic strap (2 cm above the calcaneus, directly between the medial and lateral malleolus) was used to deliver NTV (0 – 100 Hz) to the Achilles tendon. The elastic strap was secured in position and held at a snug, but comfortable tension prior to performing any study procedures and remained in place for the entire experiment. The device consisted of a custom-manufactured 3D printed housing (**Figure 1**) with onboard motor (Haptuator BM3C, Tactile Lab, Canada), amplifier (PAM8403), and accelerometer (ADXL354-CZ, Analog Devices Ltd., USA). Motor command signals were generated using custom LabVIEW software (National Instruments, USA) and output at 10 kHz from a real-time PXI system (PXIe-1062Q chassis; PXI-8105 embedded controller) with a multifunction data acquisition card (PXI-6289) and A/D board (BNC-2090). For the input signal to all subsequent NTV analyses, we computed acceleration magnitude as the square root of the sum of squared x, y, and z axis accelerations (in units of G). Using vibration magnitude is favorable in this case, as the wearable device vibrates in three-dimensions and we sought to fully capture this motion – this is unlike previous studies that used bulkier motors which moved only in one axis (e.g., Mildren et al., 2017).

2.2.6. Electrical nerve stimulation

Electrical stimulation of the tibial nerve was required to generate H-reflexes in the triceps surae muscles. These reflexes are initiated along the peripheral nerve and therefore bypass the muscle spindle. Following previous studies, H-reflexes were used in conjunction with NTV-reflexes to assess site-specific adaptations (Davis et al., 2011, Horslen et al., 2013). Surface electrodes (1-cm Ag/AgCl; MediTrace 133, Kendall, Technical products Toronto, Ontario, Canada) were used for the anode and cathode. The anode was positioned in the centre of the popliteal crease in the popliteal fossa (Zehr, 2002), targeting the tibial nerve as it travels proximally and distally. The cathode was positioned 2 cm superior to the patella on the anterior thigh. A constant-current stimulator (STMISOLA; Biopac, USA) was used to deliver square-waveform electrical pulses (1 ms duration) using LabVIEW software (National Instruments, USA). The amperage (1-100 mA) of each stimuli was manually manipulated to achieve the stimulator output producing the desired physiological response.

2.3. Data processing

Mean EDA, expressed in microSiemens (μS), was extracted during time windows encompassing the NTV and electrical stimulation periods. Force plate signals were digitally low pass filtered (dual-pass) at 10 Hz (4th order low-pass Butterworth filter) and COP was calculated from moments (Mx and My) and vertical force (Fz). The root mean squared (RMS) displacement and mean power frequency (MPF) of COP in the anteroposterior (AP) and mediolateral (ML) directions were calculated during each trial. EMG data were DC removed and full wave rectified for NTV analysis. Signals were digitally filtered at 1,000 Hz (dual-pass, 4th order, low-pass Butterworth filter). Background muscle activity was determined for the SOL, MG, and TA during H-reflex and NTV assessment by extracting the RMS amplitude during time windows encompassing the stimulation period plus one second on either end. For H-reflex assessment in the SOL, EMG signals were trigger-averaged to the stimulus delivery onset within a window of 10 ms preceding to 100 ms following tibial nerve stimulation. This resulted in a trace representing the mean response from the five H-waves evoked during each condition trial. Values were extracted for stimulus-triggered average peak-to-peak amplitude and latency from the stimulus to the first peak in the response wave (**Figure 2**). For NTV assessment in the SOL, coherence analysis was performed using the NeuroSpec2.0 software package (Rosenberg et al., 1989; Halliday et al., 1995) for MATLAB (Mathworks). This approach was adopted from prior studies investigating the time and frequency characteristics of stochastic stimuli-evoked physiological responses (Dakin et al., 2007, 2010, 2011; Mildren et al., 2017). Coherence values represent normative estimates of the frequency coupling strength between two signals (Mildren et al., 2017). Thus, coherence functions were used to assess the strength of the linear relationship between the input (NTV magnitude) and the output

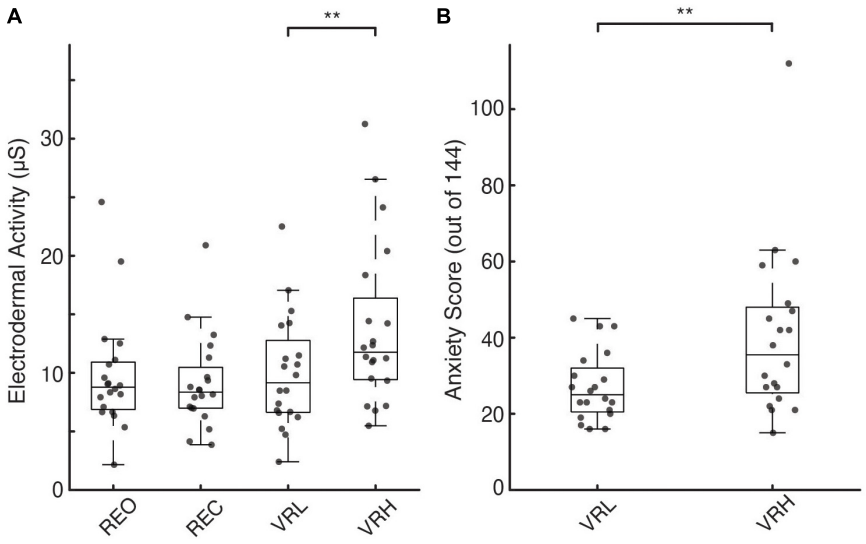


FIGURE 2
Physiological and perceptual indicators of postural threat-induced fear and anxiety. **(A,B)** Individual participant (gray circles) and box plots showing median and interquartile range for EDA under real **(A)** and virtual environments. **(B)** Individual and average participant questionnaire anxiety rating scores (out of a total of 0-144 points) under the different VR conditions. Asterisks denote significance at the $p < 0.01$** level.

(SOL EMG) signals in the frequency domain. Coherence was estimated as the magnitude of the input-output signal cross spectra squared divided by the product of the input and output autospectra (Dakin et al., 2007; Mildren et al., 2017). Temporal characteristics of coherent frequencies were detected by using the inverse Fourier transform of the input-output signal cross spectra to compute cross-covariance (cumulant density), which was normalized by the product of the vector norms of the input and output signals. Resulting cross covariance values were normalized to a range of −1 to 1 (Dakin et al., 2010) and approximate the signal coupling strength in the time domain. Therefore, a positive correlation could be produced by either increased probe magnitude being coupled with an increase in EMG, or by decreased probe magnitude being coupled with a decrease in EMG. Independence of the two signals (input and output) are assumed under the constructed 95% confidence limits for coherence (positive threshold) and cross covariance (positive and negative thresholds) (Halliday et al., 1995). Values exceeding these limits are indicative of a significant linear relationship between the input (stimulus) and output (response) signals. Coherence spectra and cumulant densities were computed for each individual participant, as well as by pooling the data from each condition across all 20 participants (FFT windows = 1s; frequency resolution 1 Hz). Peak-to-peak and the timing of the cumulant density peak was extracted for statistical analysis. Additionally, we conducted pairwise Difference of Coherence tests (Amjad et al., 1997) to determine significant differences in pooled coherence spectra (see **Figure 3**).

2.4. Statistical analysis

Analysis was performed using custom MATLAB (MathWorks, MA, USA) and SPSS software (version 26.0; IBM SPSS Statistics Inc., Chicago, Il, USA) with α set to 5%. General linear mixed effects models were used to determine the main effect of visual condition on dependent variables. This statistical model provided comparisons between all conditions in real (REO and REC) and virtual (VRL and VRH) environments with pairwise estimates using sequential Bonferroni adjustments for multiple comparisons. Dependent variables in both real and VR environments included mean EDA (assessed independently during NTV and electrical stimulation), mean rectified background root mean squared (RMS) EMG in each muscle (SOL, MG, and TA), RMS and mean power frequency (MPF) of the anterior-posterior (AP) and medial-lateral (ML) COP, H-reflex peak-to-peak amplitude in the SOL, and NTV-reflex cumulant density peak-to-peak amplitude in the SOL. Pairwise difference of coherence tests compared coupling strength as a function of frequency for the identical planned comparisons. Subjective balance confidence, fear, stability, and anxiety data from questionnaires were assessed for VRL versus VRH using (one-tailed) paired *t*-tests. Mean EDA during each type of stimulus delivery (electrical vs. NTV) was compared for each condition using paired *t*-tests to determine if participant arousal was consistent during H-reflex and NTV-reflex assessment. Cohen's *d* effect sizes were calculated, and magnitude of the effects may be interpreted as small (ES = 0.2), medium (ES = 0.5), or large (ES = 0.8) (Cohen, 1988).

3. Results

3.1. General results

Force plate analysis was conducted using data from 19 subjects; due to technical issues, these data were unavailable for one participant. Data from one participant was also excluded from the final H-Reflex analysis due to equipment malfunction. Of the remaining 19 participants, all had strong SOL muscle responses to H-reflex, while a weaker or absent response was observed in

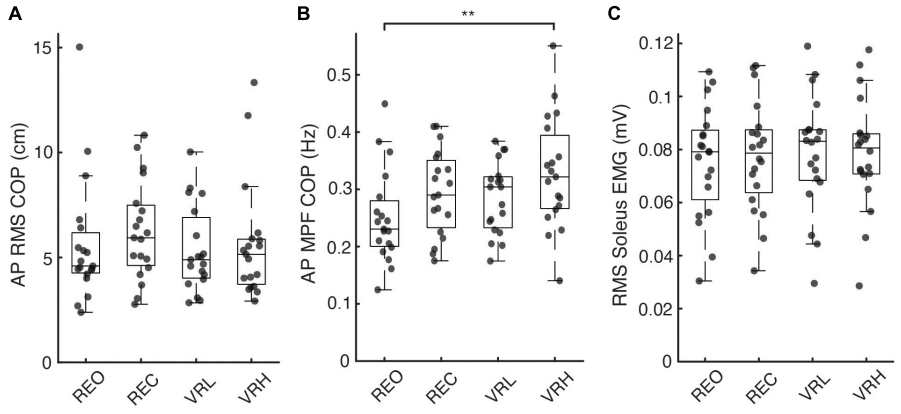


FIGURE 3
Postural sway and soleus muscle activity across real and virtual conditions. **(A)** Root-mean-squared (RMS) amplitude and **(B)** mean power frequency of anteroposterior (AP) centre of pressure (COP) excursions, as well as **(C)** RMS soleus muscle activity, across each visual condition. Asterisks denote significance at the $p < 0.01$** level.

MG (18 responders) and TA (15 responders) muscles. Similarly, all 20 participants showed strong SOL and MG reflex responses (coherence) during NTV, while none definitively produced reflex responses in the TA. Since SOL consistently demonstrated the strongest response and was present in all individuals during H-reflex and NTV delivery, we focused the analysis exclusively on SOL. This approach is aligned with prior H-Reflex (Horslen et al., 2013; Young et al., 2018) and mechanical tendon stimulation (Horslen et al., 2013, 2018; Mildren et al., 2016, 2017) studies which focused on SOL for analysis.

3.2. Electrodermal activity

There were significant main effects for condition on EDA during both H-reflex assessment ($F(3, 76) = 6.769$; $p < 0.001$) and NTV ($F(3, 76) = 7.003$; $p < 0.001$). Pairwise contrasts, adjusted for multiple comparisons using the sequential Bonferroni correction, determined differences between specific conditions. As shown in **Figure 2A**, during NTV, significantly greater EDA was observed in the VRH condition compared to REO (41.6%; $t(76) = 3.557$; $p = 0.003$; CI: 1.045 – 7.08), REC (51.9%; $t(76) = 4.126$; $p = 0.001$; CI: 1.63 – 7.818), and VRL (38.4%; $t(76) = 3.358$; $p = 0.005$; CI: 0.912 – 6.756). There were no significant differences in EDA across stimulation method (electrical vs. NTV) for any condition. During electrical stimulation (H-reflex assessment), there was a similar increase in sympathetic arousal in the VRH condition compared to REO (39.6%; t(76) = 3.741; $p = 0.002$; CI: 1.324 – 7.692), REC (40.9%; t(76) = 3.826; $p = 0.002$; CI: 1.346 – 7.875), and VRL (35.1%; t(76) = 3.422; $p = 0.004$; CI: 1.041 – 7.207).

3.3. Psychosocial questionnaires

Questionnaires were administered prior to and following VR conditions only. We observed significant mean differences between the high and low conditions across all psychosocial measures. Compared to the VRL condition, participants reported significantly lower confidence in their ability to maintain balance prior to VRH (14.3%; $t(19) = 2.63$, $p = 0.008$, ES = 0.588). Similarly, during the VRH versus the VRL condition, participants felt more unstable (11.5%; $t(19) = 2.87$, $p = 0.005$, ES = 0.642) and more fearful of falling (17.3%; $t(19) = -2.89$, $p = 0.005$, ES = 0.646), while experiencing greater levels of anxiety (10.6%; $t(319) = -7.3$, $p < 0.001$, ES = 0.408; **Figure 2B**).

3.4. Background muscle activity

A general pattern of increased muscle activity emerged within the VR environment, most notably in VRH, in each muscle measured. *Soleus:* In SOL, there was a significant main effect for condition on background muscle activity ($F(3, 76) = 2.832$; $p = 0.044$). Pairwise contrasts, adjusted for multiple comparisons using the sequential Bonferroni correction, did not identify any significant difference between any conditions; however, non-significant trends for increased EMG were observed in the VR environment, with both VRL (3.8%; $t(76) = 2.49$; $p = 0.075$; CI: 0.000 – 0.004) and VRH (3.9%; $t(76) = 2.557$; $p = 0.075$; CI: 0.000 – 0.004) approaching significance compared to baseline (REO) (**Figure 3**). *Medial Gastrocnemius:* In MG, there was a significant main effect for condition on background muscle activity ($F(3, 76) = 2.962$; $p = 0.037$). Pairwise contrasts demonstrated an increase in EMG during VRH compared to baseline (REO) (10.4%; $t(76) = 2.826$; $p = 0.036$; CI: 0.000 – 0.007). *Tibialis Anterior:* Similarly, there was a significant main effect for condition on background muscle activity in TA (F(3, 76) = 3.151; p = 0.030). Pairwise contrasts demonstrated an increase in EMG during VRH compared to baseline REO (6.5%; $t(76) = 2.953$; $p = 0.025$; CI: 0.000 – 0.004).

3.5. Centre of pressure

Standing at virtual height impacted postural sway characteristics in the AP direction, indicated by a significant

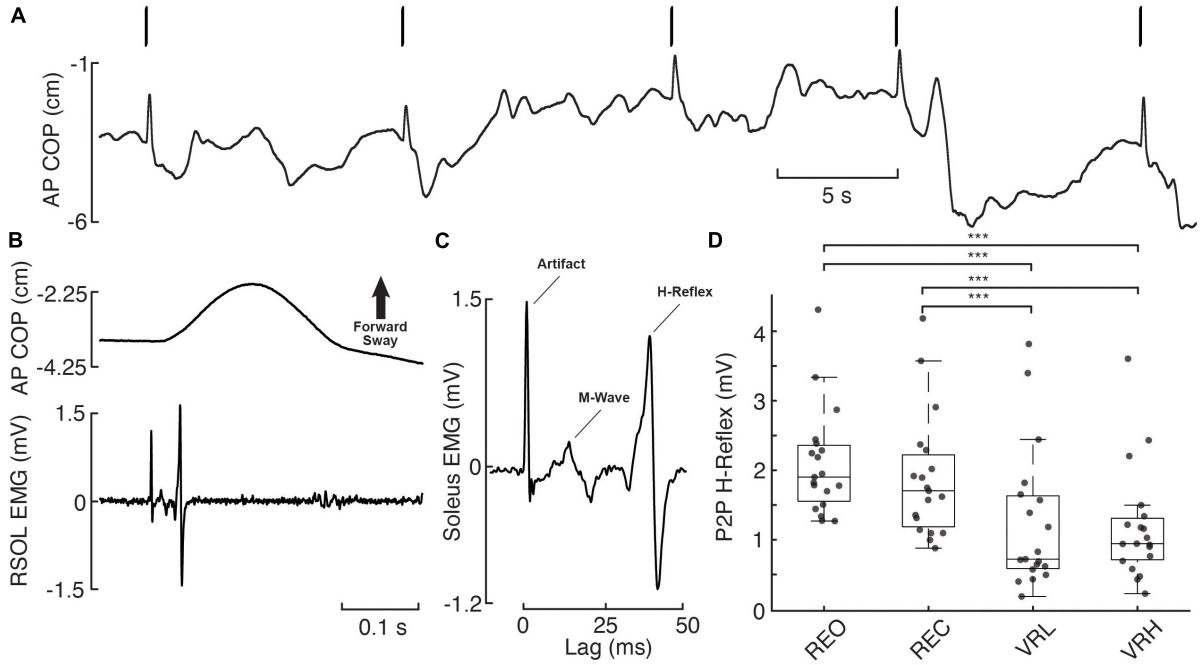


FIGURE 4
H-reflex responses across real and virtual visual conditions. **(A)** Anteroposterior (AP) centre of pressure (COP) during 5 consecutive electrical stimulation pulses (bars overtop), delivered to the tibial nerve at the popliteal fossa (S29, REO). A notable, forward-directed transient in AP COP postural sway resulting from rapid plantar flexion is observed in response to each electrical pulse. **(B)** Exemplary postural (AP COP) and muscular (RSOL EMG) response to electrical stimulation from a single pulse (initial pulse in panel **A**). **(C)** Stimulus-triggered average of the 5 consecutive pulses from this trial, with the RSOL stimulus artifact, M-Wave, and H-Reflex labeled. **(D)** Individual participant (gray dashed lines/diamonds) and average (black bars) peak-to-peak (P2P) H-reflex amplitudes. Asterisks denote significance at the $p < 0.001$*** level.

main effect for condition on AP COP MPF ($F(3, 72) = 4.605$; $p = 0.005$). Increases in AP COP MPF versus baseline (REO) were significant during the VRH condition only (31.1%; $t(72) = 3.693$, $p = 0.003$, CI: 0.021 – 0.135). There were no differences in COP RMS in the AP or ML direction, or in ML COP MPF, between any conditions. **Figure 3** illustrates COP responses across conditions.

3.6. H-reflexes

There was a significant main effect for condition on H-reflex P2P amplitude ($F(72) = 13.829$; $p < 0.001$). Significant pairwise contrasts demonstrate patterns of reduced H-reflex P2P amplitude within the virtual environment versus during real world conditions. Indeed, both VRL (40.8%; $t(72) = -4.934$, $p < 0.001$; CI: −1.315 – −0.397) and VRH (43.4%; $t(72) = -5.247$, $p < 0.001$; CI: −1.381 – −0.44) evoked smaller H-reflexes compared to baseline (REO). Similarly, H-reflex P2P amplitude was also reduced compared to REC during both VRL (34%; $t(72) = -3.676$, $p = 0.001$; CI: −1.063 – −0.212) and VRH (36.9%; $t(72) = -3.988$, $p = 0.001$; CI: −1.136 – −0.247). There was no difference in H-reflex peak-to-peak amplitude between VR height conditions ($t(72) = -0.312$, $p = 0.756$; CI: −0.4 – 0.292). A mean latency of 36 ms was observed, defined as the time between the stimulus delivery and the initial peak of the bi-phasic reflex wave (**Figure 4**).

3.7. Noisy tendon vibration reflexes

A mean latency of 44.3 ms was observed, defined as the time between the stimulus delivery and the initial peak of the bi-phasic reflex wave. However, there were no significant condition effects on NTV-reflex P2P amplitudes. However, pairwise difference of coherence tests found that coherence between NTV magnitude and SOL muscle activity increased for a subset of frequency components (see **Figure 5**) for VRH vs. VRL, VRL vs. REO, and VRH vs. REO comparisons.

4. Discussion

4.1. Summary

The primary objective of this study was to investigate adaptations in arousal, postural control, and muscle stretch reflexes under different conditions of real and virtual visual feedback. We observed that mechanically evoked muscle stretch reflex amplitude remained unchanged across visual conditions, however, there were significant reductions in H-reflex amplitude in virtual reality. This phenomenon occurred not only when exposed to virtual height-induced threat, but also while standing at virtual ground level (VRL), a condition mimicking the REO baseline condition. Thus, merely altering visual input from a real setting to a virtual environment appears to alter the excitability of

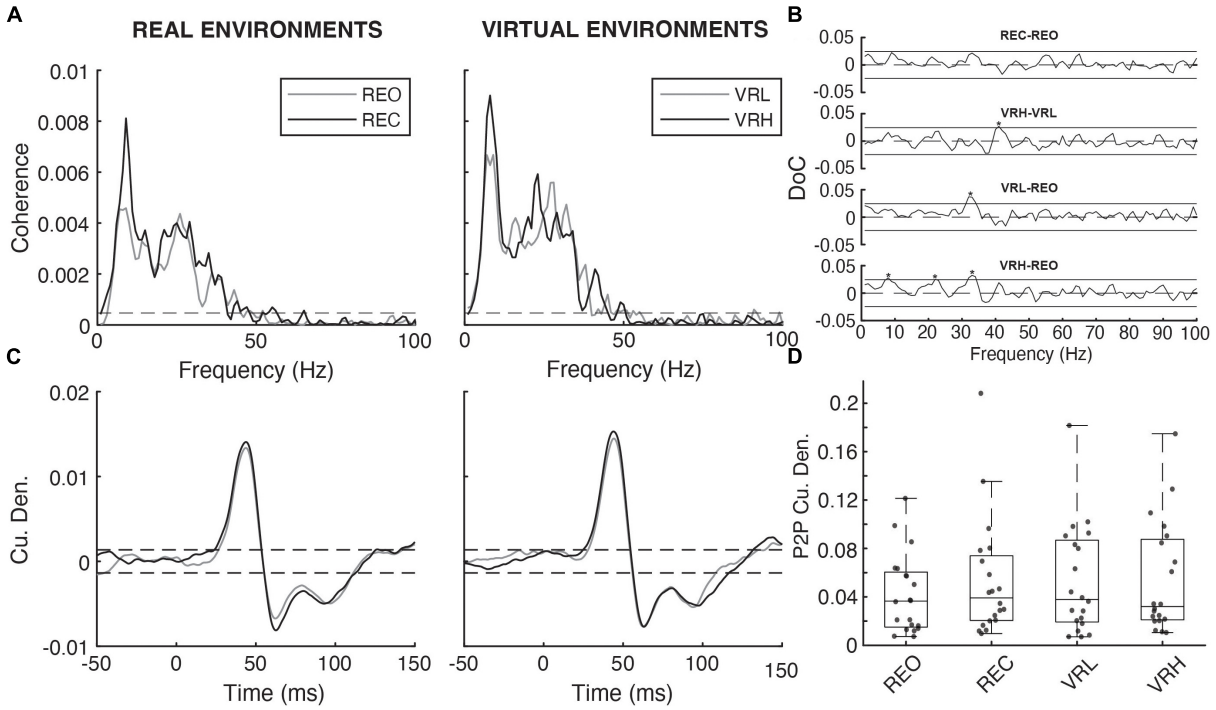


FIGURE 5
Noisy tendon vibration evoked responses across real and virtual conditions. **(A)** Pooled coherence spectral estimates (N = 20) under real (REO vs. REC) and virtual (VRL vs. VRH) visual environments. **(B)** Pairwise difference of coherence (DoC) tests comparing the pooled coherence spectra under real (REC-REO) and virtual (VRH-VRL) conditions. Comparisons between baseline (REO) and virtual (VRL-REO, VHR-REO) visual conditions are also shown. Solid horizontal black lines denote the 95% CI for each pairwise DoC test and values exceeding these lines (asterisks) indicate that, in all cases, NTV-muscular coherence was higher under conditions of greater postural threat. **(C)** Pooled cumulant density estimates for the relationship between vibration magnitude and rectified soleus activity under real and virtual visual environments. **(D)** Individual participant (gray dashed lines/diamonds) and average (black bars) peak-to-peak (P2P) cumulant density amplitudes.

spinal reflex pathways. This lends support to existing literature suggesting that a specific adaptation may occur within the muscle stretch reflex circuity, such that, in VR, there is a global decrease in neuro-excitability at the level of the spine. Our results also suggest there may be a concomitant increase in muscle spindle sensitivity, which offsets this decrease in spinal excitability, keeping mechanically evoked muscle stretch reflexes unchanged (or even slightly increased) in amplitude. To appropriately address our primary research objective, this study also aimed to elicit physiological and psychological responses to postural threat using VR. A physiological arousal response to postural threat was observed, with significant EDA increases occurring with transition to VRH from baseline (REO), REC, and VRL conditions. Similarly, subjective anxiety and fear of falling increased, while balance confidence decreased in the VRH condition compared to VRL. Postural muscles of the lower limb also demonstrated increased co-contraction, which reached significance when transitioning to the VR height environment, likely explaining the increase in the mean power frequency of postural sway observed during this condition.

4.2. Psychosocial and physiological modulation

Significant changes in physiological and psychological responses in the virtual height condition compared to ground level are indicative of increased sympathetic arousal. Skin conductance (EDA) in the VRH condition was elevated compared to VRL, and concomitant reductions in perceived stability and balance confidence, as well as increased fear of falling were also documented. These findings support prior indications that VR height simulations can stimulate sympathetic arousal such as increased skin conductance, heart rate, and perceived anxiety and fear (Meehan et al., 2002; Cleworth et al., 2012) to a similar degree to real-world elevation. Additionally, no differences were observed in EDA between the real (REO) and virtual (VRL) baseline conditions. This indication that transitioning into VR does not increase participant arousal has been previously reported (Simeonov et al., 2008; Cleworth et al., 2012). The invariability in arousal between REO and VRL suggests that participants were not exposed to physiological stress during this transition, alluding that any associated changes in reflex excitability may be attributed to visual feedback rather than perceived threat. Overall, this study was successful in using VR to generate measurable physiological and psychosocial changes in sympathetic arousal.

4.3. Altered background muscle activity and postural sway

Here we observed a general pattern of increased muscle activity with the addition of virtual reality induced threat. Patterns

of increased muscle activity in VR were significant during the height condition (VRH) compared to baseline (REO) for medial gastrocnemius and tibialis anterior. Increases in activity for these muscles at virtual height is consistent with the established increase in lower-limb co-contraction observed in real-world height induced postural threat. Simultaneously, postural sway frequency was also affected by VR; compared to baseline (REO), AP MPF increased in the VRH condition, indicative of a tensing response at the ankle joint. This likely represents a neural adaptation strategy to reduce instability (maintain balance) under conditions where the risk or consequence of a fall is heightened. A linear increase in COP frequency with height-induced threat has been previously reported (Adkin et al., 2000). In real height environments, increased muscular co-contraction alongside greater frequency of AP sway is observed.

Previous studies, which directly compared standing at ground level and height in both real and VR environments did not report significant differences in postural control between matched real and virtual conditions when standing on a firm surface (Simeonov et al., 2005; Cleworth et al., 2012), although Simeonov et al. (2005) did observe increased postural sway when standing on an unstable surface in VR versus a real environment for both ground level and height conditions. However, VR has been reported to increase postural instability during quiet stance on both firm and foam surfaces compared to similar real-world conditions; in fact, Horlings et al. (2009) observed that when compared to real-world quiet standing, the effects of VR on postural sway were similar to an eyes closed condition, consistent with the current study, which found no change in postural control between REC and VRL. It is possible that differences in technology or virtual visual environment account for the contradicting findings regarding postural control in VR. For example, the former studies created VR scenes that very closely matched their real-world laboratory (Simeonov et al., 2005; Cleworth et al., 2012), while Horlings et al. (2009) and the current study did not match the VR scene to the real-world eyes open condition. Participant VR experience may have also played a role in this discrepancy. While other studies do not report on this, participants entered the current study with very limited prior exposure to VR. It is possible that the novel nature of the VR experience could affect the results.

Other studies also did not report on changes in muscle activity in VR; however, the changes in postural control in this study are supported by the associated increase in muscular co-contraction described above. A review by Adkin and Carpenter (2018) identified eight studies which investigated height-induced postural and neurophysiological outcomes. A pattern emerged in the literature that is supported by the present research – each study reported evidence of increased COP frequency (Adkin and Carpenter, 2018). These previous studies also reported a reduction (6 studies) or no change (2 studies) in sway amplitude when threatened with height (Adkin and Carpenter, 2018). Recent research has substantiated these findings of altered balance control in response to standing in virtual elevation. Raffegeau et al. (2020) recorded postural deviations using inertial sensors while subjects were elevated in VR from ground level to 15 m height. The authors reported that AP postural responses (sway acceleration) were significantly greater at height, but only when standing parallel to the threat (i.e., on the edge of the platform compared to standing perpendicular to the edge).

4.4. Modulation in muscle stretch reflex responses

We observed that vibration evoked stretch reflexes remained unchanged, or marginally increased with virtual reality induced postural threat. Given that H-reflexes, and presumably, spinal excitability was diminished in virtual reality, our NTV results suggest that a compensatory neural mechanism is at play to maintain nearly constant mechanically evoked stretch responses. Indeed, in conditions where H-reflexes were reduced, NTV reflexes remained unchanged or slightly increased on average. This is consistent with previous reports of threat-induced modulations to postural control. Indeed, T-reflex amplification has been demonstrated to occur during static (Davis et al., 2011; Horslen et al., 2013) and dynamic reactive (Horslen et al., 2018) standing at height, as well as alternative sources of arousing stimuli (Bonnet et al., 1995; Both et al., 2003, 2005; Hjortskov et al., 2005; Kamibayashi et al., 2009).

As mentioned, an opposing trend to vibration reflex responses emerged for electrically evoked muscle stretch reflexes. H-reflex amplitudes were attenuated in VR conditions (VRL and VRH). No difference was observed in H-reflex amplitude between virtual height conditions in the present study. Interestingly, compared to REO, H-reflex amplitudes were reduced in the VRL condition, suggesting a more general effect of changing visual input from real to virtual, despite these conditions being seemingly similar. This demonstrates that the neural control of standing is altered in VR, even in the absence of any virtual height manipulation, which to our knowledge has not been reported previously (i.e., standing on ground level in the real- and virtual-world results are not equivalent). Reduced H-reflex amplitudes in response to threatening scenarios is consistent with previous reports. Subjects have exhibited weaker H-reflex amplitudes in response to an increased risk (Llewellyn et al., 1990; Horslen et al., 2013; Miranda et al., 2019) or consequence (Sibley et al., 2007; Horslen et al., 2013) of a fall, as well as during alternative scenarios stimulating sympathetic arousal (Kamibayashi et al., 2009; Tanaka, 2015). This suggests an inhibitory response at the spinal level that may be facilitated by descending commands from cortical or supraspinal structures (Llewellyn et al., 1990; Adkin et al., 2008). Tanaka et al. (2013) support this theory by using single-pulse transcortical stimulation to demonstrate increased corticospinal excitability to the internal oblique muscle during postural-challenging tasks at height. It is feasible that, when encountered with added sympathetic arousal or altered visual input, in general, the central nervous system downregulates spinal excitability to stifle excess motor outputs which may interfere with its ability to produce appropriate and efficient sensory-evoked responses to the stimuli. The net effect of this neural adaptation would be that mechanical reflex excitability remains unchanged, as observed in the present study.

4.5. Neural mechanisms for reflex modulation

We are under the assumption that both vibration and H-reflexes are short-latency responses using the same spinal circuitry except for initiation method and location. The H-reflex is initiated by stimulating the Ia afferent fibers directly (bypassing activation of muscle spindles), and therefore, its pathway is incomplete compared to its mechanically induced counterpart. Vibration reflex initiation occurs via muscle spindle stimulation resulting from a mechanical stretch applied to the muscle or tendon, encapsulating the entirety of the Ia afferent pathway. It is postulated, based on results of this study, that specific adaptations local to the muscle spindle may occur in the presence of virtual reality induced postural threat. This inference is consistent with prior research (Bonnet et al., 1995; Both et al., 2003, 2005; Hjortskov et al., 2005; Kamibayashi et al., 2009; Davis et al., 2011; Horslen et al., 2013, 2018) suggesting a heightened spindle-specific role in postural control in arousing situations. Furthermore, because significant H-reflex reduction signifies heightened inhibition at the spinal or supra-spinal level, the increase in spindle sensitivity alongside postural threat must be sufficient to overcome and exceed these inhibitions to register a net-positive increase in overall reflex. Therefore, it is likely that the muscle spindle response to threat is underestimated despite the significant enhancements observed here and in previous reports. The role of fusimotor control in posture and movement, and its specific adaptations when threats are perceived, demands further investigation by directly measuring spindle activity via single unit recordings with human microneurography.

Golgi Tendon Organ (GTO) inhibition may also play a role in NTV-reflex facilitation in threatening scenarios. Horslen et al. (2017) used electrical stimulation of the Achilles tendon to demonstrate Ib-inhibition in the triceps surae muscles during upright standing at both low and high height conditions compared to lying prone at ground level. Thus, it seems GTO-Ib reflexes are also dependent on task and threat perception, and their inhibition under these conditions may enable amplified short latency Ia-muscle spindle responses. Similar reductions in Ib inhibition have been previously demonstrated (i.e., Faist et al., 2006; Van Doornik et al., 2011) and are postulated to reflect a shift toward excitatory, rather than inhibitory reflexes during postural tasks compared to unloaded positions; thus, it seems reasonable to suggest this effect may be amplified as basic postural tasks (i.e., standing) advance to more threatening conditions (i.e., standing at height) with greater levels of muscle activation. However, axons of Ib afferents innervating GTOs are known to be smaller in diameter than Ia afferents supplying muscle spindles. Thus, their stimulation threshold requires a greater current – likely much higher than that delivered during this study, which used stimulation intensities in lower ranges to target muscle spindles specifically (Pierrot-Deseilligny and Burke, 2005). The concurrent roles of Ib-GTO inhibition and Ia-muscle spindle excitation in response to sympathetic arousal must be assessed further to confidently infer their contributions to postural tasks in threatening states.

4.6. Limitations and future considerations

One interpretation of our findings is that postural threat in VR increased fusimotor outflow, which maintained vibration reflex strength during a concomitant increase of spinal inhibition. A limitation of this research, however, is that we are inferring potential changes in fusimotor outflow and muscle spindle feedback indirectly using reflexive muscle activity as a proxy. As such, we cannot presently make firm conclusions regarding muscle spindle responses or potential mechanisms of fusimotor control. In future studies, these results should be bolstered by direct recordings from individual muscle spindles with human microneurography in threatening conditions. Even so, direct recordings from single fusimotor neurons are not feasible in humans; therefore, γMN activity will still need to be inferred from changes in muscle spindle firing, all while keeping αMN (muscle background activity) constant across conditions. This challenge of inferring fusimotor drive from muscle spindle activation is a well-appreciated in the literature (Burke, 2021; Dimitriou, 2021).

Another limitation to this study is the lack of a real-world high height condition. VR has previously been shown to be capable to eliciting similar responses to real-world conditions involving quiet standing at low and high heights (Cleworth et al., 2012). The current study aimed to build on this work by probing the state of neural circuitry under similar conditions. It would have been beneficial to directly compare these outcomes in real and virtual height conditions as well as in the real and virtual ground level conditions explored in this study. However, due to practical limitations, a real height condition was replaced with an eyes closed condition in this study. Nonetheless, closing the eyes enabled an interesting comparison to virtual environments, and has been used previously in studies involving real-world and VR visual environments (Horlings et al., 2009). Finally, comparisons between REO and VR conditions in current study would have been better controlled had the virtual environment been designed to match the real-world visual scene of the lab. Previous studies comparing real and virtual conditions have implemented near identical visual scenes within their virtual display (Simeonov et al., 2005; Cleworth et al., 2012), and this discrepancy may have played a role in some conflicting findings between studies.

Our data reinforces that VR can elicit sympathetic arousal and subjective fear responses with manipulations in virtual elevation. Future research should investigate how these outcomes scale in different, potentially even more threatening, virtual environments. Beyond basic neural mechanisms of balance control, our findings further suggest that VR may be useful for therapeutic purposes. For example, VR exposure therapy has been presented as a promising concept for managing phobias and anxieties, such as towards heights or flying, by gradually exposing patients to a negative stimulus (Rothbaum et al., 1997; Krijn et al., 2004; Bush, 2008; Albakri et al., 2022). There is already research in this area suggesting benefits in terms of fear of falling and fall risk in older adults (Rendon et al., 2012; Levy et al., 2016), and clinically in people with acrophobia (Coelho et al., 2009; Rimer et al., 2021), or anxiety disorders (Meyerbröker and Emmelkamp, 2011).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by University of Calgary's Conjoint Health Research Ethics Board (CHREB). The patients/participants provided their written informed consent to participate in this study.

Author contributions

DH wrote the first draft of the manuscript and performed the statistical analysis. JK and RP created the figures. DH, JK, and OD performed the data collection. All authors contributed to the study design, interpretation of the results, and editing of the final manuscript.

Funding

This research was funded by a Natural Sciences and Engineering Discovery Grant (DGECR-2020-00124) awarded to RP.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adkin, A., Campbell, A., Chua, R., and Carpenter, M. (2008). The influence of postural threat on the cortical response to unpredictable and predictable postural perturbations. *Neurosci. Lett.* 435, 120–125. doi: 10.1016/j.neulet.2008.02.018

Adkin, A., and Carpenter, M. (2018). New insights on emotional contributions to human postural control. *Front. Neurol.* 9:789. doi: 10.3389/fneur.2018.00789

Adkin, A., Frank, J., Carpenter, M., and Peysar, G. (2002). Fear of falling modifies anticipatory postural control. *Exp. Brain Res.* 143, 160–170. doi: 10.1007/s00221-001-0974-8.

Adkin, A., Frank, J., Carpenter, M., and Peysar, G. (2000). Postural control is scaled to level of postural threat. *Gait Posture* 12, 87–93. doi: 10.1016/S0966-6362(00)00057-6

Albakri, G., Bouaziz, R., Alharthi, W., Kammoun, S., Al-Sarem, M., Saeed, F., et al. (2022). Phobia exposure therapy using virtual and augmented reality: a systematic review. *Appl. Sci.* 12:1672. doi: 10.3390/app12031672

Amjad, A., Halliday, D., Rosenberg, J. R., and Conway, B. A. (1997). An extended difference of coherence test for comparing and combining several independent coherence estimates: theory and applications to the study of motor units and physiological tremor. *J. Neurosci. Methods* 73, 69–79. doi: 10.1016/S0165-0270(96)02214-5

Bonnet, M., Bradley, M., Lang, P., and Requin, J. (1995). Modulation of spinal reflexes: arousal, pleasure, action. *Psychophysiology* 32, 367–372. doi: 10.1111/j.1469-8986.1995.tb01219.x.

Both, S., Everaerd, W., and Laan, E. (2003). Modulation of spinal reflexes by aversive and sexually appetitive stimuli. *Psychophysiology* 40, 174–183. doi: 10.1111/1469-8986.00019.

Both, S., Van Boxtel, G., Stekelenburg, J., Everaerd, W., and Laan, E. (2005). Modulation of spinal reflexes by sexual films of increasing intensity. *Psychophysiology* 42, 726–731. doi: 10.1111/j.1469-8986.2005.00364.x.

Boucsein, W., Fowles, D., Grimnes, S., Ben-Shakhar, G., Roth, W., Dawson, M., et al. (2012). Society for psychophysiological research Ad Hoc committee on electrodermal measures. Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 1017–1034. doi: 10.1111/j.1469-8986.2012.01384.x.

Burke, D., Gandevia, S., and McKeon, B. (1983). The afferent volleys responsible for spinal proprioceptive reflexes in man. *J. Physiol.* 339, 535–552. doi: 10.1113/jphysiol.1983.sp014732

Burke, D. (2021). Crosstalk opposing view: independent fusimotor control of muscle spindles in humans: there is little to gain. *J. Physiol.* 599, 2505–2508. doi: 10.1113/JP281337

Bush, J. (2008). Viability of virtual reality exposure therapy as a treatment alternative. *Comput. Hum. Behav.* 24, 1032–1040. doi: 10.1016/j.chb.2007.03.006

Capaday, C., and Stein, R. (1986). Amplitude modulation of the soleus H-reflex in the human during walking and standing. *J. Neurosci.* 6, 1308–1313 doi: 10.1523/JNEUROSCI.06-05-01308.1986

Carpenter, M., Frank, J., and Silcher, C. (1999). Surface height effects on postural control: a hypothesis for a stiffness strategy for stance. *J. Vestib. Res.* 9, 277–286. doi: 10.3233/VES-1999-9405

Carpenter, M., Frank, J., Adkin, A., Paton, A., and Allum, J. (2004). Influence of postural anxiety on postural reactions to multi-directional surface rotations. *J. Neurophysiol.* 92, 3255–3265. doi: 10.1152/jn.01139.2003

Cattagni, T., Martin, A., and Scaglioni, G. (2014). Is spinal excitability of the triceps surae mainly affected by muscle activity or body position? *J. Neurophysiol.* 111, 2525–2532. doi: 10.1152/jn.00455.2013

Cleworth, T. W., Chua, J. R., Inglis, T., and Carpenter, M. G. (2016). Influence of virtual height exposure on postural reactions to support surface translations. *Gait Posture* 47, 96–102. doi: 10.1016/j.gaitpost.2016.04.006

Cleworth, T., Horslen, B., and Carpenter, M. (2012). Influence of real and virtual heights on standing balance. *Gait Posture* 36, 172–176. doi: 10.1016/j.gaitpost.2012.02.010

Coelho, C., Waters, A., Hine, T., and Wallis, G. (2009). The use of virtual reality in acrophobia research and treatment. *J. Anxiety Disord.* 23, 563–574. doi: 10.1016/j.janxdis.2009.01.014

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge. doi: 10.4324/9780203771587

Dakin, C., Inglis, J., and Blouin, J. (2011). Short and medium latency muscle responses evoked by electrical vestibular stimulation are a composite of all stimulus frequencies. *Exp. Brain Res.* 209, 345–354. doi: 10.1007/s00221-011-2549-7

Dakin, C., Luu, B., van den Doel, K., Inglis, J., and Blouin, J. (2010). Frequency specific modulation of vestibular-evoked sway responses in humans. *J. Neurophysiol.* 103, 1048–1056. doi: 10.1152/jn.00881.2009

Dakin, C., Son, G., Inglis, J., and Blouin, J. (2007). Frequency response of human vestibular reflexes characterized by stochastic stimuli. *J. Physiol.* 583, 1117–1127. doi: 10.1113/jphysiol.2007.133264

Davis, J., Horslen, B., Nishikawa, K., Fukushima, K., Chua, R., Inglis, J., et al. (2011). Human proprioceptive adaptations during states of height-induced fear and anxiety. *J. Neurophysiol.* 106, 3082–3090. doi: 10.1152/jn.01030.2010

Dimitriou, M. (2021). Crosstalk proposal: there is much to gain from the independent control of human muscle spindles. *J. Physiol.* 599, 2501–2504. doi: 10.1113/JP281338

Dimitriou, M. (2022). Human muscle spindles are wired to function as controllable signal-processing devices. *eLife* 11:e78091. doi: 10.7554/eLife.78091

Dimitriou, M., and Edin, B. (2010). Human muscle spindles act as forward sensory models. *Curr. Biol.* 20, 1763–1767. doi: 10.1016/j.cub.2010.08.049

Faist, M., Hoefer, C., Hodapp, M., Dietz, V., Berger, W., and Duysens, J. (2006). In humans Ib facilitation depends on locomotion while suppression of Ib inhibition requires loading. *Brain Res.* 1076, 87–92. doi: 10.1016/j.brainres.2005.12.069

Halliday, D., Rosenberg, J., Amjad, A., Breeze, P., Conway, B., and Farmer, S. (1995). A framework for the analysis of mixed time series/point process data–theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Prog. Biophys. Mol. Biol.* 64, 237–278. doi: 10.1016/S0079-6107(96)00009-0

Hauck, L. (2011). *Understanding the Influence of Fear of Falling on Clinical Balance Control – Efforts in Fall Prediction and Prevention*. Waterloo, ON: University of Waterloo Library.

Hayashi, R., Tako, K., Tokuda, T., and Yanagisawa, N. (1992). Comparison of amplitude of human soleus H-reflex during sitting and standing. *Neurosci. Res.* 13, 227–233. doi: 10.1016/0168-0102(92)90062-H

Hjortskov, N., Skotte, J., Hye-Knudsen, C., and Fallentin, N. (2005). Sympathetic outflow enhances the stretch reflex response in the relaxed soleus muscle in humans. *J. Appl. Physiol.* 98, 1366–1370. doi: 10.1152/japplphysiol.00955.2004

Horlings, C., Carpenter, M., Kung, U., Honegger, F., Wiederhold, B., and Allum, J. (2009). Influence of virtual reality on postural stability during movements of quiet stance. *Neurosci. Lett.* 451, 227–231. doi: 10.1016/j.neulet.2008.12.057

Horslen, B., Dakin, C., Inglis, J., Blouin, J., and Carpenter, M. (2014). Modulation of human vestibular reflexes with increased postural threat. *J. Physiol.* 592, 3671–3685. doi: 10.1113/jphysiol.2014.270744

Horslen, B., Inglis, J., Blouin, J., and Carpenter, M. (2017). Both standing and postural threat decrease Achilles' tendon reflex inhibition from tendon electrical stimulation. *J. Physiol.* 595, 4493–4506. doi: 10.1113/JP273935

Horslen, B., Murnaghan, C., Inglis, J., Chua, R., and Carpenter, M. (2013). Effects of postural threat on spinal stretch reflexes: evidence for increased muscle spindle sensitivity? *J. Neurophysiol.* 110, 899–906. doi: 10.1152/jn.00065.2013

Horslen, B., Zaback, M., Inglis, J., Blouin, J., and Carpenter, M. (2018). Increased human stretch reflex dynamic sensitivity with height-induced postural threat. *J. Physiol.* 596, 5251–5265. doi: 10.1113/JP276459

Hulliger, M. (1984). The mammalian muscle spindle and its central control. *Rev. Physiol. Biochem. Pharmacol.* 101, 1–110.

Kamibayashi, K., Nakazawa, K., Ogata, H., Obata, H., Akai, M., and Shinohara, M. (2009). Invariable H-reflex and sustained facilitation of stretch reflex with heightened sympathetic outflow. *J. Electromyogr. Kinesiol.* 19, 1053–1060. doi: 10.1016/j.jelekin.2008.11.002

Krijn, M., Emmelkamp, P., Olafsson, R., and Biemond, R. (2004). Virtual reality exposure therapy of anxiety disorders: a review. *Clin. Psychol. Rev.* 24, 259–281. doi: 10.1016/j.cpr.2004.04.001

Levy, F., Leboucher, P., Rautureau, G., Komano, O., Millet, B., and Jouvent, R. (2016). Fear of falling: efficacy of virtual reality associated with serious games in elderly people. *Neuropsychiatr. DisTreat.* 12, 877–881. doi: 10.2147/NDT.S97809

Liddell, E., and Sherrington, C. (1924). Reflexes in response to stretch (myotatic reflexes). *Proc. R. Soc. Lond. B* 96, 212–242. doi: 10.1098/rspb.1924.0023

Liddell, E., and Sherrington, C. (1925). Further observations on myotatic reflexes. *Proc. R. Soc. Lond. B* 97, 267–283. doi: 10.1098/rspb.1925.0002

Lim, S., Cleworth, T., Horslen, B., Blouin, J., Inglis, T., and Carpenter, M. (2017). Postural threat influences vestibular-evoked muscular responses. *J. Neurophysiol.* 117, 604–611. doi: 10.1152/jn.00712.2016

Llewellyn, M., Yang, J., and Prochazka, A. (1990). Human H-reflexes are smaller in difficult beam walking than in normal treadmill walking. *Exp. Brain Res.* 83, 22–28. doi: 10.1007/BF00232189

Matthews, P. (1991). The human stretch reflex and the motor cortex. *Trends Neurosci.* 14, 87–91. doi: 10.1016/0166-2236(91)90064-2

McIlroy, W., Bishop, D., Staines, W., Nelson, A., Maki, B., and Brooke, J. (2003). Modulation of afferent inflow during the control of balancing tasks using the lower limbs. *Brain Res.* 961, 73–80. doi: 10.1016/S0006-8993(02)03845-3

Meehan, M., Insko, B., Whitton, M., and Brooks, F. (2002). Physiological measures of presence in stressful virtual environments. *ACM Trans. Graph.* 21, 645–652. doi: 10.1145/566654.566630

Meyerbröker, K., and Emmelkamp, P. (2011). "Virtual reality exposure therapy for anxiety disorders: the state of the art," in *Advanced Computational Intelligence Paradigms in Healthcare 6. Virtual Reality in Psychotherapy, Rehabilitation, and Assessment. Studies in Computational Intelligence*, Vol. 337, eds S. Brahnam and L. C. Jain (Berlin: Springer).

Mildren, R., Peters, R., Hill, A., Blouin, J., Carpenter, M., and Inglis, J. (2017). Frequency characteristics of human muscle and cortical responses evoked by noisy Achilles tendon vibration. *J. Appl. Physiol.* 122, 1134–1144. doi: 10.1152/japplphysiol.00908.2016

Mildren, R., Zaback, M., Adkin, A., Frank, J., and Bent, L. (2016). Reliability of the Achilles tendon tap reflex evoked during stance using a pendulum hammer. *Gait Posture* 43, 182–186. doi: 10.1016/j.gaitpost.2015.09.020

Miranda, Z., Pham, A., Elgbeili, G., and Barthélemy, D. (2019). H-reflex modulation preceding changes in soleus EMG activity during balance perturbation. *Exp. Brain Res.* 237, 777–791. doi: 10.1007/s00221-018-5459-0

Nafati, G., Rossi-Durand, C., and Schmied, A. (2004). Proprioceptive control of human wrist extensor motor units during an attention-demanding task. *Brain Res.* 1018, 208–220. doi: 10.1016/j.brainres.2004.05.066

Naranjo, E., Allum, J., Inglis, J., and Carpenter, M. (2015). Increased gain of vestibulospinal potentials evoked in neck and leg muscles when standing under height-induced postural threat. *J. Neurosci.* 293, 45–54. doi: 10.1016/j.neuroscience.2015.02.026

Naranjo, E., Cleworth, T., Allum, J., Inglis, J., Lea, J., Westerberg, B., et al. (2016). Vestibulo-spinal and vestibulo-ocular reflexes are modulated when standing with increased postural threat. *J. Neurophysiol.* 115, 833–842. doi: 10.1152/jn.00626.2015

Nielsen, E., Cleworth, T., and Carpenter, M. (2022). Exploring emotional-modulation of visually evoked postural responses through virtual reality. *Neurosci. Lett.* 777:136586. doi: 10.1016/j.neulet.2022.136586

Palmieri, R., Ingersoll, C., and Hoffman, M. (2004). The hoffmann reflex: methodologic considerations and applications for use in sports medicine and athletic training research. *J. Athl. Train.* 39, 268–277.

Pierrot-Deseilligny, E., and Burke, D. (2005). *The Circuitry of the Human Spinal Cord : Its Role in Motor Control and Movement Disorders*. Cambridge: Cambridge University Press.

Raffegeau, T., Fawver, B., Young, W., Williams, A., Lohse, K., and Fino, P. (2020). The direction of postural threat alters balance control when standing at virtual elevation. *Exp. Brain Res.* 238, 2653–2663. doi: 10.1007/s00221-020-05917-5

Rendon, A., Lohman, E., Thorpe, D., Johnson, E., Medina, E., and Bradley, B. (2012). The effect of virtual reality gaming on dynamic balance in older adults. *Age Ageing* 41, 549–552. doi: 10.1093/ageing/afs053

Rimer, E., Husby, L., and Solem, S. (2021). Virtual reality exposure therapy for fear of heights: clinicians' attitudes become more positive after trying VRET. *Front. Psychol.* 12:671871. doi: 10.3389/fpsyg.2021.671871

Roman-Liu, D. (2018). Age-related changes in the range and velocity of postural sway. *ArchGerontol. Geriatr.* 77, 68–80. doi: 10.1016/j.archger.2018.04.007

Rosenberg, J., Amjad, A., Breeze, P., Brillinger, D., and Halliday, D. (1989). The Fourier approach to the identification of functional coupling between neuronal spike trains. *Prog. Biophys. Mol. Biol.* 53, 1–31. 0079-6107(89)90004-7 doi: 10.1016/

Rothbaum, B., Hodges, L., and Kooper, R. (1997). Virtual reality exposure therapy. *Int. J. Psychother. Pract. Res.* 6, 219–226.

Sibley, K., Carpenter, M., Perry, J., and Frank, J. (2007). Effects of postural anxiety on the soleus H-reflex. *Hum. Mov. Sci.* 26, 103–112. doi: 10.1016/j.humov.2006.09.004

Simeonov, P., Hsiao, H., Dotson, B., and Ammons, D. (2005). Height effects in real and virtual environments. *Hum. Factors* 47, 430–438.

Simeonov, P., Hsiao, H., Powers, J., Ammons, D., Amendola, A., Kau, T.-Y., et al. (2008). Footwear effects on walking balance at elevation. *Ergonomics* 12, 1885–1905. doi: 10.1080/00140130802562625

Smith, R., Smoll, F., and Schutz, R. (1990). Measurement and correlates of sport-specific cognitive and somatic trait anxiety: the sport anxiety scale. *Anxiety Disord. Res.* 2, 263–280. doi: 10.1080/08917779008248733

Tanaka, T., Matsugi, A., Kamata, N., and Hiraoka, K. (2013). Postural threat increases corticospinal excitability in the trunk flexor muscles in the upright stance. *J. Psychophysiol.* 27, 165–172. doi: 10.1027/0269-8803/a000101

Tanaka, Y. (2015). Spinal reflexes during postural control under psychological pressure. *Motor Control* 19, 242–249. doi: 10.1123/mc.2013-0104

Van Doornik, J., Azevedo Coste, C., Ushiba, J., and Sinkjær, T. (2011). Positive afferent feedback to the human soleus muscle during quiet standing. *Muscle Nerve* 43, 726–732. doi: 10.1002/mus.21952

Young, J., Spence, A., and Behm, D. (2018). Roller massage decreases spinal excitability to the soleus. *J. Appl. Physiol.* 124, 950–959. doi: 10.1152/japplphysiol.00732.2017

Zaback, M., Reiter, E., Adkin, A., and Carpenter, M. (2021). Initial experience of balance assessment introduces 'first trial' effects on emotional state and postural control. *Gait Posture* 88, 116–121. doi: 10.1016/j.gaitpost.2021.05.013

Zehr, E. (2002). Considerations for use of the Hoffmann reflex in exercise studies. *Eur. J. Appl. Physiol.* 86, 455–468. doi: 10.1007/s00421-002-0577-5

Zehr, E., and Stein, R. (1999). Interaction of the Jendrássik maneuver with segmental presynaptic inhibition. *Exp. Brain Res.* 124, 474–480. doi: 10.1007/s002210050643

frontiers | Frontiers in Surgery

TYPE General Commentary
PUBLISHED 30 June 2023
DOI 10.3389/fsurg.2023.1218308

OPEN ACCESS

EDITED BY
Philipp Taussky,
Harvard Medical School, United States

REVIEWED BY
Alexander Thomas Yahanda,
Washington University in St. Louis, United States

*CORRESPONDENCE
Andrew Willett
acwill25@louisville.edu

RECEIVED 06 May 2023
ACCEPTED 20 June 2023
PUBLISHED 30 June 2023

CITATION
Willett A, Haq M, Holland J and Bridwell E (2023) Commentary: Augmented reality in neurosurgery, state of art and future projections. A systematic review. Front. Surg. 10:1218308. doi: 10.3389/fsurg.2023.1218308

Commentary: Augmented reality in neurosurgery, state of art and future projections. A systematic review

Andrew Willett*, Mohammad Haq, Joseph Holland and Elizabeth Bridwell

University of Louisville School of Medicine, Center for Health Process and Innovation, Bluegrass Biodesign, Louisville, KY, United States

KEYWORDS

augmented (virtual) reality, neurosurgery, pedicle screw accuracy, neuronavigation, neurooncology

A Commentary on:
Augmented Reality in Neurosurgery, State of Art and Future Projections. A Systematic Review

By Cannizzaro D, Zaed I, Safa A, Jelmoni AJM, Composto A, Bisoglio A, Schmeizer K, Becker AC, Pizzi A, Cardia A, Servadei F. (2022). Front Surg. 9:864792. doi: 10.3389/fsurg.2022.864792

Innovation and medicine are inseparable, and technologies such as augmented reality (AR) may transform the modern neurosurgical armamentarium. Per insights shared by Cannizzaro and colleagues in their review article "*Augmented Reality in Neurosurgery, State of Art and Future Projections. A Systematic Review*" (1), AR-assisted neurosurgery is a promising, albeit complex and challenging advancement. Over the past year our team of medical students has engaged in a biomedical innovation curriculum offered by a select cohort of accredited medical schools throughout the United States. We were tasked with evaluating the future of AR in neurosurgery. Thanks to our immersive experience with this unique technology, we felt inclined to comment on this article and share our perspectives in hopes that others will engage in these conversations and further promote AR in medicine, specifically within the field of neurosurgery. We found the review article to be robust and believe that it warrants a significant amount of consideration.

The authors highlight how AR has been largely investigated in spine surgeries, composing 18.2% of their literature review (1). This neurosurgical subspeciality has grown in minimally invasive techniques, a process that has been amplified with the utilization of neuronavigation systems. Studies suggest that AR-assisted pedicle screw placement is legitimate, with some reports sharing 100% accuracy (2) while others share 97.8%–98.5% accuracy (3). Although these numbers are encouraging, we must thoroughly question how this technology challenges more traditional surgical techniques. Why invest in AR if it offers no *significant* benefit? Dennler et al. (4) show that supplemental anatomical information provided via AR may help novice surgeons match the efficacy of expert surgeons with pedicle screw placement. However, to our knowledge, no large-scale randomized control trials to date have compared AR assisted pedicle screw placement vs. traditional pedicle screw placement. Moreover, we do not thoroughly understand if AR-assisted surgeries are cost-effective. We can glean insights from literature that compares the use of free hand

techniques with robotic devices; although robotics may have more accuracy in placing pedicle screws and can help decrease postoperative complications, the initial costs (which can approach upwards of $850,000) and increased OR time are thought to drastically outweigh its current benefits (5). Contrastingly, AR systems like Xvision have comparable profiles of improved accuracy but offer lower upfront cost (6), suggesting an opportunity for feasible integration. Other groups must not only reinforce the efficacy and practicality of AR, but also clearly analyze the fiscality of these technologies if veteran and novice surgeons alike are to adopt a new way of operating.

AR is an arguably more alluring tool for neurosurgical oncologists and neurosurgeons working in remote areas seeking to collaborate with distant colleagues, which is a realm we would have appreciated for Cannizzaro and colleagues to further explore. Our review of the available AR neuro-oncology literature has been exciting—particularly, a statistically significant improvement in percent of complete glioma resection in a test group (69.6%) compared to the control group (36.4%) ($p < .01$) has been reported (7). Other studies acknowledge the postoperative improvements associated with AR guided surgeries as patients who underwent AR-assisted tumor resection experienced shorter length of hospital stay and improved postoperative quality of life in comparison to non-AR guided resections (8). Moving forward, AR enthusiasts should emphasize a need to attend to extraoperative components affiliated with AR surgeries—if patients spend less time in the hospital and report higher quality of life following AR-guided surgery, then investing in these technologies becomes clearer. Thus, if we can better pinpoint where AR aligns with the needs of both the surgeon and patient, then integration can be met with less resistance.

Additionally, neurosurgical education is a valuable avenue to pursue, explaining why this field composed 18.2% of the authors' literature review (1). Rather than focusing on the traditional sense of education (i.e., training residents, revamping models of surgical instruction, etc.), developers in AR neurosurgery should emphasize how this technology enhances collaboration among neurosurgeons. AR expedites consultations between remote neurosurgeons, encouraging long-distance conferences regarding complex cases in real time (9). Accordingly, current platforms such as virtual interactive presence and augmented reality (VIPAR) place a surgeon in front of a stereoscopic display capable of remote interaction with a workstation situated in a different region or country (10, 11). The surgeon can share their insights on the respective cases and address gaps in access to care. What we find particularly intriguing about this utilization of AR in neurosurgery is that it does not drastically change workflow, which will encourage acceptance within the field.

Nonetheless, as Cannizzaro and colleagues (1) note, there are multiple obstacles that AR must overcome before diffuse integration into the neurosurgical operating room. There are two points we want to highlight from their speculations: (1) as AR will presumably have a role in the display of intraoperative images, there will be a growing need for strong microcomputers that facilitate access to data without compromising quality and (2) AR surgical headsets and hardware will need to be widely distributable, comfortable, non-obstructive, yet customizable to individual needs (1). We too find these concepts paramount to the future of virtually assisted surgeries. If AR can accurately augment intraoperative neuroimaging, it has the potential to significantly decrease the number of times a neurosurgeon must look away from the operating field, diminishing and possibly eliminating the need to constantly reorient from the patient to a distant screen in the operating room and back (12). Although some technologies allow for this, they are quite cumbersome for the surgeon, requiring them to wear a large, bulky headset (13). To further complicate matters, currently available AR technologies may not fully integrate with a surgeon's established hospital, software, and/or current neuronavigation systems, which makes a future with AR burdensome and laborious.

Naturally, these conversations will spark excitement and criticism, but we cannot let this outweigh the need to continue expanding and diversifying the future of medicine. We thank Cannizzaro and colleagues (1) for their thoughtful contributions to the current discourse and are excited to see how AR and neurosurgery will co-evolve.

Author contributions

The following authors were significant contributors to both the conceptual and physical design of the commentary, each participating in the relevant literature review, drafting and editing of the commentary, and the ultimate approval of the commentary for submission: AW, MH, JH, EG. We agree to be accountable for this work, and will openly welcome any questions or additional commentary from Cannizzarro and colleagues. All authors contributed to the article and approved the submitted version.

Acknowledgments

We thank the Bluegrass Biodesign Program affiliated with the University of Louisville School of Medicine. We also thank Joseph Neimat, Brent Wright, and In Kim for their input in this project and their continued mentorship.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Cannizzaro D, Zaed I, Safa A, Jelmoni AJM, Composto A, Bisoglio A, et al. Augmented reality in neurosurgery, state of art and future projections. A systematic review. *Front Surg*. (2022) 9:864792. doi: 10.3389/fsurg.2022.864792

2. Yahanda AT, Moore E, Ray WZ, Pennicooke B, Jennings JW, Molina CA. First in-human report of the clinical accuracy of thoracolumbar percutaneous pedicle screw placement using augmented reality guidance. *Neurosurg Focus*. (2021) 51(2):E10. doi: 10.3171/2021.5.FOCUS21217

3. Liu A, Jin Y, Co6rill E, Khan M, Westbroek E, Ehresman J, et al. Clinical accuracy and initial experience with augmented reality-assisted pedicle screw placement: the first 205 screws. *J Neurosurg Spine*. (2021) 36(3):1–7. doi: 10.3171/2021.2.SPINE202097

4. Dennler C, Jaberg L, Spirig J, Agten C, Gotschi T, Furnstahl P, et al. Augmented reality-based navigation increases precision of pedicle screw insertion. *J Orthop Surg Res*. (2020) 15(1):174. doi: 10.1186/s13018-020-01690-x

5. Ezeokoli EU, Pfennig M, John J, Gupta R, Khalil JG, Park DK. Index surgery cost of fluoroscopic freehand versus robotic-assisted pedicle screw placement in lumbar instrumentation: an age, sex, and approach-matched cohort comparison. *J Am Acad Orthop Surg Glob Res Rev*. (2022) 6(12): 3–6. doi: 10.5435/JAAOSGlobal-D-22-00137

6. Rush AJ III, Shepard N, Nolte M, Siemionow K, Phillips F. Augmented reality in spine surgery: current state of the art. *Int J Spine Surg*. (2022) 16(S2):S22–7. doi: 10.14444/8273

7. Sun GC, Wang F, Chen XL, Yu XG, Ma XD, Zhou DB, et al. Impact of virtual and augmented reality based on intraoperative magnetic resonance imaging and functional neuronavigation in glioma surgery involving eloquent areas. *World Neurosurg*. (2016) 96:375–82. doi: 10.1016/j.wneu.2016.07.107

8. Yang DL, Xu QW, Che XM, Wu JS, Sun B. Clinical evaluation and follow-up outcome of presurgical plan by dextroscope: a prospective controlled study in patients with skull base tumors. *Surg Neurol*. (2009) 72(6):682–9; discussion 689. doi: 10.1016/j.surneu.2009.07.040

9. Kockro RA, Stadie A, Schwandt E, Reisch R, Charalampaki C, Ng I, et al. A collaborative virtual reality environment for neurosurgical planning and training. *Neurosurgery*. (2007) 61(5 Suppl 2):379–91; discussion 391. doi: 10.1227/01.neu.0000303997.12645.26

10. Shenai MB, Dillavou M, Shum C, Ross D, Tubbs RS, Shih A, et al. Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance. *Neurosurgery*. (2011) 68(Suppl 1):200–7; discussion 207. doi: 10.1227/NEU.0b013e3182077efd

11. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. *World Neurosurg*. (2016) 86:103–11. doi: 10.1016/j.wneu.2015.08.053

12. Matsukawa K, Yato Y. Smart glasses display device for fluoroscopically guided minimally invasive spinal instrumentation surgery: a preliminary study. *J Neurosurg Spine*. (2020) 34(1):1–6. doi: 10.3171/2020.6.SPINE20644

13. Xvision: the future of surgery is within sight. Available at: https://augmedics.com

frontiers | Frontiers in Human Neuroscience

TYPE Systematic Review
PUBLISHED 04 August 2023
DOI 10.3389/fnhum.2023.1239374

OPEN ACCESS

EDITED BY
Yu Huang,
Soterix Medical Inc., United States

REVIEWED BY
Mohammad Mofatteh,
Queen's University Belfast, United Kingdom
Rabia Saleem,
University of Leicester, United Kingdom

*CORRESPONDENCE
Thomas C. Booth
✉ thomas.booth@kcl.ac.uk

RECEIVED 15 June 2023
ACCEPTED 20 July 2023
PUBLISHED 04 August 2023

CITATION
Robertshaw H, Karstensen L, Jackson B, Sadati H, Rhode K, Ourselin S, Granados A and Booth TC (2023) Artificial intelligence in the autonomous navigation of endovascular interventions: a systematic review. *Front. Hum. Neurosci.* 17:1239374. doi: 10.3389/fnhum.2023.1239374

Artificial intelligence in the autonomous navigation of endovascular interventions: a systematic review

Harry Robertshaw[1], Lennart Karstensen[2,3], Benjamin Jackson[1], Hadi Sadati[1], Kawal Rhode[1], Sebastien Ourselin[1], Alejandro Granados[1] and Thomas C. Booth[1,4]*

[1]School of Biomedical Engineering & Imaging Sciences, Kings College London, London, United Kingdom, [2]Fraunhofer IPA, Mannheim, Germany, [3]AIBE, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany, [4]Department of Neuroradiology, Kings College Hospital, London, United Kingdom

Background: Autonomous navigation of catheters and guidewires in endovascular interventional surgery can decrease operation times, improve decision-making during surgery, and reduce operator radiation exposure while increasing access to treatment.

Objective: To determine from recent literature, through a systematic review, the impact, challenges, and opportunities artificial intelligence (AI) has for the autonomous navigation of catheters and guidewires for endovascular interventions.

Methods: PubMed and IEEEXplore databases were searched to identify reports of AI applied to autonomous navigation methods in endovascular interventional surgery. Eligibility criteria included studies investigating the use of AI in enabling the autonomous navigation of catheters/guidewires in endovascular interventions. Following Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA), articles were assessed using Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2). PROSPERO: CRD42023392259.

Results: Four hundred and sixty-two studies fulfilled the search criteria, of which 14 studies were included for analysis. Reinforcement learning (RL) (9/14, 64%) and learning from expert demonstration (7/14, 50%) were used as data-driven models for autonomous navigation. These studies evaluated models on physical phantoms (10/14, 71%) and *in-silico* (4/14, 29%) models. Experiments within or around the blood vessels of the heart were reported by the majority of studies (10/14, 71%), while non-anatomical vessel platforms "idealized" for simple navigation were used in three studies (3/14, 21%), and the porcine liver venous system in one study. We observed that risk of bias and poor generalizability were present across studies. No procedures were performed on patients in any of the studies reviewed. Moreover, all studies were limited due to the lack of patient selection criteria, reference standards, and reproducibility, which resulted in a low level of evidence for clinical translation.

Conclusion: Despite the potential benefits of AI applied to autonomous navigation of endovascular interventions, the field is in an experimental proof-of-concept stage, with a technology readiness level of 3. We highlight that reference standards with well-identified performance metrics are crucial to allow for comparisons of data-driven algorithms proposed in the years to come.

Systematic review registration: identifier: CRD42023392259.

KEYWORDS

artificial intelligence, machine learning, endovascular intervention, autonomy, navigation

1. Introduction

Cardiovascular (CV) diseases are the most common cause of death across Europe, accounting for more than four million deaths each year, with coronary heart disease (44.2%) and cerebrovascular disease (25.4%) emerging as the predominant contributors to CV-related mortality across all ages and genders (Townsend et al., 2016). Endovascular catheter-based interventions such as percutaneous coronary intervention (PCI), pulmonary vein isolation (PVI), and mechanical thrombectomy (MT) have become an established treatment for CV diseases (Thukkani and Kinlay, 2015; Goyal et al., 2016; Giacoppo et al., 2017; Lindgren et al., 2018). During such a procedure, an operator navigates a guidewire and catheter from an insertion point (typically the common femoral or radial artery) to the area of interest to perform the intervention. Intraoperative fluoroscopy is used intermittently throughout the navigation and intervention to guide the catheter and guidewire through the vasculature. Once the target site has been reached, the treatment can be performed through the catheter. This is typically thrombus removal in the case of MT, stent deployment in the case of PCI, and ablation for PVI (Brilakis, 2020).

In acute CV disease, time from symptom onset to treatment is often crucial for effective endovascular interventions. For example, the benefits of MT become non-significant after 7.3 h of stroke for non-stratified patients (Saver et al., 2016). As a result, in the UK for example, only 1.4% of stroke admissions benefit from MT despite the 10% of patients that are eligible for treatment (McMeekin et al., 2017). Other challenges for endovascular interventions relate to occasional complications including perforation, thrombosis, and dissection in the parent artery, as well as distal embolization of thrombus (Hausegger et al., 2001). Moreover, angiography requires intravascular contrast agent administration, which can occasionally lead to nephrotoxicity (Rudnick et al., 1995). For operators and their teams, the high cumulative dose of x-ray radiation from angiography is a risk factor for cancer and cataracts (Klein et al., 2009). Although exposure can be minimized with current radiation protection practice, some measures involve operators wearing heavy protective equipment which is a risk factor for orthopedic complications, and so alternative methods of exposure reduction are beneficial (Ho et al., 2007; Madder et al., 2017).

It is hoped that robotic surgical systems can either mitigate or eliminate some of the challenges currently presented by endovascular interventions. For example, robotic systems could be set up in hospitals nationwide and tele-operated remotely from a central location, increasing the speed of access to treatments such as MT beyond what is possible currently (Crinnion et al., 2022). Additionally, robotic systems might eliminate any operator physiological tremors or fatigue and allow endovascular interventions to be performed in an optimum ergonomic position while potentially increasing procedural precision (for example, procedure time), and thereby improving overall performance scores and reducing complication rates (Riga et al., 2010). Furthermore, as operators would not be required to stand next to the patient, their radiation exposure would be reduced and the need to wear heavy protective equipment would be obviated.

Commercial robotic systems are currently available to perform endovascular interventions. Hansen Medical developed the *Magellan*™ system (Auris Health, Redwood City, USA), the first commercially available robotic system to be used for PVI, and more recently used to successfully perform carotid artery stenting in 13 patients (Duran et al., 2014; Jones et al., 2021). This system comprises a steerable guide catheter inside a steerable sheath allowing movement in three dimensions, and a separate remote guidewire manipulator allowing linear and rotational movement. The *Corpath GRX*® (Corindus Vascular Robotics, USA), the next-generation system of the *Corpath*® 200 robot, has successfully been used for PCI and PVI. This system has performed diagnostic cerebral angiography procedures and ten carotid artery stenting procedures (Nogueira et al., 2020; Sajja et al., 2020; Weinberg et al., 2021). Furthermore, it has been recently used to perform robot-assisted, neuroendovascular interventions including aneurysm embolization and epistaxis embolization (Pereira et al., 2020; Cancelliere et al., 2022; Saber et al., 2022). These systems use a controller-operator structure, where operators remotely control and navigate a robot through a patient's vasculature to the target site. In currently available systems, the operator has complete control over the robot and makes all of the decisions.

While these robotic systems help alleviate some of the challenges of endovascular interventions, they have limitations. The controller-operator structure requires a reasonably high cognitive workload, can still result in human error and means that the procedure is limited to an individual operator's skill set (Mofatteh, 2021). These robotic systems also consist of user interfaces such as buttons and joysticks, requiring skills that are different to those used in current clinical practice. Additionally, a lack of haptic feedback from robotic systems might result in difficulties to receive tactile feedback from the catheters and guidewires as they interact with vessel walls (Crinnion et al., 2022).

One emerging method of mitigating these challenges is using artificial intelligence (AI) techniques in conjunction with robotic systems. AI, and in particular, machine learning (ML), has accelerated in recent years in its applications for data analysis and learning (Sarker, 2021), with many areas of healthcare already making use of this technology for disease prediction and diagnosis (Fatima and Pasha, 2017; Silahtaroğlu and Yılmaztürk, 2021). ML algorithms can be divided into three main groups: supervised, unsupervised, and reinforcement learning (RL). Supervised learning is the most common form of ML and involves constructing a model trained on a dataset with labels (the corresponding correct outputs). The model can then accurately predict the labels of new, unknown instances based on the patterns learned from the training data (Kotsiantis, 2007).

Unsupervised learning involves training an algorithm to represent particular input features in a way that reflects the structure of the overall collection of input patterns (Dayan, 2017). In contrast to other types of ML, the dataset is unlabeled and there are no explicit target outputs or environmental evaluations associated with each input.

RL is a form of ML, whereby an agent learns by interacting with the environment and receiving feedback in the form of rewards. The goal of RL is to maximize the cumulative reward over time by learning a policy that optimizes the agent's current state for a set of actions (Arulkumaran et al., 2017). Similar to the natural way of human learning, robotic RL automatically acquires the skills

through "trials and errors" (Sutton and Barto, 2018). Applications of RL are becoming more expansive, as numerous research areas aim to use the method, for example, in precision medicine, medical imaging, and rehabilitation (Lowery and Faisal, 2013; Naros and Gharabaghi, 2015; Ghesu et al., 2018).

Learning from demonstration (LfD) is a variant of supervised learning, where input data is provided by an expert demonstrator. This can also act as a precursor for RL, whereby the agent can further improve its behavior through interaction with the environment. Table 1 describes the ML methods that are referred to later in this paper, each of which can be used to improve performance across the three types of ML described above. LfD has been separated from the other types of ML in this case, as it can be used in the context of both supervised learning and RL.

The use of these ML techniques for autonomy in medical robotics presents several challenges. To help in the consideration of regulatory, ethical, and legal barriers imposed, a six-level autonomy framework has been proposed, ranging from no autonomy at level 0, up to level 5 which involves full autonomy with no human intervention (Yang et al., 2017). This study aims to systematically review the methodology, performance and autonomy level of AI applied to the autonomous navigation of catheters and guidewires for endovascular interventions. Understanding the current developments in the field will help determine the impact, challenges, and opportunities required to direct future translational research and ultimately guide clinical practice.

2. Methods

This systematic review is PROSPERO (International prospective register of systematic reviews) registered (CRD42023392259). The review followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021).

2.1. Selection criteria

2.1.1. Eligibility criteria

Included reports consisted of primary research studies, which investigated the use of AI in enabling the autonomous navigation of catheters and/or guidewires in endovascular interventions. Excluded studies did not use AI methods to achieve autonomous navigation of catheters/guidewires or looked at path planning for endovascular interventions rather than the navigation itself. Additionally, studies without an English translation were not included (Nussbaumer-Streit et al., 2020).

2.1.2. Information sources and search strategy

PubMed and IEEEXplore were used to capture original research articles, published anytime until the end of January 2023, with the following search query: "(Artificial Intelligence OR Machine Learning OR Reinforcement Learning OR Deep Learning OR Autonomous OR Learning-based) AND (Endovascular OR Vascular Intervention OR Catheter OR Guidewire) AND (Navigation OR Guidance)." Pre-prints and non-peer-reviewed articles were excluded.

2.1.3. Selection and data collection process

A medical robotics data scientist, H.R. (3 years of research experience), searched for studies as defined in the search strategy and followed the selection process as shown in Figure 1. A medical robotics data scientist, L.K. (4 years experience in autonomous endovascular navigation using AI), independently reviewed the manuscripts against the eligibility criteria. In the case of discrepancy, consensus was reached by discussion between the two reviewers. If consensus was not reached, the multi-disciplinary authorship would make the final arbitration. The relevant data items, as defined in the following section, were extracted.

2.2. Data items, effect measures, and synthesis methods

Information extracted from each study included: the AI method used and more granular model details (where available), the current level of autonomy, the type of experiment (*in vivo*, *in vitro*, *in silico*), the method of tracking the catheter and/or guidewire position, the method of catheter and/or guidewire manipulation, description of the navigation path, performance measures, and key performance outcomes (where available).

The levels of autonomy followed (Yang et al., 2017). Briefly, these are level 0: no autonomy, level 1: robot assistance, level 2: task autonomy, level 3: conditional autonomy, level 4: high autonomy, and level 5: full autonomy. It should be noted that if the autonomy level was not described in the study, an appropriate level was assigned based on the content of the paper.

2.3. Study risk of bias, reporting bias, and certainty assessment

Where appropriate, both Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) methodology alongside AI metrics from the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) were used to assess the risk of bias for each study (Rutjes et al., 2011; Mongan et al., 2020).

3. Results

3.1. Studies

As shown in Figure 1, 462 studies met the search criteria, and 21 full-text studies were assessed against the eligibility criteria. A total of 14 were identified for review (Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020; Behr et al., 2019; You et al., 2019; Zhao et al., 2019; Kweon et al., 2021; Meng et al., 2021, 2022; Cho et al., 2022; Karstensen et al., 2022; Wang et al., 2022). The characteristics of the fourteen studies are listed in Table 2.

According to QUADAS-2 methodology, all studies reviewed gave a high or unclear "risk of bias" and "concerns regarding

TABLE 1 Description of ML methods.

Name	ML Type	Description
A3C	RL	An algorithm that employs multiple agents working in parallel to learn policies in an environment (Mnih et al., 2016).
Behavior cloning	LfD	technique where an agent learns a policy by imitating expert behavior. It learns from labeled examples provided by experts, mapping input observations to corresponding actions to replicate the demonstrated behavior. It can be used as a pre-training step in RL allowing the agent to learn by imitating the behavior of an expert (Codevilla et al., 2019).
CNN	Supervised learning	Type of deep neural network specifically designed for image processing and pattern recognition tasks. CNNs leverage spatial hierarchies through convolutional layers that extract local features and preserve spatial relationships, enabling effective image classification, object detection, and image segmentation tasks (O'Shea and Nash, 2015).
DDPG	RL	An algorithm that merges RL and policy optimization. It iteratively refines the policy based on estimated value distributions, to find an optimal strategy (Lillicrap et al., 2016).
DQN	RL	Leverages a deep neural network to learn optimal policies through Q-learning (see Q-learning explanation below). It enables agents to make decisions by maximizing the expected cumulative rewards, facilitating dynamic environment interaction (Mnih et al., 2013).
Dueling DQN	RL	An extension of DQN that separates the estimation of state value and action advantages. By independently approximating these values, the agent can learn the value of being in a particular state while also considering the advantages of each action (Wang et al., 2016).
GAIL	LfD	Method where an agent learns a policy by imitating expert behavior using a generative adversarial framework. It involves a generator network that aims to replicate the expert and a discriminator network that distinguishes between expert and generated behavior (Ho and Ermon, 2016).
GMM	Unsupervised learning	A statistical model that assumes data is generated by a mixture of several Gaussian distributions (Reynolds, 2015).
HD	LfD	Term that encompasses the process of an expert performing a task. Human demonstration can be used as a means to collect data for LfD (Nair et al., 2017).
HER	RL	Allows an agent to learn from "failed" experiences by redefining the goal of a task (Andrychowicz et al., 2017).
HMM	Unsupervised learning	A statistical model that assumes observations are generated by a hidden sequence of states that follow a Markov process (Rabiner, 1989).
PI^2	RL	Optimization algorithm which aims to find the optimal policy by iteratively improving the policy through gradient-based optimization methods, maximizing the expected return (Theodorou et al., 2010).
PPO	RL	An algorithm that optimizes policies iteratively while ensuring small policy updates. It balances exploration and exploitation, enhancing stability, and sample efficiency during training (Schulman et al., 2017).
Q-learning	RL	Algorithm that learns the optimal action-value function (Q-value function) for sequential decision-making. It updates Q-values iteratively based on observed rewards and the maximum expected future rewards (Jang et al., 2019).
Rainbow	RL	Extension of DQN that combines multiple improvements to enhance performance, by incorporating techniques such as prioritized experience replay, distributional value estimation, and multi-step learning to improve overall learning stability and efficiency (Hessel et al., 2017).
YOLO	Supervised learning	Object detection algorithm that can detect and classify objects in real-time. It uses a single neural network to directly predict bounding boxes and class probabilities for objects in an image, providing fast and accurate object detection (Redmon et al., 2015).

A3C, asynchronous advantage actor critic; CNN, convolutional neural network; DDPG, deep deterministic policy gradient; DQN, deep Q-network; GAIL, generative adversarial imitation learning GMM, Gaussian mixture modeling; HD, human demonstration; HER, hindsight experience replay; HMM, hidden Markov models; LfD, learning from demonstration; PI2, policy improvement with path integrals; PPO, proximal policy optimization; RL, reinforcement learning; YOLO, you only look once.

applicability" in all domains. No studies performed procedures on patients and therefore had no clearly defined patient selection criteria, reference standards, or index tests. Despite the low level of evidence, there is value in discussing these individual studies as they represent the current state of the art and form a baseline for further research.

3.2. AI models

3.2.1. RL methods

RL was used in nine studies (9/14, 64%) with algorithms including A3C, DDPG, DQN, Dueling DQN, HER, PI^2, PPO, and Rainbow (Chi et al., 2018a, 2020; Behr et al., 2019; You et al., 2019; Kweon et al., 2021; Meng et al., 2021, 2022; Cho et al., 2022; Karstensen et al., 2022). Demonstrator data in some form (GAIL, Behavior Cloning, or HD) was used as a precursor in four of the studies (4/14, 29%) during training (LfD), in conjunction with other RL algorithms (Chi et al., 2018a; Behr et al., 2019; Kweon et al., 2021; Cho et al., 2022). The SOFA framework (Inria, Strasbourg, France; Faure et al., 2012) was used for training RL models in four studies (4/14, 29%; Behr et al., 2019; Cho et al., 2022; Karstensen et al., 2022; Meng et al., 2022), the Unity engine (Unity Technologies, San Francisco, USA) was used in two studies (2/14, 14%; You et al., 2019; Meng et al., 2021), while the platform used for training was not specified in three

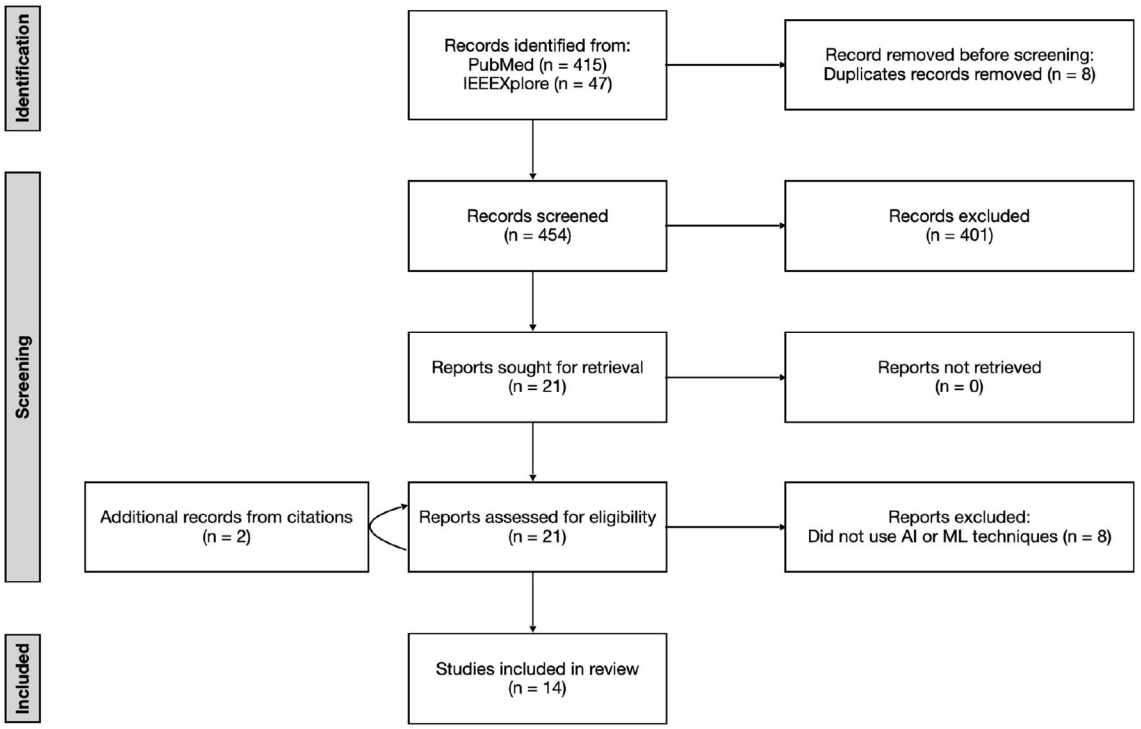


FIGURE 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram showing the number of articles searched and excluded at each stage of the literature search after screening titles, abstracts, and full texts.

studies (3/14, 21%; Chi et al., 2018a, 2020; Kweon et al., 2021).

3.2.2. Other ML types

RL was not used in five studies (5/14, 36%) which employed LfD (not as a precursor for RL), unsupervised (GMM and HMM) and supervised (CNN and YOLO) methods alone or in combination (Rafii-Tari et al., 2013, 2014; Chi et al., 2018b; Zhao et al., 2019; Wang et al., 2022). The most common method was LfD, which was used in three studies (3/14, 21%; Rafii-Tari et al., 2013, 2014; Chi et al., 2018b). Two of these studies (2/14, 14%) used a GMM to generate the probabilistic representation of the dataset provided by a demonstrator (Rafii-Tari et al., 2013; Chi et al., 2018b), while the other study utilized HMMs to model each movement primitive (Rafii-Tari et al., 2014). The other two non-RL studies (2/14, 14%) used solely CNNs or YOLOV5s (Zhao et al., 2019; Wang et al., 2022).

3.3. Level of autonomy

Conditional autonomy (level 3) was performed in seven studies (7/14, 50%; Behr et al., 2019; You et al., 2019; Zhao et al., 2019; Kweon et al., 2021; Cho et al., 2022; Karstensen et al., 2022; Wang et al., 2022). Here, a target in the vasculature is selected by an operator and the subsequent navigation to the target of the guidewire and/or catheter takes place autonomously. Task autonomy (level 2) was performed across five studies (5/14, 36%), whereby the robotic driver automates the catheter motion and an operator manipulates the guidewire for assistance (Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020). Robot assistance (level 1) was demonstrated in two studies (2/14, 14%), where experiments were performed entirely in simulation and under continuous supervision of an operator (Meng et al., 2021, 2022).

3.4. Experimental design

Clinical trials were not performed in any of the studies reviewed. Physical phantoms were used in the majority of studies (11/14, 79%) reviewed (Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020; Behr et al., 2019; You et al., 2019; Zhao et al., 2019; Kweon et al., 2021; Cho et al., 2022; Wang et al., 2022). Of these studies, seven used 3D vascular phantoms (Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020; You et al., 2019; Wang et al., 2022), three used 2D phantoms (Behr et al., 2019; Zhao et al., 2019; Cho et al., 2022), and one study used both 2D and 3D phantoms (Kweon et al., 2021). Commercial phantoms were used in six studies (6/14, 43%): 3D silicone-based, transparent, anthropomorphic phantoms (Elastrat Sarl, Geneva, Switzerland) were used in 5/14 (36%) studies (Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020); and the study using both 2D and 3D phantoms used firstly, a 2D PCI trainer for beginners (Medi Alpha Co., Ltd., Tokyo, Japan) and secondly, a silicone 3D Embedded Coronary Model (Trandomed 3D Medical Technology Co., Ltd., Ningbo, China), respectively. Five studies

(5/14, 36%) appeared to use in-house phantoms: one study used a silicone-based 3D printed heart model and inferior vena cava (You et al., 2019), and one used a 10 mm vessel diameter phantom made of polymethyl methacrylate (PMMA; Behr et al., 2019), while sufficient phantom detail is not provided by the other three studies (Zhao et al., 2019; Cho et al., 2022; Wang et al., 2022).

In silico methods were used in two of the studies (2/14, 14%; one used SOFA framework and one used Unity engine; Meng et al., 2021, 2022). *Ex vivo* experiments using porcine liver vasculature were reported by one study (Karstensen et al., 2022). Here, *in silico* methods were used for training models before the *ex vivo* experiments.

Figure 2 shows the anatomical regions where each study focuses. Experiments within or around the blood vessels of the heart were reported by the majority of studies (10/14, 71%; Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020; You et al., 2019; Kweon et al., 2021; Meng et al., 2021, 2022; Wang et al., 2022), with the study with the longest path length starting at the femoral artery and finishing at the coronary artery (Wang et al., 2022). Non-anatomical vessel platforms "idealized" for simple navigation were used in three studies (3/14, 21%; Behr et al., 2019; Zhao et al., 2019; Cho et al., 2022), and the porcine liver venous system in one study (Karstensen et al., 2022).

3.5. Evaluation

Passive tracking relies on external sensors to detect the catheter's position, active tracking involves the use of sensors located at the distal end of the catheter for real-time position tracking, and magnetic tracking utilizes external magnetic fields to guide the catheter's movement and track its position. A passive, tracking-based, method for catheter manipulation was used in eight studies (8/14, 57%; Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020; You et al., 2019; Meng et al., 2021, 2022), whereas a passive, image-based, method for catheter manipulation was used in the other six studies (6/14, 43%; Behr et al., 2019; Zhao et al., 2019; Kweon et al., 2021; Cho et al., 2022; Karstensen et al., 2022; Wang et al., 2022). None of the studies reviewed reported active or magnetic steering methods.

A top-down camera for tracking the location of the guidewire and/or catheter was implemented in five of the studies (5/14, 36%) where transparent phantoms allowed real-time video to provide software-generated tracking data (Behr et al., 2019; Zhao et al., 2019; Kweon et al., 2021; Cho et al., 2022; Wang et al., 2022). Electromagnetic (EM) position sensors were employed in six studies (6/14, 43%; Rafii-Tari et al., 2013, 2014; Chi et al., 2018a,b, 2020; You et al., 2019). An Aurora control unit and EM Generator of Aurora electromagnetic tracking system (NDI, Waterloo, Canada) were used in one of these studies (You et al., 2019), whilst custom-designed sensors (Rafii-Tari et al., 2013) were used in the other five. These five studies also employed a top-down camera simultaneously enabled through the use of transparent phantoms during data collection pre-training. One study employed continuous fluoroscopy, capturing 7.5 images per second, and used a CNN to segment the guidewire from real-time fluoroscopy images to track data that included the coordinates (Karstensen et al., 2022).

Two studies (2/14, 14%) were performed entirely *in silico*, and hence no tracking method was required (Meng et al., 2021, 2022).

Quantitative performance measures used in the studies were heterogeneous which may reflect the low technology readiness level (TRL; Mankins, 1995) of AI applied to autonomous navigation of endovascular interventions shown by the studies in this systematic review. Common performance measures used were success rate of navigation task (7/14, 50%) and time to complete procedure (5/14, 36%). Other performance measures shared across studies were: measures of force (6/14, 43%); acceleration (4/14, 29%); various measures of speed (4/14, 29%); and path length (4/14, 29%). Half of the studies (7/14, 50%) reviewed compared manual performance against their autonomous navigation performance. The key performance outcomes of the 14 studies are listed in Table 3.

Where possible, critical outcome data for success rate, procedure time and path length were extracted from the study. Three of the 14 studies (3/14, 21%) did not measure any of these performance measures (Rafii-Tari et al., 2013; Meng et al., 2021; Wang et al., 2022). Of the seven studies (7/14, 50%) that measured success rate, the value was over 90% in four studies (4/14, 29%; Chi et al., 2018a, 2020; Zhao et al., 2019; Kweon et al., 2021).

4. Discussion

4.1. Summary of findings

There is no high-level evidence (Howick et al., 2011) to demonstrate that AI autonomous navigation of catheters and guidewires in endovascular intervention is non-inferior or superior to manual procedures. Currently, AI autonomous navigation of catheters and guidewires in endovascular intervention has not surpassed TRL 3. There has been no clinical validation nor has there been comprehensive laboratory validation. Over half of the studies (9/14, 64%) employed RL methodologies, particularly in recent years, where most studies used RL (8/10, 80% published beyond 2018). There are no standardized *in silico*, *in vitro*, or *ex vivo* experimental reference standard designs, nor are there standardized performance measures, meaning comparison of studies quantitatively is of limited value.

4.2. Strengths and limitations

4.2.1. Strengths

The primary strength of the studies reviewed came from the range of ML techniques employed. Most focused on finding a ML technique that would improve upon previous work, rather than using similar algorithms and extending the experimental environment. This is demonstrated well within the nine studies (9/14, 64%) which used RL, where a different ML-based methodology was used in every case except for two (where the simulation environment and output measurements were changed between studies). Exploring various techniques is advantageous for research, especially in the rapidly evolving field of ML, as the fast pace of development increases the likelihood that more effective algorithms are created. For example, autonomous endovascular

TABLE 2 Studies resulting from our search and eligibility criteria proposing AI models for the autonomous navigation of catheters/guidewires in endovascular interventions.

References	ML method*	Level of autonomy	Validation environment	Tracking type	Tracking method	Navigation path	Performance measures
RL							
Chi et al. (2018b)	PI^2 + LfD	Level 2	*In vitro* (phantom)	Passive (tracking-based)	EM position sensor	Origin of LCA, to BCA or LSA	Max acceleration of catheter tip, Mean/max/STDEV/impact area of contact force, Path-following error (RMSE), Mean/STDEV speed of catheter tip, Path length, Procedure time
Behr et al. (2019)	DDPG, DQN, HER, + HD	Level 3	*In vitro* (phantom)	Passive (image-based)	Top-down camera	Idealized vessel platform with a bifurcation followed by a bi-and trifurcation in one plane	SR of navigation task
You et al. (2019)	Dueling DQN	Level 3	*In vitro* (phantom)	Passive (tracking-based)	EM position sensor	Insertion at heart RA's IVC, target is nerve nodes around CS and TA of heart RA	Path length, SR of navigation task
Chi et al. (2020)	PPO + GAIL	Level 2	*In vitro* (phantom)	Passive (tracking-based)	EM position sensor	Position in aorta (proximal to major branches), to BCA or LCCA	Mean/max force between endovascular instruments and vascular phantom, Mean/STDEV speed of catheter tip, Path length, Procedure time, SR of navigation task
Cho et al. (2022)	DDPG + Behavior Cloning	Level 3	*In vitro* (phantom)	Passive (image-based)	Top-down camera	Idealized vessel platform with a bifurcation followed by a bifurcation in one plane	Procedure time
Meng et al. (2021)	A3C	Level 1	*In silico*	Passive (tracking-based)	Simulation-based	Traversing descending aorta, through aortic arch, cannulation of LCA, LSA, or innominate artery	Limited information available
Kweon et al. (2021)	Rainbow + HD	Level 3	*In vitro* (phantom)	Passive (image-based)	Top-down camera	Proximal point in left anterior descending artery to target location in main or side branch	Procedure time, SR of navigation task
Meng et al. (2022)	A3C	Level 1	*In silico*	Passive (tracking-based)	Simulation-based	Traversing descending aorta, through aortic arch, cannulation of LCA, LSA, or innominate artery	Contact force, Procedure time
Karstensen et al. (2022)	DDPG, HER	Level 3	*Ex vivo* (porcine liver)	Passive (image-based)	Fluoroscopy	Vena cava inferior to vena hepatica dextra, vena hepatica intermedia or and vena hepatica sinistra (porcine liver)	Number of failures due to wrong branch/entanglement, SR of navigation task
Non-RL							
Rafii-Tari et al. (2013)	GMM + LfD	Level 2	*In vitro* (phantom)	Passive (tracking-based)	EM position sensor	Traversing descending aorta, through aortic arch, cannulation of innominate artery	Mean/max acceleration of catheter tip, Mean/max speed of catheter tip
Rafii-Tari et al. (2014)	HMM + LfD	Level 2	*In vitro* (phantom)	Passive (tracking-based)	EM position sensor	Cannulation of LSA and RCCA	Mean/max acceleration of catheter tip, Path length
Chi et al. (2018a)	GMM + LfD	Level 2	*In vitro* (phantom)	Passive (tracking-based)	EM position sensor	Origin of LCA, to bifurcation site between RCCA and RSA	Mean/max acceleration of catheter tip, Mean/max/STDEV/impact area of contact force, Mean/max/STDEV speed of catheter tip, Path length, SR of navigation task
Zhao et al. (2019)	CNN	Level 3	*In vitro* (phantom)	Passive (image-based)	Top-down camera	Medical and designed vessel models	Procedure time, SR of navigation task
Wang et al. (2022)	YOLOV5s	Level 3	*In vitro* (phantom)	Passive (image-based)	Top-down camera	Femoral to coronary artery	Average Precision

Clinical: BCA, brachiocephalic artery; CS, coronary sinus; EM, electromagnetic; IVC, inferior vena cava; LCA, left coronary artery; LCCA, left common carotid artery; LSA, left subclavian artery; RA, right atrium; RCCA, right common carotid artery; RSA, right subclavian artery; TA, transaortic.

Technical: A3C, asynchronous advantage actor critic; CNN, convolutional neural network; DDPG, deep deterministic policy gradient; DQN, deep Q-network; GAIL, generative adversarial imitation learning GMM, Gaussian mixture modeling; HD, human demonstration; HER, hindsight experience replay; HMM, hidden Markov models; LfD, learning from demonstration; PI^2, policy improvement with path integrals; PPO, proximal policy optimization; RL, reinforcement learning.

Evaluation: RMSE, root-mean-squared error; SR, success rate; STDEV, standard deviation.

*LfD was used as a ML method in cases where no further information about the type of LfD was available.

Descriptions of each type of ML method can be found in Table 1.

10.3389/fnhum.2023.1239374

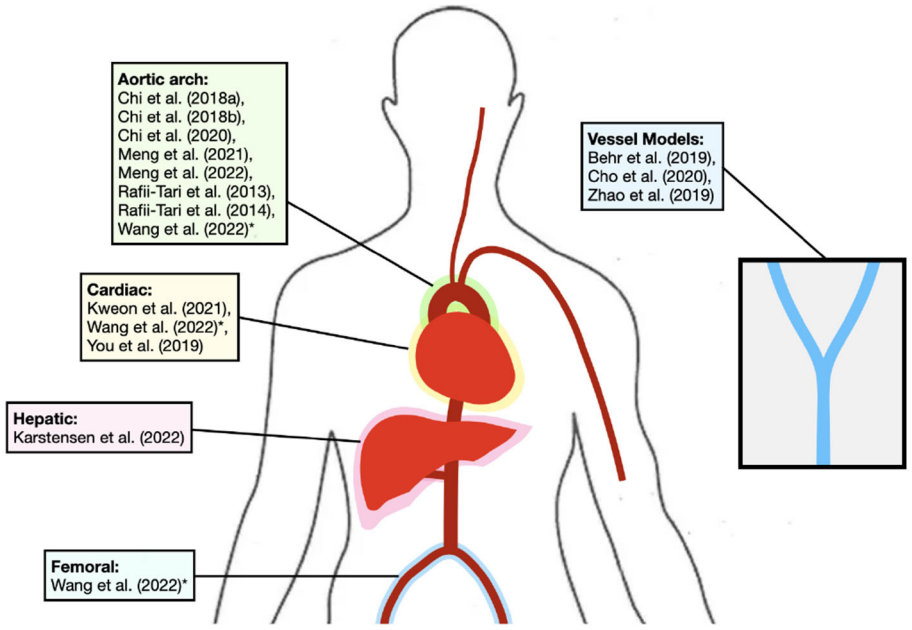


FIGURE 2
Diagram depicting the general vessels of interest for each study. *Study is in more than one area. Studies using non-anatomical platforms are also shown.

TABLE 3 Key performance outcomes from studies reviewed.

References	Path length	Procedure time	Success rate
Behr et al. (2019)	n/a	n/a	70% for DDPG
Chi et al. (2018a)	Median: 360.5 mm manual, 281.2 mm robot	n/a	Expert model: 100% under dry condition, 94.4% under continuous flow, 55.6% under pulsatile flow
Chi et al. (2018b)	367.8 mm pre-RL, 211.6 mm RL	74.5 ± 11.6 s manual, 137.7 ± 7.1 s pre-RL, 121.3 ± 9.5 s RL	n/a
Chi et al. (2020)	Type-I Aortic Arch, BCA: 55.7 ± 9.4 mm automation, 51.4 ± 8.3 mm manual	Type-I Aortic Arch, BCA: 52.1 ± 9.9 s automation, 6.36 ± 1.4 s manual	Type-I Aortic Arch: 94.4% for BCA cannulation, 88.9% for LCCA cannulation
Cho et al. (2022)	n/a	Real vessel phantom: 34.06 s own algorithm, 63.2 s expert algorithm	n/a
Karstensen et al. (2022)	n/a	n/a	30% (*ex-vivo* surgical task)
Kweon et al. (2021)	n/a	Proximal targets: 9.29 ± 6.00 s autonomous, 82.1 ± 34.2 s manual	>95% after 646 episodes (distal-main target)
Meng et al. (2021)	n/a	n/a	n/a
Meng et al. (2022)	n/a	97.35 s manual, 68.61 s training	n/a
Rafii-Tari et al. (2013)	n/a	n/a	n/a
Rafii-Tari et al. (2014)	2.9 m LSA manual intermediate, 0.44 m LSA robot intermediate	n/a	n/a
Wang et al. (2022)	n/a	n/a	n/a
You et al. (2019)	n/a	n/a	73% no noise model (phantom)
Zhao et al. (2019)	n/a	n/a	Medical vessel model: 94%, Designed vessel model: 92%

BCA, brachiocephalic artery; DDPG, deep deterministic policy gradient; LCCA, left common carotid artery; LSA, left subclavian artery; RL, reinforcement learning.

intervention progress has been catalyzed by combining two recent approaches (LfD and RL; Chi et al., 2018a, 2020; Kweon et al., 2021; Cho et al., 2022). Here, using demonstrator data in a third of the RL studies allowed expert operator skill in complex endovascular procedures to be incorporated. This proficiency can be leveraged effectively to accelerate the RL training process. The combined approach, therefore, shortens the transition from a simulated training environment to a physical testing environment which typically presents significant challenges, as evidenced by the findings of Karstensen et al. (2022). Another benefit of accelerating the process is that in some scenarios thousands of mechanical experimental training cycles may no longer be required leading to reduced mechanical wear on the experimental equipment.

4.2.2. Limitations

The limitations of the studies assessed encompassed three areas:

(1) Whilst it was a strength that most studies focused on finding a ML technique that would improve upon previous endovascular navigation, the lack of focus on using similar or fixed algorithms and extending the experimental environment was a limitation. The challenge of fixing many experimental variables whilst changing another, is compounded by the lack of standardized *in silico*, *in vitro* or *ex vivo* experimental reference standard designs for endovascular navigation, as well as a lack of standardized performance measures. As such, the ability to compare studies quantitatively was limited by confounding. For example, although some performance measures (e.g., "success rate" and "procedure time") were common to several studies, study comparison was limited due to experimental variations between studies. Firstly, the navigation path used to test the models varied. Secondly, some studies defined "success rate" only if a task was completed within a certain time frame, whereas others had no time limit for completion. Thirdly, "procedure time" was measured using different starting points and target sites.

(2) Another limitation, also concerned with reference standards, is the importance of comparing the endovascular navigation with an autonomous system against the endovascular navigation without an autonomous system, to determine any incremental benefit through autonomy. Critically, the endovascular navigation without an autonomous system should ideally be operated by a relevant expert operating with minimal technical constraint to derive the reference standard (baseline) allowing comparison. Half the studies (7/14, 50%) reviewed did compare endovascular navigation with and without an autonomous system; however, in some cases, the operator was technically constrained by using a novel robotic system rather than using the equipment used and processes they would typically employ, during an endovascular procedure in the clinic. For instance, the reviewed robotic systems failed to replicate crucial haptic feedback experienced during manual procedures. These include viscous forces between catheters and blood, friction forces between catheters and the vessel wall, impact forces from catheter tips and guidewire, and contact with the vessel wall (Crinnion et al., 2022). Additionally, an expert is not able to use their previous experience with standard equipment and may be unfamiliar with these controls, meaning that performance at a given task will likely be affected.

(3) There were no clinical studies of autonomous endovascular navigation which is a reflection of the nascent field and current TRL of the technology. The majority of studies (11/14, 79%) were *in vitro* and are valuable for development and testing as they limit the number of failures during subsequent *in vivo* testing (Ionita et al., 2014). However, these studies did not consider whether construct, face, and context validity of endovascular navigation systems was acceptable to allow TRL progression toward the clinic. In particular, in many of the studies reviewed, there were translational concerns regarding how the guidewires and/or catheters are tracked within the vasculature, as the alternative to using fluoroscopy with standard off-the-shelf catheters and guidewires is to create entirely new tracking methods. For example, several papers (6/14, 43%) used EM-tracking to visualize the catheter in real-time, which has been shown to allow better real-time 3D orientation, facilitating navigation, reducing cannulation and total fluoroscopy times, and improving motion consistency and efficiency (Schwein et al., 2017). However, clinical translation using this method would require the introduction of new systems with specialized catheters and guidewires, resulting in additional costs and training. Furthermore, other studies (5/14, 36%) employed an experimental set-up involving a tabletop with a transparent phantom and a top-down camera. In its current state, this tracking method would not be suitable for future clinical studies, as a top-down camera would not be able to provide images of the guidewire and/or catheter through patient tissue. Nonetheless, it is noted that top-down cameras have a narrower clinical translation gap than EM-tracking, as they pose the same 2D challenges as fluoroscopy.

4.3. Final thoughts and future research

Using AI, it may be possible to create a robotic system capable of autonomously navigating catheters and wires through a patient's vasculature to the target site, requiring minimal assistance from an operator. If proven to be safe and effective in clinical trials, the benefits of autonomous navigation are numerous. It is plausible that in clinical specialities facing a shortage of highly-trained operators, there may be a reduced need for their expertise, potentially leading to greater accessibility of endovascular treatments globally, such as MT. For example, components of MT such as complex navigation tasks could be performed autonomously. Furthermore, autonomous systems are not limited by human factors such as fatigue or loss of focus, potentially making procedures safer and quicker (Mirnezami and Ahmed, 2018).

The concept of fully autonomous navigation in endovascular interventions is promising; however, with a TRL level of 3 (Mankins, 1995), the technology is yet to complete validation even in a laboratory environment. Due to the inadequate evidence supporting its use (the limited number of studies and its low-level;

Howick et al., 2011), it is far from being used in clinical practice. It first must be demonstrated that it can reliably provide benefits over currently available treatments before it can progress toward clinical trials.

Importantly, reference standards for endovascular navigation models need to be established to allow new models to be compared. This would allow effective comparison of different AI methods to determine the most effective model for autonomous endovascular navigation. These reference standards need to be established judiciously at the *in silico*, *in vitro*, and *ex vivo* level with carefully-defined environments for different endovascular tasks such as PCI, PVI, and MT. It is noteworthy that at the *in silico* level, where there are continuous advancements in modeling research and increased computational power, other areas of clinically-orientated ML research have successfully employed reference standards to enable reproducibility of results and comparability between competing models (Russakovsky et al., 2015; Stubbs et al., 2019). This includes computer vision (ImageNet Large Scale Visual Recognition Challenge) and natural language processing (National NLP Clinical Challenges). Furthermore, a set of minimum reporting standards of performance should be defined for studies investigating the use of AI in the autonomous navigation of endovascular interventions. In combination with a reference standard, this would allow complete comparison between ML algorithms designed for this specific task.

Clear regulation is required to determine how the community designs systems for the autonomous navigation in endovascular interventions. In the seven studies (7/14, 50%) which proposed a system with "level 3" autonomy, there is an expert operator in place who can intervene in the autonomous task if needed ("human in the loop"). At higher levels of autonomy where the robot can make decisions, particularly 'level 5' and potentially 'level 4', it is unclear how systems will be regulated. Therefore, it may be prudent, for now, for researchers to focus on optimizing systems with 'level 1–3' autonomy. As such future researchers may wish to optimize simple task autonomy, for example the autonomous navigation from the puncture point to the target site, in a system where an operator can stop the procedure and take over at any time. It is envisaged that as autonomous technology and regulations mature over time, systems will then be updated to carry out more difficult tasks.

Various AI methods have been used to investigate the possibility of autonomous navigation in endovascular interventions. Although it is plausible that autonomous navigation may eventually benefit patients while reducing occupational hazards for staff, there is currently no high-level evidence to support this assertion. For the technology to progress, reference standards and minimum reporting standards need to be established to allow meaningful comparisons of new system development.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

HR, AG, and TB contributed to the study conception and design. HR carried out the paper identification and screening. LK reviewed the manuscripts selected by HR against the eligibility criteria. The first draft of the manuscript was written by HR and edited by AG and TB. All authors commented on versions of the manuscript, read, and approved the manuscript.

Funding

Partial financial support was received from the WELLCOME TRUST under Grant Agreement No. 203148/A/16/Z and from the Engineering and Physical Sciences Research Council Doctoral Training Partnership (EPSRC DTP) under Grant Agreement No. EP/R513064/1.

Acknowledgments

The authors acknowledge support from the Department of Health and Social Care (DHSC) through the National Institute for Health and Care Research (NIHR) MedTech Co-operative award for Cardiovascular Diseases to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., et al. (2017). "Hindsight experience replay," in *Advances in Neural Information Processing Systems, Vol. 30* (Long Beach, CA).

Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A. (2017). A brief survey of deep reinforcement learning. *IEEE Signal Process. Mag.* doi: 10.1109/MSP.2017.2743240

Behr, T., Pusch, T. P., Siegfarth, M., Hüsener, D., Mörschel, T., and Karstensen, L. (2019). Deep reinforcement learning for the navigation of neurovascular catheters. *Curr. Direct. Biomed. Eng.* 5, 5–8. doi: 10.1515/cdbme-2019-0002

Brilakis, E. (2020). *Manual of Percutaneous Coronary Interventions, 1st Edn.* Elsevier.

Cancelliere, N. M., Lynch, J., Nicholson, P., Dobrocky, T., Swaminathan, S. K., Hendriks, E. J., et al. (2022). Robotic-assisted intracranial aneurysm treatment: 1 year follow-up imaging and clinical outcomes. *J. Neurointervent. Surg.* 14, 1229–1233. doi: 10.1136/neurintsurg-2021-017865

Chi, W., Dagnino, G., Kwok, T. M. Y., Nguyen, A., Kundrat, D., Abdelaziz, M. E. M. K., et al. (2020). "Collaborative robot-assisted endovascular catheterization with generative adversarial imitation learning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (Paris), 2414–2420. doi: 10.1109/ICRA40945.2020.9196912

Chi, W., Liu, J., Abdelaziz, M. E. M. K., Dagnino, G., Riga, C., Bicknell, C., et al. (2018a). "Trajectory optimization of robot-assisted endovascularcatheterization with reinforcement learning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid), 3875–3881. IEEE. doi: 10.1109/IROS.2018.8593421

Chi, W., Liu, J., Rafii-Tari, H., Riga, C., Bicknell, C., and Yang, G. Z. (2018b). Learning-based endovascular navigation through the use of non-rigid registration for collaborative robotic catheterization. *Int. J. Comput. Assist. Radiol. Surg.* 13, 855–864. doi: 10.1007/s11548-018-1743-5

Cho, Y., Park, J.-H., Choi, J., and Chang, D. E. (2022). "SIM-to-real transfer of image-based autonomous guidewire navigation trained by deep deterministic policy gradient with behavior cloning for fast learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Kyoto), 3468–3475. doi: 10.1109/IROS47612.2022.9982168

Codevilla, F., Santana, E., López, A. M., and Gaidon, A. (2019). "Exploring the limitations of behavior cloning for autonomous driving," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul), 9328–9337. doi: 10.1109/ICCV.2019.00942

Crinnion, W., Jackson, B., Sood, A., Lynch, J., Bergeles, C., Liu, H., et al. (2022). Robotics in neurointerventional surgery: a systematic review of the literature. *J. Neurointervent. Surg.* 14, 539–545. doi: 10.1136/neurintsurg-2021-018096

Dayan, P. (2017). *Unsupervised Learning*. Springer US. doi: 10.1016/B978-0-12-811654-8.00004-X

Duran, C., Lumsden, A. B., and Bismuth, J. (2014). A randomized, controlled animal trial demonstrating the feasibility and safety of the Magellan endovascular robotic system. *Ann. Vasc. Surg.* 28, 470–478. doi: 10.1016/j.avsg.2013.07.010

Fatima, M., and Pasha, M. (2017). Survey of machine learning algorithms for disease diagnostic. *J. Intell. Learn. Syst. Appl.* 9, 1–16. doi: 10.4236/jilsa.2017.91001

Faure, F., Duriez, C., Delingette, H., Allard, J., Gilles, B., Marchesseau, S., et al. (2012). *SOFA, a Multi-Model Framework for Interactive Physical Simulation*. Berlin; Heidelberg: Springer. doi: 10.1007/8415_2012_125

Ghesu, F. C., Georgescu, B., Grbic, S., Maier, A., Hornegger, J., and Comaniciu, D. (2018). Towards intelligent robust detection of anatomical structures in incomplete volumetric data. *Med. Image Anal.* 48, 203–213. doi: 10.1016/j.media.2018.06.007

Giacoppo, D., Colleran, R., Cassese, S., Frangieh, A. H., Wiebe, J., Joner, M., et al. (2017). Percutaneous coronary intervention vs. coronary artery bypass grafting in patients with left main coronary artery stenosis: a systematic review and meta-analysis. *JAMA Cardiol.* 2, 1079–1088. doi: 10.1001/jamacardio.2017.2895

Goyal, M., Menon, B. K., Zwam, W. H. V., Dippel, D. W., Mitchell, P. J., Demchuk, A. M., et al. (2016). Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet* 387, 1723–1731. doi: 10.1016/S0140-6736(16)00163-X

Hausegger, K. A., Schedlbauer, P., Deutschmann, H. A., and Tiesenhausen, K. (2001). Complications in endoluminal repair of abdominal aortic aneurysms. *Eur. J. Radiol.* 39, 22–33. doi: 10.1016/S0720-048X(01)00339-4

Hessel, M., Modayil, J., van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., et al. (2017). "Rainbow: combining improvements in deep reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32* (San Francisco, CA). doi: 10.1609/aaai.v32i1.11796

Ho, J., and Ermon, S. (2016). "Generative adversarial imitation learning," in *Advances in Neural Information Processing Systems*.

Ho, P., Cheng, S. W., Wu, P. M., Ting, A. C., Poon, J. T., Cheng, C. K., et al. (2007). Ionizing radiation absorption of vascular surgeons during endovascular procedures. *J. Vasc. Surg.* 46, 455–459. doi: 10.1016/j.jvs.2007.04.034

Howick, J., Chalmers, I., Lind, J., Glasziou, P., Greenhalgh, T., Heneghan, C., et al. (2011). *Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document)*. Available online at: http://www.cebm.net/index.aspx?o=5653

Ionita, C. N., Mokin, M., Varble, N., Bednarek, D. R., Xiang, J., Snyder, K. V., et al. (2014). "Challenges and limitations of patient-specific vascular phantom fabrication using 3d polyjet printing," in *Medical Imaging 2014: Biomedical Applications in Molecular, Structural, and Functional Imaging, Vol. 9038* (San Francisco, CA), 90380M. SPIE. doi: 10.1117/12.2042266

Jang, B., Kim, M., Harerimana, G., and Kim, J. W. (2019). Q-learning algorithms: a comprehensive classification and applications. *IEEE Access* 7, 133653–133667. doi: 10.1109/ACCESS.2019.2941229

Jones, B., Riga, C., Bicknell, C., and Hamady, M. (2021). Robot-assisted carotid artery stenting: a safety and feasibility study. *CardioVasc. Intervent. Radiol.* 44, 795–800. doi: 10.1007/s00270-020-02759-0

Karstensen, L., Ritter, J., Hatzl, J., Pätz, T., Langejürgen, J., Uhl, C., et al. (2022). Learning-based autonomous vascular guidewire navigation without human demonstration in the venous system of a porcine liver. *Int. J. Comput. Assist. Radiol. Surg.* 17, 2033–2040. doi: 10.1007/s11548-022-02646-8

Klein, L. W., Miller, D. L., Balter, S., Laskey, W., Haines, D., Norbash, A., et al. (2009). Occupational health hazards in the interventional laboratory: time for a safer environment. *Soc. Intervent. Radiol.* 250, 538–544. doi: 10.1148/radiol.2502082558

Kotsiantis, S. B. (2007). Supervised machine learning: a review of classification techniques. *Informatica* 31, 249–268.

Kweon, J., Kim, K., Lee, C., Kwon, H., Park, J., Song, K., et al. (2021). Deep reinforcement learning for guidewire navigation in coronary artery phantom. *IEEE Access* 9, 166409–166422. doi: 10.1109/ACCESS.2021.3135277

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., et al. (2016). "Continuous control with deep reinforcement learning," in *4th International Conference on Learning Representations*, (*ICLR) 2016*, eds Y. Bengio and Y. LeCun (San Juan).

Lindgren, A., Vergouwen, M. D., van der Schaaf, I., Algra, A., Wermer, M., Clarke, M. J., et al. (2018). Endovascular coiling versus neurosurgical clipping for people with aneurysmal subarachnoid haemorrhage. *Cochrane Database Syst. Rev.* 15:CD003085. doi: 10.1002/14651858.CD003085.pub3

Lowery, C., and Faisal, A. A. (2013). "Towards efficient personalized anaesthesia using continuous reinforcement learning for propofol infusion control," in *2013 6th International IEEE EMBS Conference on Neural Engineering (NER)* (San Diego, CA: Institute of Electrical and Electronics Engineers IEEE Engineering in Medicine and Biology Society), 1414–1417. doi: 10.1109/NER.2013.6696208

Madder, R. D., VanOosterhout, S., Mulder, A., Elmore, M., Campbell, J., Borgman, A., et al. (2017). Impact of robotics and a suspended lead suit on physician radiation exposure during percutaneous coronary intervention. *Cardiovasc. Revasc. Med.* 18, 190–196. doi: 10.1016/j.carrev.2016.12.011

Mankins, J. C. (1995). *Technology Readiness Level, a White Paper*. NASA, Office of Space Access and Technology, Advanced Concepts Office.

McMeekin, P., White, P., James, M. A., Price, C. I., Flynn, D., and Ford, G. A. (2017). Estimating the number of UK stroke patients eligible for endovascular thrombectomy. *Eur. Stroke J.* 2, 319–326. doi: 10.1177/2396987317733343

Meng, F., Guo, S., Zhou, W., and Chen, Z. (2021). "Evaluation of a reinforcement learning algorithm for vascular intervention surgery," in *2021 IEEE International Conference on Mechatronics and Automation, ICMA 2021* (Takamatsu: Institute of Electrical and Electronics Engineers Inc.), 1033–1037. doi: 10.1109/ICMA52036.2021.9512675

Meng, F., Guo, S., Zhou, W., and Chen, Z. (2022). "Evaluation of an autonomous navigation method for vascular interventional surgery in virtual environment," in *2022 IEEE International Conference on Mechatronics and Automation, ICMA 2022* (Columbus, OH: Institute of Electrical and Electronics Engineers Inc.), 1599–1604. doi: 10.1109/ICMA54519.2022.9856107

Mirnezami, R., and Ahmed, A. (2018). Surgery 3.0, artificial intelligence and the next-generation surgeon. *Br. J. Surg.* 105, 463–465. doi: 10.1002/bjs.10860

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., et al. (2016). "Asynchronous methods for deep reinforcement learning," in *Proceedings of the 33rd International Conference on Machine Learning - Volume 48* (New York, NY), 1928–1937.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). Playing atari with deep reinforcement learning. *CoRR, abs/1312.5602*. abs/1312.5602.

Mofatteh, M. (2021). Neurosurgery and artificial intelligence. *AIMS Neurosci.* 8, 477–495. doi: 10.3934/Neuroscience.2021025

Mongan, J., Moy, L., and Kahn, C. E. (2020). Checklist for artificial intelligence in medical imaging (claim): a guide for authors and reviewers. *Radiology* 2:e200029. doi: 10.1148/ryai.2020200029

Nair, A., McGrew, B., Andrychowicz, M., Zaremba, W., and Abbeel, P. (2017). "Overcoming exploration in reinforcement learning with demonstrations," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Singapore), 6292–6299. doi: 10.1109/ICRA.2018.8463162

Naros, G., and Gharabaghi, A. (2015). Reinforcement learning of self-regulated beta-oscillations for motor restoration in chronic stroke. *Front. Hum. Neurosci.* 9:391. doi: 10.3389/fnhum.2015.00391

Nogueira, R. G., Sachdeva, R., Al-Bayati, A. R., Mohammaden, M. H., Frankel, M. R., and Haussen, D. C. (2020). Robotic assisted carotid artery stenting for the treatment of symptomatic carotid disease: technical feasibility and preliminary results. *J. NeuroIntervent. Surg.* 12, 341–344. doi: 10.1136/neurintsurg-2019-015754

Nussbaumer-Streit, B., Klerings, I., Dobrescu, A. I., Persad, E., Stevens, A., Garritty, C., et al. (2020). Excluding non-english publications from evidence-syntheses did not change conclusions: a meta-epidemiological study. *J. Clin. Epidemiol.* 118, 42–54. doi: 10.1016/j.jclinepi.2019.10.011

O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. *ArXiv, abs/1511.08458*.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021:372. doi: 10.31222/osf.io/v7gm2

Pereira, M., Cancelliere, N., Nicholson, P., Radovanovic, I., Drake, K. E., Sungur, J.-M., et al. (2020). First-in-human, robotic-assisted neuroendovascular intervention. *J. NeuroIntervent. Surg.* 12, 338–340. doi: 10.1136/neurintsurg-2019-015671.rep

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286. doi: 10.1109/5.18626

Rafii-Tari, H., Liu, J., Lee, S.-L., Bicknell, C., and Yang, G.-Z. (2013). "Learning-based modeling of endovascular navigation for collaborative robotic catheterization," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013* (Berlin; Heidelberg: Springer), 369–377. doi: 10.1007/978-3-642-40763-5_46

Rafii-Tari, H., Liu, J., Payne, C. J., Bicknell, C., and Yang, G.-Z. (2014). "Hierarchical hmm based learning of navigation primitives for cooperative robotic endovascular catheterization," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2014* (Japan: Nagoya University; Springer International Publishing), 496–503. doi: 10.1007/978-3-319-10404-1_62

Redmon, J., Divvala, S. K., Girshick, R. B., and Farhadi, A. (2015). "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV), 779–788. doi: 10.1109/CVPR.2016.91

Reynolds, D. (2015). "Gaussian Mixture Models", in *Encyclopedia of Biometrics*, eds S. Z. Li and A. K. Jain (Boston, MA: Springer). doi: 10.1007/978-1-4899-7488-4_196

Riga, C. V., Cheshire, N. J., Hamady, M. S., and Bicknell, C. D. (2010). The role of robotic endovascular catheters in fenestrated stent grafting. *J. Vasc. Surg.* 51, 810–820. doi: 10.1016/j.jvs.2009.08.101

Rudnick, M. R., Goldfarb, S., Wexler, L., Ludbrook, P. A., Murphy, M. J., Halpern, E. F., et al. (1995). Nephrotoxicity of ionic and nonionic contrast media in 1196 patients: a randomized trial. *Kidney Int.* 47, 254–261. doi: 10.1038/ki.1995.32

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Rutjes, A., Westwood, M., Deeks, J. J., Whiting, P. F., Weswood, M. E., Rutjes, A. W., et al. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies evaluation of quadas, a tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* 155, 529–536. doi: 10.7326/0003-4819-155-8-201110180-00009

Saber, H., Beaman, C., and Tateshima, S. (2022). Complete robotic intervention for acute epistaxis in a patient with COVID-19 pneumonia: technical considerations and device selection tips. *J. NeuroIntervent. Surg.* 14, 500–502. doi: 10.1136/neurintsurg-2021-018582

Sajja, K. C., Sweid, A., Saiegh, F. A., Chalouhi, N., Avery, M. B., Schmidt, R. F., et al. (2020). Endovascular robotic: feasibility and proof of principle for diagnostic cerebral angiography and carotid artery stenting. *J. NeuroIntervent. Surg.* 12, 345–349. doi: 10.1136/neurintsurg-2019-015763

Sarker, I. H. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Comput. Sci.* 2:160. doi: 10.1007/s42979-021-00592-x

Saver, J. L., Goyal, M., Lugt, A. V. D., Menon, B. K., Majoie, C. B., Dippel, D. W., et al. (2016). Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *J. Am. Med. Assoc.* 316, 1279–1288. doi: 10.1001/jama.2016.13647

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv: abs/1707.06347*.

Schwein, A., Kramer, B., Chinnadurai, P., Walker, S., O'Malley, M., Lumsden, A., and Bismuth, J. (2017). Flexible robotics with electromagnetic tracking improves safety and efficiency during *in vitro* endovascular navigation. *J. Vasc. Surg.* 65, 530–537. doi: 10.1016/j.jvs.2016.01.045

Silahtaroğlu, G., and Yılmaztürk, N. (2021). Data analysis in health and big data: a machine learning medical diagnosis model based on patients' complaints. *Commun. Stat.* 50, 1547–1556. doi: 10.1080/03610926.2019.1622728

Stubbs, A., Filannino, M., Soysal, E., Henry, S., and Özlem Uzuner (2019). Cohort selection for clinical trials: N2c2 2018 shared task track 1. *J. Am. Med. Inform. Assoc.* 26, 1163–1171. doi: 10.1093/jamia/ocz163

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction, 2nd Edn*. Cambridge: The MIT Press.

Theodorou, E. A., Buchli, J., and Schaal, S. (2010). A generalized path integral control approach to reinforcement learning. *J. Mach. Learn. Res.* 11, 3137–3181. doi: 10.1109/ROBOT.2010.5509336

Thukkani, A. K., and Kinlay, S. (2015). Endovascular intervention for peripheral artery disease. *Circul. Res.* 116, 1599–1613. doi: 10.1161/CIRCRESAHA.116.303503

Townsend, N., Wilson, L., Bhatnagar, P., Wickramasinghe, K., Rayner, M., and Nichols, M. (2016). Cardiovascular disease in Europe: epidemiological update 2016. *Eur. Heart J.* 37, 3232–3245. doi: 10.1093/eurheartj/ehw334

Wang, S., Liu, Z., Shu, X., Cao, Y., Zhang, L., and Xie, L. (2022). "Study on autonomous delivery of guidewire based on improved yolov5s on vascular model platform," in *2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (Xishuangbanna), 1–6. doi: 10.1109/ROBIO55434.2022.10011829

Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., and de Freitas, N. (2016). "Dueling network architectures for deep reinforcement learning," in *Proceedings of the 33rd International Conference on Machine Learning - Volume 48* (New York, NY), 1995–2003.

Weinberg, J. H., Sweid, A., Sajja, K., Gooch, M. R., Herial, N., Tjoumakaris, S., et al. (2021). Comparison of robotic-assisted carotid stenting and manual carotid stenting through the transradial approach. *J. Neurosurg.* 135, 21–28. doi: 10.3171/2020.5.JNS201421

Yang, G. Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P. E., et al. (2017). Medical robotics-regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci. Robot.* 2:eaam8638. doi: 10.1126/scirobotics.aam8638

You, H., Bae, E. K., Moon, Y., Kweon, J., and Choi, J. (2019). Automatic control of cardiac ablation catheter with deep reinforcement learning method. *J. Mech. Sci. Technol.* 33, 5415–5423. doi: 10.1007/s12206-019-1036-0

Zhao, Y., Guo, S., Wang, Y., Cui, J., Ma, Y., Zeng, Y., et al. (2019). A CNN-based prototype method of unstructured surgical state perception and navigation for an endovascular surgery robot. *Med. Biol. Eng. Comput.* 57, 1875–1887. doi: 10.1007/s11517-019-02002-0

frontiers | Frontiers in Surgery

TYPE Systematic Review
PUBLISHED 21 August 2023
DOI 10.3389/fsurg.2023.1245851

OPEN ACCESS

EDITED BY
Mohammad Mofatteh,
Queen's University Belfast, United Kingdom

REVIEWED BY
Mark Ter Laan,
Radboud University Medical Centre, Netherlands
Sabino Luzzi,
University of Pavia, Italy

*CORRESPONDENCE
Anya Ragnhildstveit
✉ amr210@cam.ac.uk

†These authors have contributed equally to this work and share senior authorship

RECEIVED 23 June 2023
ACCEPTED 04 August 2023
PUBLISHED 21 August 2023

CITATION
Ragnhildstveit A, Li C, Zimmerman MH, Mamalakis M, Curry VN, Holle W, Baig N, Uğuralp AK, Alkhani L, Oğuz-Uğuralp Z, Romero-Garcia R and Suckling J (2023) Intra-operative applications of augmented reality in glioma surgery: a systematic review. Front. Surg. 10:1245851. doi: 10.3389/fsurg.2023.1245851

Intra-operative applications of augmented reality in glioma surgery: a systematic review

Anya Ragnhildstveit[1,2]*, Chao Li[3,4], Mackenzie H. Zimmerman[1], Michail Mamalakis[2], Victoria N. Curry[1,5], Willis Holle[1,6], Noor Baig[1,7], Ahmet K. Uğuralp[1], Layth Alkhani[1,8], Zeliha Oğuz-Uğuralp[1], Rafael Romero-Garcia[2,9†] and John Suckling[2†]

[1]Integrated Research Literacy Group, Draper, UT, United States, [2]Department of Psychiatry, University of Cambridge, Cambridge, England, [3]Department of Clinical Neurosciences, University of Cambridge, Cambridge, England, [4]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England, [5]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States, [6]Department of Physics and Astronomy, The University of Utah, Salt Lake City, UT, United States, [7]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, United States, [8]Department of Biology, Stanford University, Stanford, CA, United States, [9]Instituto de Biomedicina de Sevilla (IBiS) HUVR/CSIC/Universidad de Sevilla/CIBERSAM, ISCIII, Dpto. de Fisiología Médica y Biofísica

Background: Augmented reality (AR) is increasingly being explored in neurosurgical practice. By visualizing patient-specific, three-dimensional (3D) models in real time, surgeons can improve their spatial understanding of complex anatomy and pathology, thereby optimizing intra-operative navigation, localization, and resection. Here, we aimed to capture applications of AR in glioma surgery, their current status and future potential.

Methods: A systematic review of the literature was conducted. This adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline. PubMed, Embase, and Scopus electronic databases were queried from inception to October 10, 2022. Leveraging the Population, Intervention, Comparison, Outcomes, and Study design (PICOS) framework, study eligibility was evaluated in the qualitative synthesis. Data regarding AR workflow, surgical application, and associated outcomes were then extracted. The quality of evidence was additionally examined, using hierarchical classes of evidence in neurosurgery.

Results: The search returned 77 articles. Forty were subject to title and abstract screening, while 25 proceeded to full text screening. Of these, 22 articles met eligibility criteria and were included in the final review. During abstraction, studies were classified as "development" or "intervention" based on primary aims. Overall, AR was qualitatively advantageous, due to enhanced visualization of gliomas and critical structures, frequently aiding in maximal safe resection. Non-rigid applications were also useful in disclosing and compensating for intra-operative brain shift. Irrespective, there was high variance in registration methods and measurements, which considerably impacted projection accuracy. Most studies were of low-level evidence, yielding heterogeneous results.

Conclusions: AR has increasing potential for glioma surgery, with capacity to positively influence the onco-functional balance. However, technical and design limitations are readily apparent. The field must consider the importance of consistency and replicability, as well as the level of evidence, to effectively converge on standard approaches that maximize patient benefit.

KEYWORDS

augmented reality, brain tumor, glioma, mixed reality, neuronavigation, neurosurgery, systematic review, virtual reality

Introduction

Gliomas account for 78% of primary malignant brain tumors (1). They originate from glial progenitor cells, namely astrocytes or oligodendrocytes, that constitute a significant portion of the mammalian brain. As such, gliomas are highly heterogeneous, known for their diverse histopathology, molecular genetics, and clinical behavior. In the United States, the incidence of gliomas varies from 4.7 to 5.7 per 100,000 persons (2), representing more than 18,500 new cases and 13,000 deaths annually (3). Glioblastoma multiforme (GBM), the most common and aggressive form of glioma, has a median survival of 16 months (4), carrying a five-year post-diagnosis survival rate of 6.8% (5). While the pathogenesis differs considerably, low-grade glioma has a more favorable timespan, ranging from 5.6 to 13.3 years, depending on several prognostic factors (6). Nonetheless, 70% of these tumors transform to GBM within 10 years (7). This eventually causes disability and premature death (8, 9).

Currently, the primary care pathway for gliomas is surgical resection followed by chemoradiotherapy, with concomitant temozolomide or other alkylating drugs (10, 11). Maximizing the extent of resection (EOR), until functional borders are encountered, is central to prolonging survival, improving the efficacy of adjuvant therapies, and delaying anaplastic transformation in both low- and high-grade glioma (12–15). The rationale for performing this type of "supratotal" resection is based on evidence that gliomas infiltrate the parenchyma well beyond magnetic resonance imaging (MRI)-defined abnormalities (16). Tumor recurrence may thus arise from undetected glioma cells growing beyond signal abnormalities, typically found 1–2 cm outside of contrast enhancement, as detected on volumetric fluid-attenuated inversion recovery (FLAIR) images. However, supratotal resection is not always practical or feasible to achieve.

Diffusely infiltrating gliomas often limit radical resection strategies, which preferentially invade along myelinated fibers in white matter tracts (17); cluster in eloquent brain regions with dense functional connections, like the basal ganglia and internal capsule (18); and develop functional multi-cellular network structures (19). Therefore, surgically-acquired lesions in functionally critical areas may cause significant neurologic morbidity and mortality (20, 21). Neural plasticity is another barrier to radical resection, due to functional reorganization (22, 23). Injury to white matter tracts, dynamically interacting with gliomas, is linked to post-operative deficits, accordingly (24, 25). Hence, the true benefit of resection depends on the "onco-functional balance": (26) maximizing the extent of tumor removal while preserving patients' functional integrity and quality of life.

Augmented reality (AR) is a technology that superimposes computer-generated, three-dimensional (3D) holograms, as well as auditory and sensory feedback, on reality in real time and space. This composite view of virtual objects with the real world creates a semi-immersive environment. Dating back to 1986 (27), AR has been applied in neurosurgery for nearly 30 years, carrying several advantages over conventional approaches (28).

First, AR maps patient-specific neuroanatomy directly onto the operating field, rendering display of surface and sub-surface targets. This has proven useful in visualizing anatomical structures, vasculature and hemodynamics, and deep-seated lesions in stereotactic, neurovascular, and tumor surgery, respectively (29). It also allows surgeons to access and contextualize radiological images and pre-operative planning. Second, AR eliminates attentional shifts between patients on operating tables and screens displaying relevant clinical information. This can reduce fatigue, cognitive load, and inattention blindness among surgeons, leading to more focused and efficient procedures (30). Third and finally, AR may disclose and compensate for intra-operative brain shift (31, 32): a highly prevalent and complex phenomenon of brain deformation due to changes in gravity and hydrostatic pressure, loss of cerebrospinal fluid, tissue manipulation or removal, and other factors (33). For image- and function-guided neurosurgery, this can invalidate patient-to-image registration and reduce the accuracy of localizing and resecting intra-cranial targets, as well as positioning surgical tools (34, 35). Thus, AR can be used to update virtual scenes, when combined with multimodal imaging and functional testing, to precisely identify pathologies, probe subcortical pathways, and tailor resection plans (28, 36, 37).

To date, applications of AR in neurosurgery have been limited to early clinical research. Given this stage, a variety of AR devices have been used for image projection, including head-up displays (HUDs), head-mounted displays (HMDs), microscopes, endoscopes, smartphones, and tablets (29, 37–40). Commercial display devices have also emerged, such as Google Glass (Google LLC, Mountain View, California, USA), HoloLens (Microsoft Corp., Redmond, Washington, USA), and Magic Leap (Magic Leap Inc., Plantation, Florida, USA) (29). Such innovation underscores the growing clinical and commercial interest in AR for neurosurgical practice, with proposed roles in skin incision, craniotomy, and resection (37, 38, 40). This is of particular import for eloquent brain tumors, as serious threats to human life and health, whereby AR may positively influence maximal safe resection and functional outcomes. In this review, we summarize current applications of AR in glioma surgery, as described in the scientific literature, with the aim of characterizing emerging trends and providing avenues for future research.

Methods

We performed an in-depth systematic review, adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline (41). The review protocol was registered *a priori* with the Open Science Framework (OSF) (42), developed and maintained by the Center for Open Science (COS), which can be accessed via the digital object identifier (DOI): 10.17605/OSF.IO/DJ72P. The PRISMA 2020 Checklist (41), review strategy, and review protocol are additionally available for consultation upon reasonable request.

PubMed (National Library of Medicine), Embase (Elsevier), and Scopus (Elsevier) electronic databases were queried from inception to October 10, 2022, for relevant articles. The following

search strategy was employed: PubMed: (("augmented reality"[All Fields] OR "mixed reality"[All Fields]) AND ("glioma"[MeSH Terms] OR "glioma"[All Fields] OR "gliomas"[All Fields] OR "glioma s"[All Fields])); Embase: (("augmented reality"/exp OR "augmented reality" OR "mixed reality"/exp OR "mixed reality") AND ("glioma"/exp OR "glioma") AND [article]/lim AND [humans]/lim); and Scopus: TITLE-ABS-KEY ((("augmented reality" OR "mixed reality") AND (glioma))). No publication date or study type restrictions were applied.

Inclusion criteria consisted of (1) phantoms or patients of any age and biological sex diagnosed with glioma; (2) AR developed for or applied in glioma surgery, specifically to aid intra-operative navigation, localization, and/or resection; (3) protocol or technical note papers; case-control studies, case series, or case reports; retrospective, prospective, or concurrent cohort studies; or non-randomized, randomized, or post-hoc analyses of clinical trials; and (4) peer-reviewed studies published in the English language. In contrast, exclusion criteria comprised (1) phantoms or patients without glioma; (2) studies not developing or applying AR for/in glioma surgery, such as for patient education and surgical planning purposes; (3) reviews, editorials, expert opinion pieces, commentaries, letters to the editor, and articles with inaccessible full texts; and (4) studies not peer-reviewed and published in the English language. Duplicates were excluded prior to screening and studies that failed to meet full inclusion criteria were excluded from the overall analysis. The "Population, Intervention, Comparison, Outcomes, and Study design" (PICOS) (43) framework was applied for evaluating eligibility criteria in the qualitative synthesis.

Two independent reviewers, one with and one without prior content knowledge (N.B., M.H.Z.), screened articles against PICOS criteria, initially evaluating their titles and abstracts. Relevant studies were then selected for full text screening and assessed for eligibility. Inter-rater agreement was reported (Cohen's $k = 0.74$), with disagreements reconciled through discussion and/or by involvement of a third independent reviewer (A.R.) until a consensus was reached.

Another set of independent reviewers (W.H., A.K.U.) subsequently extracted data from eligible studies into a Microsoft Excel Spreadsheet (Microsoft Corp., Redmond, Washington, USA). Table cells were labeled as "Not applicable" (N/A) if data were missing. To ensure global data integrity, an independent reviewer (Z.O.U.) performed quality assurance checks at random. Data extracted from eligible studies included: study year, study location, study design, study type, number of total patients, number of glioma patients, number of phantom patients, glioma pathology, other pathologies, image acquisition phase, image data source, image segmentation technique, geometric modeling software, registration method, registration accuracy, display device, display brand, clinical application, primary outcomes, and any other pertinent findings. Following extraction, data were qualitatively described, using frequency (count, percentage), central tendency (mean, median, mode), and variability (range, standard deviation), as applicable, via R version 4.1.3 (44). Pooled statistical analyses, such as meta-regressions, were not performed due to heterogeneity in study designs and measured outcomes.

To assess the quality of evidence, a risk of bias assessment was conducted, using hierarchical classes of evidence in neurosurgery (45). This involved ranking the methodological rigor of each study, whereby "Level V" indicates the lowest or weakest level of evidence, such as case reports, and "Level I" indicates the highest or strongest level of evidence, such as randomized trials. As the study design becomes more rigorous, the quality of evidence increases, and the probability of bias decreases. Two independent reviewers (V.N.C., L.A.) conducted the risk of bias assessment, with inter-rater agreement reported (Cohen's $k = 0.82$).

Results

The initial search query returned 77 articles for potential inclusion in the review [PubMed ($n = 21$), Embase ($n = 25$), and Scopus ($n = 31$)]. Of these, 40 (52%) unique articles remained following removal of duplicates (35, 46%) and inaccessible texts (2, 3%). After title and abstract screening, 15 (38%) articles were excluded for being literature reviews (7, 18%), investigating non-glioma tumors (1, 3%), not involving AR (4, 10%), or falling out of scope with PICOS criteria (3, 8%). Accordingly, 25 (63%) full text articles were assessed for eligibility, with 3 (12%) deemed non-eligible for inclusion.

The final review comprised 22 articles, with the first study published in 2003 ($n = 1$), the largest number published in 2021 ($n = 6$), and the most recent published in 2022 ($n = 4$). See studies by publication year in **Figure 1**. Regarding location, the majority of studies were conducted in Asia (13, 59%) followed by Europe (7, 32%) and North America (2, 9%). The PRISMA 2020 flow diagram, describing the search strategy and selection schema, is displayed in **Figure 2**. Characteristics of the studies included in the review are summarized in **Tables 1–3**.

At the time of data abstraction, based on primary aims, studies were sub-grouped into the following categories: "*development*" and "*intervention*". Ten (45%) studies (46, 48–50, 53, 54, 57–59, 67) evaluated the technical design and suitability of AR for glioma surgery, principally assessing feasibility, accuracy, and/or reliability benchmarks (i.e., development). The remaining 12 (55%) studies (47, 51, 52, 55, 56, 60–66) investigated the clinical utility of AR for glioma surgery, with a focus on feasibility, safety, and/or efficacy profiles (i.e., intervention). Across studies, there was a total of 909 patients (41.3 ± 54.1), of which 488 were diagnosed with gliomas (22.2 ± 35.7). Other pathologies included meningioma, lymphoma, angioma, papilloma, craniopharyngioma, hemangioblastoma, and arachnoid cysts, among others. **Figure 3A–B** illustrates the number of glioma patients and imaging acquisition protocol across study types. Three studies (51, 59, 67) additionally leveraged phantoms ($n = 22$, 1.0 ± 4.3). Steps for applying AR models in glioma surgery comprised: (1) image acquisition, (2) image segmentation, (3) geometric model generation, (4) registration and tracking, and (5) intra-operative navigation via fused image overlay.

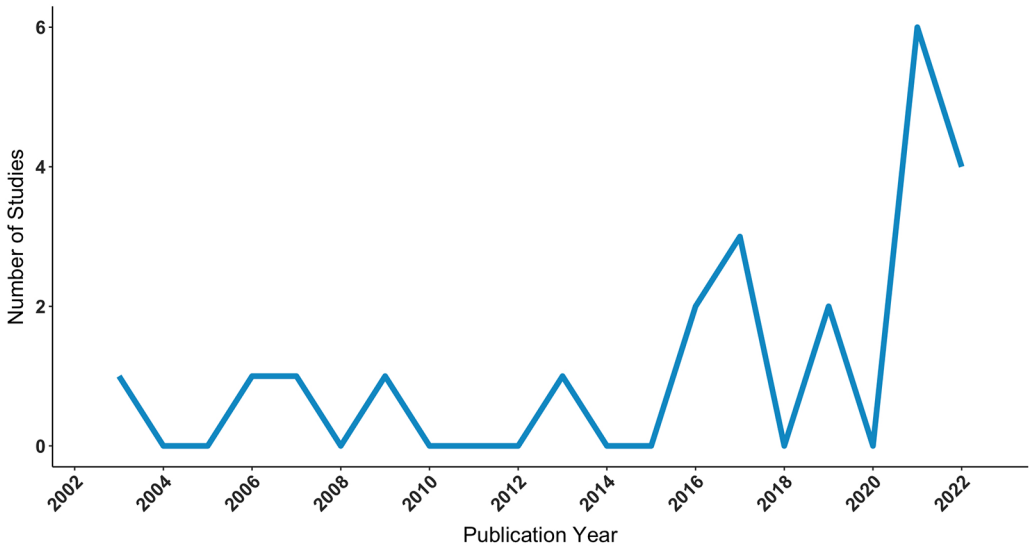


FIGURE 1
Number of studies applying AR in glioma surgery by publication year, as identified via PubMed, Embase, and Scopus electronic database searches, executed on October 22, 2022. AR, augmented reality.

1. Image acquisition

Data for AR were obtained from a variety of medical imaging sources that occasionally integrated with network analysis and brain mapping. All studies performed anatomical or volumetric magnetic resonance imaging (MRI: 22, 100%), followed by computerized tomography (CT: 16, 73%), diffusion weighted or tensor imaging (DWI/DTI: 13, 59%), functional magnetic resonance imaging (fMRI: 4, 20%), computerized tomography angiography (CTA: 3, 14%), 3D rotational angiography (3DRA: 3, 14%), and magnetic resonance spectroscopy (MR spectroscopy: 2, 9%). Across image acquisitions, pertinent factors included spatial resolution, slice thickness, signal- and contrast-to-noise ratios, and image artifact.

DTI-based tractography, ultrasound, and navigated transcranial magnetic stimulation (nTMS) were additionally carried out as complimentary techniques in 8 (36%), 1 (5%), and 1 (5%) study, respectively. Of those involving DTI-based tractography, two studies (61, 62) performed high-definition fiber tractography with sodium fluorescein (HDFT-F), motivated by the prospect of increasing tumor resectability as well as survival rates. In another study (52), ultrasound images were obtained prior to corticectomy and following resection, with the aim of delineating lesion borders and post-resection cavity. See **Figure 4** for the distribution of imaging sources used for AR. Most images were obtained pre-operatively (12, 55%), with 2 (9%) studies obtaining images intra-operatively and 8 (36%) from both phases (**Figure 3B**).

2. Image segmentation

Imaging data was partitioned, or "segmented", into anatomical regions of interest, removing unnecessary and irrelevant information, commonly exported as Digital Imaging and Communications in Medicine (DICOM) images. This included target and adjacent structures, namely tumors and surrounding blood vessels, nerves, and other tissues; in addition to cortical and sub-cortical areas, such as the postcentral gyrus and corticospinal tract. In general, segmentation techniques involved thresholding, edge pixel detection, and region growing. Liao et al. (59) specifically developed a rapid, autostereoscopic segmentation method, using fuzzy connectedness for open MRI-guided glioma surgery.

3. Model generation

Three-dimensional modeling was used to convert segmentations to virtual objects, most frequently with 3D Slicer (7, 32%) and BrainLab (5, 23%) visualization software, based on DICOM images. To achieve this, tumor and cortical surface meshes, for instance, were exported as files suitable for 3D printing and computer-aided design (CAD), generally in stereolithrography (STL) file format. Unique to studies, Ghimire et al. (52) transformed positive motor responses, acquired from pre-operative nTMS, in 3D objects projected onto a tractography model of the corticospinal tract.

4. Registration and tracking

Registration was performed to format, align, and superimpose virtual objects—including drilling axis and cutting planes—onto patients' real anatomy. For studies using neuronavigation, this typically involved registering the patient to the system and co-registering the display device (e.g., surgical microscope) to determine the necessary transformation. In manual registration, virtual objects were scaled, translated, and/or rotated in relation to patients' head or brain, by the user, based on anatomical landmarks and fiducial markers. While this was the simplest approach, it was also the most time intensive and susceptible to human error, requiring continuous interaction between surgeons and technicians to update the AR scene. In automatic registration, landmarks and fiducials were often used as starting

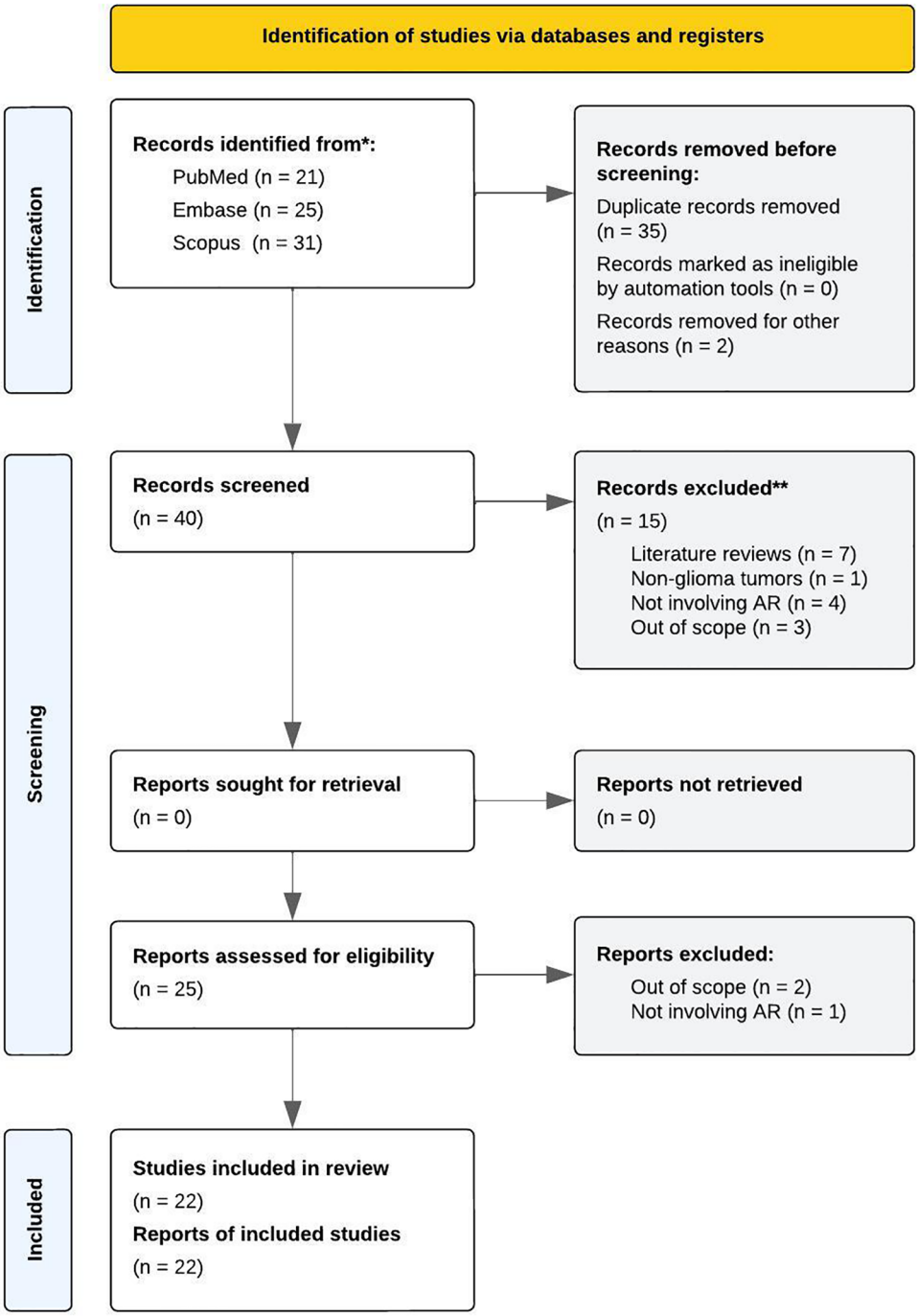


FIGURE 2
PRISMA 2020 flow diagram, describing the search strategy and selection schema of the review process.

points with further processing and no user interaction. This approach was more expeditious, though relied on the quality of machine learning methods (i.e., training datasets). Registration also involved rigid, non-rigid, or hybrid surface interactions. Here, an assumption was made whether virtual and real objects —to be aligned—had the same shape, relating them by a single or multiple rigid transformations.

Given the variety and inconsistency in registration techniques, fused image overlay greatly varied across studies. Multiple terms were also used to express positional accuracy, or the estimate of error, as an indication of the system's ability to guide surgical targets. These terms included the target registration error (TRE), fiducial registration error (FRE), root-mean-squared error (RMSE), and target deviation (D). The TRE and FRE were the most widely calculated. Other factors contributing to variation comprised: geometric and optical distortions, such as incorrect tracking and display abnormalities; "swimming" effects, like bed movement and brain shift; and glioma presentation. Mascitelli et al. (63), for example, reported greater accuracy for superficial lesions compared to deep-seated ones

TABLE 1 Characteristics of included studies.

Sample size	*n* (m)
Glioma patients	488 (22.2)
Phantom patients	22 (1.0)
Study type	*n* (%)
Development	10 (45.5)
Intervention	12 (54.5)
Imaging source	*n* (%)
3DRA	3 (13.6)
CT	16 (72.7)
CTA	3 (13.6)
DWI/DTI	10 (45.5)
fMRI	4 (18.2)
MRI	22 (100.0)
MR spectroscopy	2 (9.1)
Imaging phase	*n* (%)
Pre-operative	12 (54.5)
Intra-operative	2 (9.1)
Both stages	8 (36.4)
Modeling software	*n* (%)
3D Slicer	7 (31.8)
Amira	2 (9.1)
Avizo lite	3 (13.6)
BrainLab	5 (22.7)
Other	5 (22.7)
Not reported	1 (4.5)
Display Device	*n* (%)
Camera	5 (22.7)
Endoscope	1 (4.5)
HMD	3 (13.6)
HUD[a]	2 (9.1)
Microscope	7 (31.8)
Smartphone	2 (9.1)
Tablet	2 (9.1)
Video	1 (4.5)
Not reported	1 (4.5)
Study location	*n* (%)
China	5 (22.7)
England	1 (4.5)
Germany	3 (13.6)
Italy	2 (9.1)
Japan	8 (36.4)
Switzerland	1 (4.5)
United States	2 (9.1)
Levels of evidence[b]	*n* (%)
Level II	1 (4.5)
Level III	2 (9.1)
Level IV	13 (59.1)
Level V	6 (27.3)

3DRA, three-dimensional rotational angiography; CT, computerized tomography; CTA, computerized tomography angiography; DTI, diffusion tensor imaging, DWI, diffusion-weighted imaging; fMRI, functional magnetic resonance imaging; HMD, head-mounted display; HUD, head-up display; MRI, magnetic resonance; MRI, magnetic resonance imaging.
[a]HUDs were used as integration tools for AR visualization in surgical microscopes.
[b]Levels of evidence used by neurosurgeons in clinical practice.

(88.0% vs. 64.4%, $p = .029$). They also disabled HUDs in 59.6% of cases due to lack of use, distraction, and inaccuracy. Considering the aforementioned issues, several studies developed novel registration methods.

Fick et al. (50) designed a custom reference array, as an adjunct to HMDs, to correct initial registration for bed movements in GBM. While their technical workflow functioned as desired, and improved spatial understanding for surgeons, their registration accuracy was sub-optimal for clinical use (FRE = 8.55 mm). Relatedly, Archip et al. (46) evaluated a volumetric, non-rigid registration scheme to compensate for intra-operative brain shift. In 11 patients with eloquent supratentorial glioma, they revealed significant improvement in alignment accuracy compared to rigid-based, state-of-the-art technology ($p < 0.001$), with a mean residual displacement of D = 1.82 mm. Another study (67) leveraged a markerless spatial drift registration method to precisely align real and virtual objects. Pre-operatively, this aided surgeons in diagnosis and surgical planning; whereas intra-operatively, it helped them distinguish lesion boundaries and localize nerves, thereby increasing accurate resection of gliomas (RMSE = 1.86 mm). Liao et al. (59) similarly developed a spatial image registration method for integral videography image overlay, demonstrating satisfactory accuracy (TRE = 0.90 ± 0.21 mm). Nonetheless, there was high heterogeneity in registration methods and measurements, with no standard criteria for defining nor evaluating accuracy. Others failed to measure and/or report accuracy altogether (49, 52, 61–63, 65, 66).

5. Fused image overlay

Devices used to display virtual objects on the skull or dura of phantoms and patients included surgical microscopes (7, 32%), cameras (5, 23%), HMDs (3, 14%), tablets (2, 9%), smartphones (2, 9%), videos (1, 5%), and endoscopes (1, 5%). Of the studies utilizing surgical microscopes, two integrated AR via HUD (47, 63). Display device was unspecified for one study (46), which broadly referenced surgical instruments. See display devices by study design in **Figure 5**.

Augmented reality applications

Iseki et al. (55) were among the first to evaluate AR in tumor surgery. An analysis of 42 patients with malignant gliomas, located in or adjacent to functional regions, showed markedly increased EOR (≥90%) when open MRI was simultaneously applied with real-time update navigation, which continuously refreshed intra-operative images. Using a similar concomitant method, Sun et al. (66) achieved complete resection in 69.6% of glioma patients ($n = 79$) compared to 36.4% of control patients ($n = 55$), with an average EOR of 95.2% ± 8.5% and 84.9% ± 15.7%, respectively ($p < 0.01$). The rates of post-operative recovery in motor, visual, and language function were also higher in the study group at two weeks and three months ($p < 0.05$). In both studies, intra-operative MRI disclosed and corrected for brain shift, providing surgeons with accurate and objective information as well as quality control during procedures.

A retrospective study (63) also showed potential for AR in intra-cranial surgery, detailing early experience with HUDs. For superficial and intra-axial lesions, HUD provided greater utility for skin incision, craniotomy, dural opening, and corticectomy,

TABLE 2 Design of included studies.

Author	Approach	Design[a]	Type	Sample	Glioma	Model[b]	Pathology
Archip et al. (46)	Prospective	Case Series	Development	$n = 11$	$n = 11$	$n = 0$	LGG, HGG
Carl et al. (47)	Prospective	Case Series	Intervention	$n = 10$	$n = 2$	$n = 0$	E, LGG
Chen et al. (48)	Prospective	Case Report	Development	$n = 16$	$n = 1$	$n = 0$	Glioma
De Mauro et al. (49)	Technical	Technical Note	Development	$n = 0$	$n = 0$	$n = 0$	LGG
Fick et al. (50)	Prospective	Case Report	Development	$n = 3$	$n = 1$	$n = 0$	GBM
Finger et al. (51)	Retrospective	Case Series	Intervention	$n = 28$	$n = 6$	$n = 1$	AA, DA, GG, PA
Ghimire et al. (52)	Retrospective	Case Series	Intervention	$n = 180$	$n = 145$	$n = 0$	LGG, HGG
Hou et al. (53)	Prospective	Case Series	Development	$n = 35$	$n = 6$	$n = 0$	DA, GBM, GG
Inoue et al. (54)	Prospective	Case Report	Development	$n = 3$	$n = 1$	$n = 0$	GBM
Iseki et al. (55)	Prospective	Case Series	Intervention	$n = 148$	$n = 72$	$n = 0$	Glioma
Koike et al. (56)	Prospective	Case Series	Intervention	$n = 18$	$n = 18$	$n = 0$	AA, AO, DA, GBM, OD
Koike et al. (57)	Prospective	Case Series	Development	$n = 16$	$n = 16$	$n = 0$	A, AOA, DA, GBM, OA, OD
Koike et al. (58)	Prospective	Case Series	Development	$n = 15$	$n = 14$	$n = 0$	Glioma
Liao et al. (59)	Prospective	Case Report	Development	$n = 2$	$n = 1$	$n = 1$	Glioma
Liu et al. (60)	Retrospective	Case-Control Study	Intervention	$n = 53$	$n = 30$	$n = 0$	A, GBM
Luzzi et al. (61)	Prospective	Case Report	Intervention	$n = 1$	$n = 1$	$n = 0$	GBM
Luzzi et al. (62)	Retrospective	Comparative Cohort Study	Intervention	$n = 117$	$n = 54$	$n = 0$	HGG
Mascitelli et al. (63)	Retrospective	Case Series	Intervention	$n = 79$	$n = 4$	$n = 0$	Glioma
Satoh et al. (64)	Prospective	Case Series	Intervention	$n = 20$	$n = 7$	$n = 0$	AE, DA, HGG, GG
Satoh et al. (65)	Prospective	Case Series	Intervention	$n = 5$	$n = 3$	$n = 0$	AA, GBM
Sun et al. (66)	Prospective	Comparative Study	Intervention	$n = 134$	$n = 79$	$n = 0$	LGG, HGG
Zhou et al. (67)	Prospective	Case Series	Development	$n = 16$	$n = 16$	$n = 20$	HGG

A, astrocytoma; AA, anaplastic astrocytoma; AE, anaplastic ependymoma; AO, anaplastic oligodendroglioma; AOA, anaplastic oligoastrocytoma; DA, diffuse astrocytoma; E, ependymoma; GBM, glioblastoma; GG, ganglioglioma; HGG, high-grade glioma (III, IV); LGG, low-grade glioma (I, II); OA, oligoastrocytoma; OD, oligodendroglioma; PA, pilocytic astrocytoma.

[a]Study design reflects glioma patients only.

[b]Models reflect phantom glioma patients.

whereas the device was most useful for patient positioning and bone removal in those with skull base lesions. Although their sample with low-grade gliomas was small ($n = 4$), the authors postulated that HUD would be practical for guiding localization and resection, though more robust data was needed.

Recently, Luzzie et al. (62) tested the safety and efficacy of a new multimodal AR technique. They compared patients with supratentorial high-grade glioma undergoing AR HDFT-F-based cytoreductive surgery ($n = 54$) to a cohort of patients undergoing conventional white-light surgery assisted by infrared neuronavigation ($n = 63$). See **Figure 6** for an illustrative case. Results showed higher EOR ($p = 0.019$), lower post-operative neurological deficits ($p = 0.011$), and longer progression-free survival ($p = 0.006$) in the study vs. control group. The EOR was specifically ≥98% in 85% of study cases. However, the types, grades, and percentages of complications in both groups were analogous (9.2% vs. 9.5%).

This AR HDFT-F technique was further validated in the maximal safe resection of a postcentral gyrus GBM (61). Post-operatively, the patient reported significant improvement in upper extremity motor function and regained their ability to walk, with no recurrence at nine months follow-up. In a similar case, Inoue et al. (54) applied a newly developed AR neuronavigation system that superimposed tumors and vasculature plus motor tractography. This proved useful in visualizing the patient's lesion border and corticospinal tract, yet it had limitations in depth perception and accuracy. Their tumor resided at the corpus callosum inside the resection cavity as a result. However, no new neurological deficits were observed.

Two studies leveraged a tablet-based AR navigation system, called the "trans-visible navigator" (TVN). Satoh et al. (65) first applied this apparatus to stereotactic biopsy in three cases of deep-seated lesions, allowing surgeons to confirm target point accuracy and trajectory suitability, as well as precisely advance biopsy probes. This evaded the drawbacks of frame-based stereotactic navigation, resulting in satisfactory histopathology without complication. In a separate study (64), the TVN was applied in seven surgeries of low- and high-grade glioma, to which surgeons rated its utility across the neurosurgical workflow. Based on their findings, the apparatus was most practical for resecting superficial tumors, but less so for deep-seated ones, except when using transcortical and interhemispheric approaches. Results emphasized the importance of pre-operative discussions with surgeons in maximizing the effectiveness of AR.

Given the expense of modern navigation systems, two studies examined mobile AR (mAR) for localizing low- and high-grade glioma, as alternative low-cost solutions. Hou et al. (53) compared an iPhone-based method to a frameless neuronavigation system ($n = 6$), demonstrating technical feasibility with comparable accuracy ($D \leq 5$ mm). Further, their device simplified image pre-processing, co-registration, and projection, all of which were completed under 10 min. Chen et al. (48) also examined mAR in the localization of a supratentorial glioma, leveraging the Sina Intraoperative Neurosurgical Assist app. Despite registration meriting improvement, their system was practical and reliable over standard neuronavigation ($D = 4.4 \pm 1.1$ mm). Notably, both studies used manual registration in their AR workflows, likely

TABLE 3 Summary of AR workflows in glioma surgery.

Author	Display	Brand	Technique	Registration	Accuracy	Primary outcomes
Archip et al. (46)	Surgical instruments	Not specified	Pre-operative MRI, fMRI, DTI, and/or MR spectroscopy data plus intra-operative MRI data segmented and modeled via 3D Slicer software, integrated into surgical instruments, superimposed on patient's brain	Patient-specific, volumetric non-rigid registration with anatomical landmarks and estimation for brain deformation	RD = 1.82 mm	Feasible application of non-rigid method that compensates for brain deformation within surgical time constraints; significantly increased alignment accuracy compared to rigid method; visualization of critical structural and functional brain areas
Carl et al. (47)	Microscope	PENTERO and PENTERO 900 (Carl Zeiss Meditec Inc., Oberkochen, Germany)	Pre-operative CT, CTA, and/or MRI data plus intra-operative CT data segmented and modeled via BrainLab software, integrated by HUD into surgical microscope, superimposed on patient's brain	Automatic, user-independent rigid and/or non-linear registration based on low-dose intra-operative CT via reference arrays and markers	TRE = 0.72 ± 0.24 mm	Reliable application with high accuracy; smooth integration into surgical workflow; good hand-eye coordination; intuitive depth perception and visualization of tumor extent and surrounding structures; high impression on patient anatomy, facilitating orientation
Chen et al. (48)	Smartphone	Honor 6 Plus (Huawei Technologies Co., Ltd., Shenzhen, China)	Pre-operative CT and MRI data segmented and modeled via 3D Slicer software, integrated into Android smartphone running Sina app superimposed on patient's head	Manual registration with anatomical landmarks and fiducial markers	D = 4.4 ± 1.1 mm	Practical application for visualizing and localizing supratentorial lesions, but not for infratentorial lesions; satisfactory accuracy compared to standard neuronavigation system; simple, cost-effective approach
De Mauro et al. (49)	Microscope	OPMI MD-NC1 (Carl Zeiss Meditec Inc., Oberkochen, Germany) (cited from prior work 68)	Pre-operative CT and MRI data segmented and modeled via 3D Slicer software, integrated into surgical microscope connected to infrared optical tracker of novel mixed reality system, superimposed on patient's head and brain	Manual, point-based registration with markers and reference model (cited prior work 68)	Not reported	Working prototype for visual and haptic simulation of LGG palpation; force feedback to distinguish normal from pathological tissue (VR); stereoscopic visualization with real time brain navigation and space cognition (AR)
Fick et al. (50)	HMD	Hololens 1.0 (Microsoft Corp., Redmond, Washington, USA)	Pre-operative CT and MRI data segmented and modeled via 3D Slicer software, integrated into HMD of holographic neuronavigation system, superimposed on patient's head and brain	Point-based matching rigid registration with anatomical landmarks, reference array, and visual markers	FRE = 8.55 mm	Proof-of-concept application for intraoperative patient tracking; relatively inaccurate registration and navigation accuracy, with hologram instability and drifting as well as functional difficulties
Finger et al. (51)	Endoscope	MINOP (Aesculap Inc., Tuttlingen, Germany)	Pre-operative CT or MRI data segmented and modeled via Scopus Nova Plan software, integrated into surgical endoscope of neuronavigation system, superimposed on phantom's or patient's head and brain (intraventricular space)	Hybrid registration with anatomical landmarks, surface matching, and optical reference matrix	D = 1.2 ± 0.4 mm	Feasible application with sufficient accuracy; increased precision to optimally place burr holes and apply trajectories; safe navigation and intuitive visualization of trajectories while perforating cortical and subcortical structures; helpful estimation of tumor location and surrounding structures
Ghimire et al. (52)	Microscope	KINEVO 900 (Carl Zeiss Meditec Inc., Jena, Germany)	Pre-operative MRI, fMRI, and/or DTI data segmented and modeled via StealthViz Medtronic software, combined with intra-operative ultrasound and cortical mapping data, integrated into surgical microscope of neuronavigation system, superimposed on patient's brain	Point-based registration with anatomical landmarks and visual markers	Not reported	Successful application; supplementary motor homunculus and novel subcortical motor map with accurate intra-operative identification of cortical and sub-cortical boundaries as well as localization of intercostal muscles

(*Continued*)

TABLE 3 Continued

Author	Display	Brand	Technique	Registration	Accuracy	Primary outcomes
Hou et al. (53)	Smartphone	LVL CAM (Daniel LLC, Apple Inc., App Store)	Pre-operative CT and MRI data segmented and modeled via Windows XP software, integrated into iPhone smartphone running LVL CAM iOS app, superimposed on patient's head	CT- or MRI-based registration with markers; manual co-registration of virtual images with sagittal photograph based on anatomical landmarks via FUSED app	D ≤ 5 mm	Feasible application; useful for localizing intracranial lesions at low-cost with high accuracy; suitable for shallow supratentorial lesions of moderate size, but not for infratentorial lesions
Inoue et al. (54)	Camera	Qcam Pro 9,000 with headset [QCAM-200S-HS] and Qcam Connect (Logicool Co., Tokyo, Japan)	Pre-operative MRI and DTI data segmented and modeled via 3D Slicer software, integrated into handheld or headband web camera with optical markers connected to neuronavigation system, superimposed on patient's head and brain	Point-based registration with fiducial markers and reference table	FRE = 1.8	Feasible application; useful for visualizing and navigating corticospinal tract without damage; effective for resecting surface tumors, but not for deep-seated tumors due to camera malperformance; difficulty accurately judging depth perception
Iseki et al. (55)	Camera	Not specified	Pre-operative CT and MRI data plus intra-operative MRI data segmented and modeled, integrated into high-definition couple charged device camera on liquid crystal monitor of information-guided navigation system with optical tracking and real-time update, superimposed on patient's brain	CT- and MRI-based registration with markers	ME = 0.8 mm	Successful application with excellent accuracy; significantly increased average resection rate and total removal rate of malignant gliomas using open MRI with disclosed brain deformation and shift; improved EOR when simultaneously using real-time update navigation
Koike et al. (56)	Camera	Not specified	Pre-operative CT, MRI, DTI, and 3DRA data segmented and modeled via Avizo Lite software, integrated into fusion 3DCG combined with intra-operative brain surface photograph of patient, as part of mixed reality registration system	Manual registration via paired anatomical landmark and thin-plate spline methods using fusion 3DCG as reference	TRE = 0.70 mm	Successful application integrated with 3DCG tractography model; excellent accuracy despite brain shift at time of intra-operative photograph
Koike et al. (57)	Camera	Not specified	Pre-operative CT, MRI, and 3DRA data segmented and modeled via Avizo Lite software, integrated into fusion 3DCG combined with intra-operative brain surface photograph of patient, as novel mixed reality registration method	Automatic, non-rigid registration via paired anatomical landmark, thin-plate spline, and NMI methods using fusion 3DCG as reference	TRE = 0.72 ± 0.04 mm	Working method with highly precise spatial alignment between real and virtual space with angle versatility; 3DCG useful for skin incision and craniotomy planning with ability to display functional information
Koike et al. (58)	Camera	Nikon D500 (Nikon Corp., Tokyo, Japan)	Pre-operative CT, MRI, and 3DRA data segmented and modeled via Avizo Lite software, integrated into novel mixed reality image-guided system, projecting intra-operative brain surface photograph of patient onto 3DCG	Automatic registration using the NMI method to align intra-operative brain surface photograph (target) and 3DCG (reference) textures	TRE = 1.19 ± 0.14 mm	Feasible and cost-effective application of mix reality projection mapping with accurate alignment and minimal equipment; efficiently observed pre- and intra-operative information in the same coordinate system
Liao et al. (59)	Video	Not specified	Intra-operative MRI data segmented and modeled via fuzzy connectedness and 3D Slicer software, integrated into integral videography of image overlay navigation system, superimposed on phantom's or patient's head and brain	Semi-automatic, point-based registration with donut and fiducial markers	TRE = 0.90 ± 0.21 mm	Feasible application of real-time, autostereoscopic image overlay for open MRI-guided glioma surgery with simplicity and satisfactory accuracy; potential to reduce procedure time

(*Continued*)

TABLE 3 Continued

Author	Display	Brand	Technique	Registration	Accuracy	Primary outcomes
Liu et al. (60)	HMD	HoloLens (Microsoft Inc., Redmond, Washington, USA)	Pre-operative CT, CTA, and MRI data segmented and modeled, integrated by HMD into MR holographic imaging technology system, superimposed on patient's head and brain	CT-based registration with visual markers and viewpoint tracking	Not reported	Successful application with real-time display of resection degree; significantly higher complete resection accuracy and post-operative recovery rate, as well as significantly lower post-operative complications compared to ultrasound
Luzzi et al. (61)	Microscope	KINEVO 900 (Carl Zeiss Meditec Inc., Oberkochen, Germany)	Pre-operative MRI, DWI, and DTI data segmented and modeled via BrainLab software, integrated into surgical microscope of neuronavigation system, superimposed on patient's brain	Not reported	Not reported	Safe and effective application in maximizing EOR of postcentral gyrus glioblastoma; high technique versatility; improved patient motor function; no tumor recurrence at 9-months follow-up
Luzzi et al. (62)	Microscope	OPMI Neuro-NC4 or KINEVO 900 (Carl Zeiss Meditec Inc., Oberkochen, Germany)	Pre-operative CT, MRI, fMRI, DWI, and MR spectroscopy data segmented and modeled via BrainLab software, integrated into robotic surgical microscope of neuronavigation system, superimposed on patient's brain	CT-based optical tracking registration with anatomical landmarks and surface matching	Not reported	Safe and effective application with significantly higher EOR and PFS rates compared to control group; optimized patient functional outcomes; limited accuracy and reliability due to parallax error and crowding of fiber tracts
Mascitelli et al. (63)	Microscope	PENTERO 900 (Carl Zeiss Meditec Inc., Dublin, California, USA) or Leica OH6 (Leica Microsystems Inc., Buffalo Grove, Illinois, USA)	Pre-operative CT, CTA, and MRI data segmented and modeled via BrainLab software, integrated by HUD (with variance) into surgical microscope, superimposed on patient's head and brain	Standard registration; co-registration of HUD with navigation system	Not reported	Safe application with good-to-excellent accuracy; useful for skin incision, craniotomy, dural opening, and corticectomy for intra-axial and superficial lesions; useful for bed/head positioning and extradural/intradural bone removal for skull base lesions; disabled in 59.6% of cases due to lack of use, distraction, and inaccuracy
Satoh et al. (64)	Tablet	Surface Pro (Microsoft Corp., Redmond, Washington, USA)	Pre-operative CT and MRI data segmented and modeled via Amira software, integrated into tablet PC with back-facing camera of trans-visible navigator system connected to optical markers, superimposed on patient's head and brain	Point-based registration with anatomical landmarks; co-registration of virtual and camera image	TRE = 2.31 ± 2.18 mm	Useful application for skin incisions, craniotomy, dural incisions, and superficial tumor resections; less useful for deep-seated tumor resections, except when using transcortical and interhemispheric approaches; minimal time and labor; pre-surgical discussions essential to efficacy
Satoh et al. (65)	Tablet	Surface Pro (Microsoft Corp., Redmond, Washington, USA)	Pre-operative CT and MRI data segmented and modeled via Amira software, integrated into tablet PC with back-facing camera of trans-visible navigator system combined with stereotactic frame, superimposed on patient's head and brain	Point-based registration with anatomical landmarks	Not reported	Feasible application to stereotactic biopsy of deep-seated lesions; clear trajectory and ability to advance biopsy probe precisely; avoidance of critical structures, the target point's location, and viewpoint turning
Sun et al. (66)	Microscope	Not specified	Pre-operative fMRI and DTI data plus intra-operative MRI and DTI data segmented and modeled via BrainLab software, integrated into surgical microscope of functional neuronavigation system, superimposed on patient's head and brain	Not reported	Not reported	Successful application for pre-operative planning as well as intra-operative guidance and manipulation; verified brain shift and quality control during surgery; significantly improved tumor resection rate and neurofunctional preservation

(*Continued*)

TABLE 3 Continued

Author	Display	Brand	Technique	Registration	Accuracy	Primary outcomes
Zhou et al. (67)	HMD	HoloLens (Microsoft Inc., Redmond, Washington, USA)	Pre-operative CT, MRI, and DTI data segmented and modeled, integrated by HMD into novel stereotactic mixed reality-guided surgical navigation system, superimposed on phantom's or patient's head and brain	Markerless spatial registration with spatial drift and movement compensation methods to precisely align virtual anatomy with real patient pre-operatively	RMSE = 1.18 mm (phantom), 1.86 mm (patient)	Feasible application with suitable accuracy and efficacy for clinical use and resection; intuitive diagnosis and performance of surgical planning pre-operatively as well as identification of lesion boundary intra-operatively

3D, three-dimensional; 3DCG, three-dimensional computer graphics; 3DRA, three-dimensional rotational angiography; CT, computerized tomography; CTA, computerized tomography angiography; D, deviation; DTI, diffusion tensor imaging, DWI, diffusion weighted imaging; EOR, extent of resection; fMRI, functional magnetic resonance imaging; FRE, fiducial registration error; GBM, glioblastoma; ME, mean error; HGG, high-grade glioma; HMD, head mounted display; HUD, head up display; LGG, low-grade glioma; MR, magnetic resonance; MRI, magnetic resonance imaging; NMI, normalized mutual information; nTMS, navigated transcranial magnetic stimulation; PC, personal computer; PFS, progression-free survival; RD, residual displacement, RMSE, root-mean-squared error; TRE, target registration error.

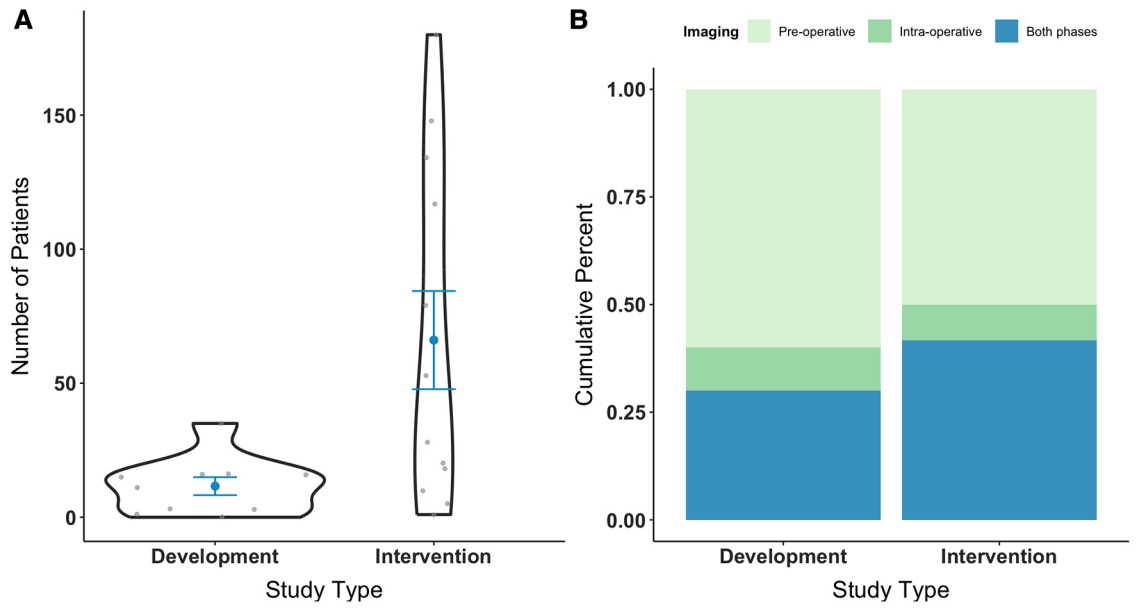


FIGURE 3
Number of patients diagnosed with glioma recruited in AR studies by study type, excluding phantom patients (A). Medical image acquisition phase by study type, spanning pre-operative and/or intra-operative stages (B). Study types are grouped by primary aim into "development" or "intervention", with respect to AR application. AR, augmented reality.

increasing time and attenuating projection alignment, which the authors noted.

Other applications included intra-ventricular neuroendoscopy and intra-dural spinal surgery. Finger et al. (51) described their first experience with AR-guided neuroendoscopy among six patients with glioma in addition to one phantom model. By integrating pre-operative information into the endoscope's field of view, they were able to optimally place burr holes, estimate tumor location and surrounding structures, and apply trajectories for surgical intervention. Carl et al. (47), on the other hand, applied microscope-based AR via HUD in two patients with intradural spinal gliomas. In both study cases, AR provided intuitive visualization of tumor extent and neighboring structures, with high registration accuracy (TRE = 0.72 ± 0.24). This was particularly useful for visualizing the cranio-caudal extent of an intra-medullary ependymoma.

Mixed reality applications

Six studies applied mixed reality (MR), a blend of augmented reality (AR) and virtual reality (VR), wherein physical and virtual objects co-exist and interact in real time. Motivated by computer-aided surgery (CAS), De Mauro et al. (49) developed a prototypical MR system for pre-operative training (VR) and intra-operative use (AR) embedded in a surgical microscope. Their VR feature simulated visual and tactile sensations of brain palpation, with force feedback interaction of soft and hard tissues. Using real patient data, this allowed surgeons to distinguish between normal and pathological tissue affected by low-grade glioma. In contrast, their AR feature enabled stereoscopic visualization of relevant 3D data for real-time brain navigation. This was specifically designed to aid image-guided neurosurgery.

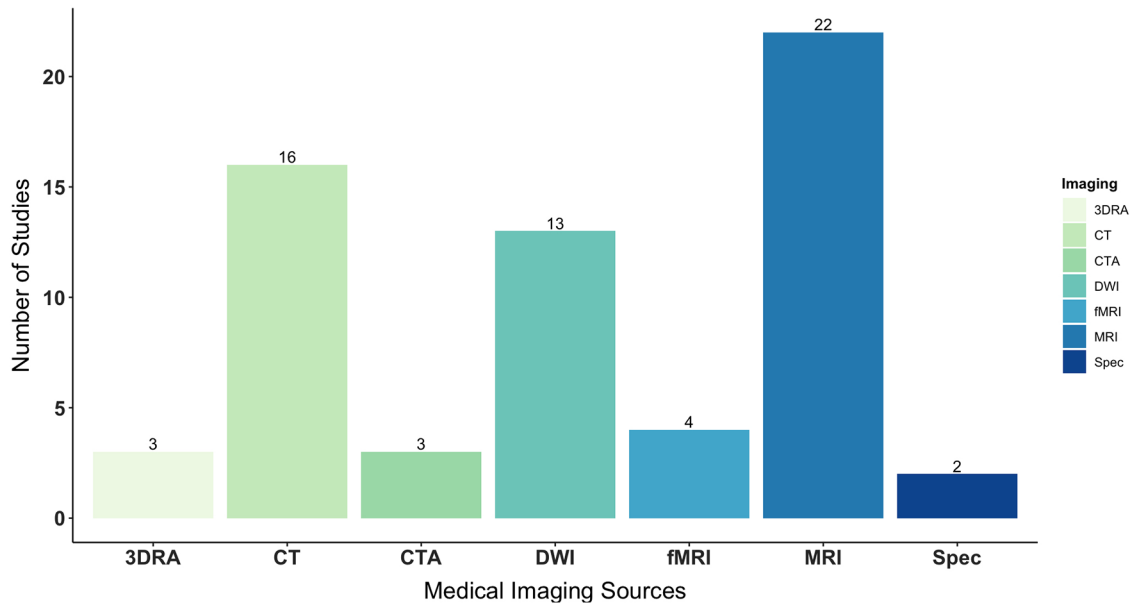


FIGURE 4
Types of medical imaging used as data sources for AR, specifically segmenting, modeling, formatting, and projecting virtual objects onto phantoms or patients' real anatomy. AR, augmented reality.

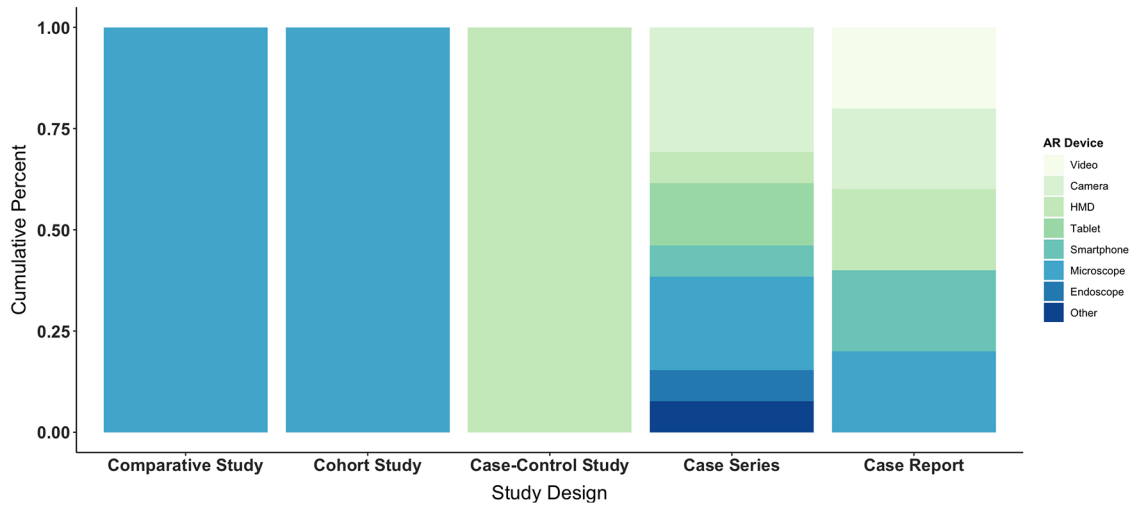


FIGURE 5
Types of augmented reality display devices used across study designs, including comparative, cohort, and case-control studies as well as case series and case reports. AR, augmented reality; HMD, head-mounted display.

Zhou et al. (67) similarly evaluated a novel MR navigation system. Using a markerless spatial registration method, they tested its ability to diagnose and perform surgical planning pre-operatively, as well as identify lesion boundaries intra-operatively. Compared to standard applications, and under ideal conditions, their system met accuracy, efficacy, and reliability benchmarks along with time requirements. This was validated in both phantom experiments ($n = 20$; RMSE = 1.18 mm; M_{time} = 6.02 min) and clinical trials ($n = 16$; RMSE = 1.86 mm; M_{time} = 7.95 min) among patients with high-grade glioma, indicating suitability for clinical use.

Moreover, using MR projection mapping (MRPM), Koike et al. (58) developed an image-guided surgery system that projects intra-operative brain surface photographs (BSPs; real space) onto high-resolution 3D computer graphics (3DCGs; virtual space). See Figure 7. This was accomplished without the need for large-scale equipment, such as neuronavigation and other computer-assisted technologies. In 14 glioma patients, their system displayed accurate alignment of patient anatomy (BSPs) and medical images (3DCGs), presented in the same coordinate system, even after brain shift due to craniotomy (TRE = 1.19 ± 0.14 mm). Alignment accuracy was evaluated by two neurosurgeons,

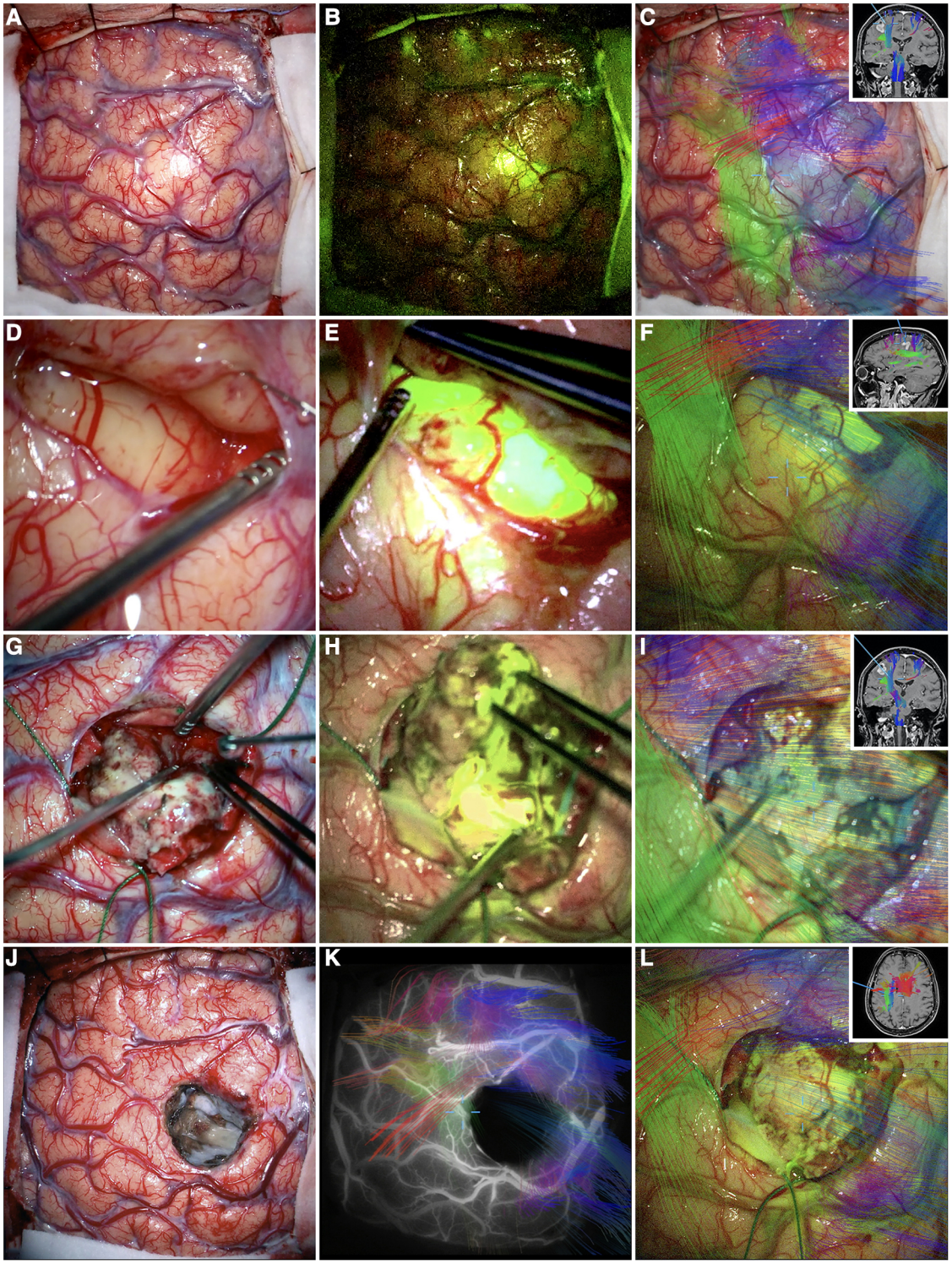


FIGURE 6
Illustrative case. Primary motor cortex glioblastoma of the dominant hemisphere. Intraoperative photographs obtained under white light (**A**), YELLOW 560 filter (**B**), and AR HDFT (**C**) before tumor resection. (**D–I**) The main steps of the surgery that were performed in large part along with the AR HDFT-F technique. (**J–L**) The surgical field at the end of the tumor resection obtained under white light (**J**), combined INFRARED 800 and AR HDFT during indocyanine green videoangiography (**K**), and AR HDFT-F (**L**). Insets in panels (**C,F,I,L**) are the screenshots obtained during the microscope focus-based neuronavigation. "Supratentorial High-Grade Gliomas: Maximal Safe Anatomical Resection Guided by Augmented Reality High-Definition Fiber Tractography and Fluorescein." This figure is protected by Copyright, is owned by The Journal of Neurosurgery Publishing Group (JNSPG), and is used with permission only within this document. Permission to use it otherwise must be secured from JNSPG. Full text of the article containing the original figure is available at thejns.org.

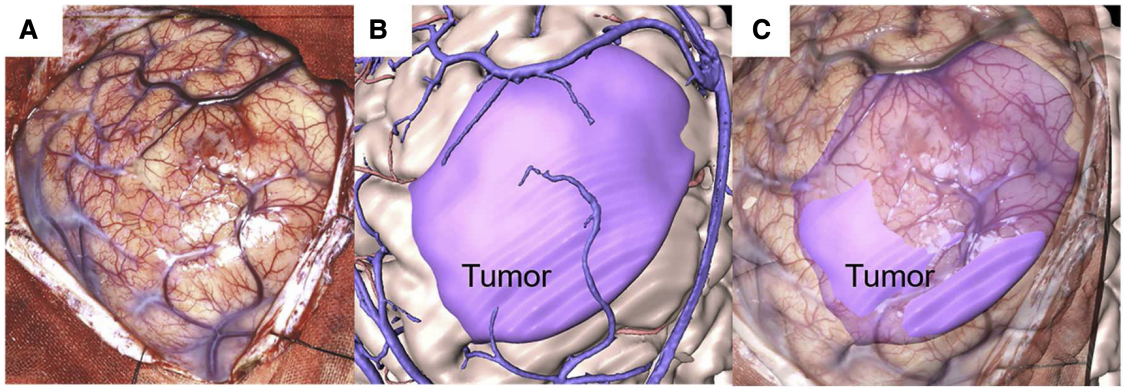


FIGURE 7
Illustrative case. A 31-year-old man with oligodendroglioma. (A) Intraoperative brain surface photograph in JPEG (Joint Photographic Experts Group) format. (B) Fusion 3-dimensional computer graphics (3DCG) created from preoperative imaging studies. The purple highlight indicates the tumor area. (C) Mixed-reality computer graphics created by aligning the intraoperative brain surface photograph and fusion 3DCG. The purple highlight indicates the tumor area. "Development of Innovative Neurosurgical Operation Support Method Using Mixed-Reality Computer Graphics." © 2021 The Author(s). Published by Elsevier Inc. Licensed under CC-BY-NC-ND.

together under consultation, who measured the difference between BSPs and 3DCGs. This was performed after dividing craniotomy areas into 16 fields. Further, MRPM made it possible for surgeons to plan trajectories for intervention based on cortical stimulation mapping. No post-operative complications were observed.

In a follow-up study ($N = 15$) (51, 52), recruiting patients with left hemispheric glioma, MRPM allowed surgeons to visualize the spatial correlation between medical images and the surgical field, specifically language-function hubs in the frontal lobe (TRE = 0.70 mm). This was likewise achieved despite brain shift at the time of the intra-operative BSP. Both studies leveraged a registration method previously developed and validated by Koike et al. (57), which demonstrated high spatial accuracy (TRE = 0.72 ± 0.04 mm).

In a retrospective, case-control study, Liu et al. (60) compared MR holographic imaging technology ($n = 30$) to ultrasound ($n = 23$) in neurosurgery for spinal cord glioma. Findings showed a significantly higher total tumor resection rate in the experimental group than in the control group (96.7%, vs. 82.6%; $p < 0.05$). This extended to more accurate complete resections (93.3%, vs. 73.5%; $p < 0.05$), far lower post-operative complications (3.3%, vs. 21.7%; $p < 0.05$), and improved recovery rates at 12 months follow-up (56.7%, vs. 41.1%; $p < 0.05$). Additionally, the tumor recurrence rate was lower in the experimental group compared to the control group at 12 months follow-up (0.0%, vs. 4.3%); however, this was not significant ($p > 0.05$). Enhanced MRI results were used to master the accuracy of intra-operative complete resections, informed by prior literature on AR and anatomic pathology (69). Post-operative MRI was then used to evaluate complete resections and their rate, as well the incidence of complications after surgery. Functional recovery was assessed via Modified McCormick Scale (MMS) grading.

Discussion

In this review, we summarized applications of AR in glioma surgery. Qualitatively, AR is a valuable tool that precisely overlays multiple imaging datasets, plus other relevant clinical information, onto the surgical field in real time. This obviates the need for surgeons to shift focus away from patients to nearby monitors for guidance, and mentally relate 2D information into 3D anatomy. Inattention blindness and interpretation error can therefore be mitigated. Further, AR enhances visualization of tumor complexity and its relationship to critical structures. This facilitates spatial understanding of neuroanatomy; aids surgeons in navigating, localizing, and resecting lesions; and may subsequently lead to improved patient outcomes.

However, at present, there is limited data that AR effectively extends the EOR and PFS as well as preserves motor, visual, and language function post-operatively. This is evidenced by the number, breadth, and quality of published studies. Of the 22 included in this review, 10 studies developed and tested AR systems, including segmentation and registration techniques, while the remaining 12 applied AR interventionally. Notably, 86% of studies were of low-level or weak evidence, based on hierarchical classes of evidence in neurosurgery (45), largely comprising case series and case reports. Accordingly, 77% of studies included 18 glioma patients or less and 73% lacked standard control groups. No studies involved random sampling nor random assignment.

Another critical limitation observed here, relevant to future research, was the inability to "double blind" patients and surgeons a key method for reducing detection bias. This problem underlies any technical advancement in surgery, namely trials of non-pharmacological treatments with physical components (70–73). As a result, surgeons who trusted in the efficacy of AR may have—unconsciously or deliberately—influenced the EOR. This might have led to an overestimation of treatment effects and perhaps more significant outcomes. Blinding of outcome assessors was additionally absent and/or unreported; yet it serves an important role in the case of soft endpoints, such as psychosocial function and quality of life. Still, this is not readily achieved (70, 72, 74). A possible solution is to apply the IDEAL

Framework (75, 76), a paradigm for incorporating evidence-based advances in neurosurgery. Here, specific study designs and reporting standards are recommended across five stages of surgical innovation: Idea, Development, Exploration, Assessment, and Long-term study. A controlled, interrupted-time series design is one acceptable alternative, suggested by the IDEAL framework, for minimizing known bias (77).

From a technical standpoint, AR workflows considerably varied across studies. Diverse approaches were employed to acquire and segment images, model virtual objects, register and track systems, and display fused data intra-operatively. This variation extended to measurements, with several terms used to express registration error, or the accuracy of overlaid virtual objects on patients' head and/or brain, each with its own shortcomings. The TRE, for instance, measures the anatomical region of interest for surgeons in 3D space, yet has poor depth perception. Other indices fail to correlate with (TRE) or underestimate (TE) true accuracy. This lack of standardization in measurement and reporting is well-documented in the field (78), and likely affects the validity of AR and its ability to guide localization and resection of intra-cranial targets, as well as the positioning of surgical instruments. Registration thus presents a significant and pressing challenge to precision neurosurgery. Along the same accord, most studies relied on software and hardware not formally vetted for pre- and/or intra-operative use. This chiefly applied to novel AR systems and techniques, which may lead to publication bias. As such, reported outcomes, at this stage, must be interpreted with caution.

One strategy towards standardization is founding a consortium of AR workflows in neurosurgery, including image acquisitions and open-source software, that could be utilized across research institutions. Compiling larger, more homogeneous, and longitudinal datasets may help identify universal methods; and allow for the development, validation, and use of machine learning algorithms (79)—and complimentary techniques (e.g., ultrasound 80, 81)—to maximize accuracy throughout procedures. This space also stands to benefit from a "readiness framework" to evaluate AR suitability for surgical implementation, drawing parallels to Tang et al.'s (82) analytic model. Their team specifically developed an evidence-based schema for assessing AR in medical education, underscoring four criteria: quality, application content, outcome, and feasibility. Adapting this model to AR in glioma surgery may address inconsistency in assessment tools and reliably gauge clinical utility. Mastering the use of AR in educational, training, and pre-operative settings will likewise increase its intra-operative value.

Limitations

This systematic review carries inherent limitations. First, no automated tools were employed during the selection and data collection process, increasing susceptibility to human error. In the event this occurred, studies returned by databases may have been overlooked, incorrectly excluded, and/or misreported in this review. Second, ascertaining the quality of evidence for each study was subjective, despite following hierarchical classes for neurosurgery. Studies may have been misclassified accordingly, especially those meeting criteria for more than one level. However, additional quality assurance measures, outlined in our methodology, were taken to ensure global data integrity. Finally, our qualitative analysis is limited by its data, which is heterogeneous at best. Standards for measurements and reporting will help improve the therapeutic value of AR in glioma surgery moving forward, and will enable meta-analytic approaches to precisely estimate both technical performance and treatment effects.

Conclusion

AR has increasing potential in the surgical management of glioma. It enables improved understanding of complex relationships between anatomy and pathology, aiding in real-time intra-operative navigation, localization, and resection. Further, there are signals of improvement in neurofunctional preservation associated with AR use, pointing to real, discernable benefit to patient care. This is evermore salient given the poor prognosis of gliomas, especially those with malignant and invasive presentations. However, technical and design limitations are readily apparent. A universal approach for developing, applying, and measuring AR systems, for instance, is critically absent. The field must consider the importance of consistency and replicability to effectively converge on standard uses of AR and its therapeutic value.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

A.R.: Conceptualization, Methodology, Investigation, Software, Formal Analysis, Visualization, Writing—Original Draft, Project Administration. C.L.: Writing—Review & Editing, Supervision. M.Z.: Investigation, Data Curation, Validation, Writing—Review & Editing. M.M.: Writing—Review & Editing, Supervision. V.C.: Data Curation, Validation, Writing—Original Draft. W.H.: Investigation, Data Curation, Writing—Review & Editing. N.B.: Investigation, Data Curation, Writing—Review & Editing. A.K.U.: Investigation, Data Curation, Software, Formal Analysis, Visualization, Writing—Review & Editing. L.A.: Data Curation, Validation, Writing—Review & Editing. Z.O.U.: Validation, Writing—Review & Editing. R.R.G.: Writing—Review & Editing, Supervision. J.S.: Writing—Review & Editing, Supervision.

Funding

All research at the Department of Psychiatry in the University of Cambridge was supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312) and the NIHR Applied Research Collaboration East of England. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Acknowledgments

We express our gratitude to the patients, clinicians, and researchers—described in this report—for their valuable contributions to science understanding the potential of augmented reality for glioma surgery.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. American Association of Neurological Surgeon. Brain Tumors. Available at: https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors (Accessed August 19, 2021).

2. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, et al. The epidemiology of glioma in adults: a "state of the science" review. *Neuro-Oncology*. (2014) 16(7):896–913. doi: 10.1093/neuonc/nou087

3. Lin D, Wang M, Chen Y, Gong J, Chen L, Shi X, et al. Trends in intracranial glioma incidence and mortality in the United States, 1975–2018. *Front Oncol*. (2021) 11:748061. doi: 10.3389/fonc.2021.748061

4. Gerritsen JKW, Broekman MLD, De Vleeschouwer S, Schucht P, Nahed BV, Berger MS, et al. Safe surgery for glioblastoma: recent advances and modern challenges. *Neurooncol Pract*. (2022) 9(5):364–79. doi: 10.1093/nop/npac019

5. Rabab'h O, Al-Ramadan A, Shah J, Lopez-Negrete H, Gharaibeh A. Twenty years after glioblastoma multiforme diagnosis: a case of long-term survival. *Cureus*. (2021) 13(6). doi: 10.7759/cureus.16061

6. Lombardi G, Barresi V, Castellano A, Tabouret E, Pasqualetti F, Salvalaggio A, et al. Clinical management of diffuse low-grade gliomas. *Cancers (Basel)*. (2020) 12 (10):3008. doi: 10.3390/cancers12103008

7. Kiran M, Chatrath A, Tang X, Keenan DM, Dutta A. A prognostic signature for lower grade gliomas based on expression of long non-coding RNAs. *Mol Neurobiol*. (2019) 56(7):4786–98. doi: 10.1007/s12035-018-1416-y

8. Jakola AS, Skjulsvik AJ, Myrmel KS, Sjåvik K, Unsgård G, Torp SH, et al. Surgical resection versus watchful waiting in low-grade gliomas. *Ann Oncol*. (2017) 28 (8):1942–8. doi: 10.1093/annonc/mdx230

9. Sanai N, Chang S, Berger MS. Low-grade gliomas in adults: a review. *JNS*. (2011) 115(5):948–65. doi: 10.3171/2011.7.JNS101238

10. Stupp R, Brada M, van den Bent MJ, Tonn JC, Pentheroudakis G. High-grade glioma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. (2014) 25:iii93–iii101. doi: 10.1093/annonc/mdu050

11. Witteler J, Kjaer TW, Tvilsted S, Schild SE, Rades D. Re-Evaluation of prognostic factors for survival after radiotherapy of cerebral gliomas: a supplementary analysis to a previous study. *Anticancer Res*. (2020) 40(11):6513–5. doi: 10.21873/anticanres.14674

12. Duffau H. Long-term outcomes after supratotal resection of diffuse low-grade gliomas: a consecutive series with 11-year follow-up. *Acta Neurochir*. (2016) 158 (1):51–8. doi: 10.1007/s00701-015-2621-3

13. Hervey-Jumper SL, Berger MS. Maximizing safe resection of low- and high-grade glioma. *J Neurooncol*. (2016) 130(2):269–82. doi: 10.1007/s11060-016-2110-4

14. Rossi M, Ambrogi F, Gay L, Gallucci M, Conti Nibali M, Leonetti A, et al. Is supratotal resection achievable in low-grade gliomas? Feasibility, putative factors, safety, and functional outcome. *J Neurosurg*. (2020) 132(6):1692–705. doi: 10.3171/2019.2.JNS183408

15. Stummer W, van den Bent MJ, Westphal M. Cytoreductive surgery of glioblastoma as the key to successful adjuvant therapies: new arguments in an old discussion. *Acta Neurochir*. (2011) 153(6):1211–8. doi: 10.1007/s00701-011-1001-x

16. Pallud J, Varlet P, Devaux B, Geha S, Badoual M, Deroulers C, et al. Diffuse low-grade oligodendrogliomas extend beyond MRI-defined abnormalities. *Neurology*. (2010) 74(21):1724–31. doi: 10.1212/WNL.0b013e3181e04264

17. Giammalva GR, Brunasso L, Costanzo R, Paolini F, Umana GE, Scalia G, et al. Brain mapping-aided SupraTotal resection (SpTR) of brain tumors: the role of brain connectivity. *Front Oncol*. (2021) 11:645854. doi: 10.3389/fonc.2021.645854

18. Mandal AS, Romero-Garcia R, Hart MG, Suckling J. Genetic, cellular, and connectomic characterization of the brain regions commonly plagued by glioma. *Brain*. (2020) 143(11):3294–307. doi: 10.1093/brain/awaa277

19. Osswald M, Jung E, Sahm F, Solecki G, Venkataramani V, Blaes J, et al. Brain tumour cells interconnect to a functional and resistant network. *Nature*. (2015) 528 (7580):93–8. doi: 10.1038/nature16071

20. Liang J, Lv X, Lu C, Ye X, Chen X, Fu J, et al. Prognostic factors of patients with gliomas—an analysis on 335 patients with glioblastoma and other forms of gliomas. *BMC Cancer*. (2020) 20(1):35. doi: 10.1186/s12885-019-6511-6

21. Gulati S, Jakola AS, Nerland US, Weber C, Solheim O. The risk of getting worse: surgically acquired deficits, perioperative complications, and functional outcomes after primary resection of glioblastoma. *World Neurosurg*. (2011) 76(6):572–9. doi: 10.1016/j.wneu.2011.06.014

22. Ius T, Angelini E, Thiebaut de Schotten M, Mandonnet E, Duffau H. Evidence for potentials and limitations of brain plasticity using an atlas of functional resectability of WHO grade II gliomas: towards a "minimal common brain.". *NeuroImage*. (2011) 56(3):992–1000. doi: 10.1016/j.neuroimage.2011.03.022

23. De Benedictis A, Duffau H, Paradiso B, Grandi E, Balbi S, Granieri E, et al. Anatomo-functional study of the temporo-parieto-occipital region: dissection, tractographic and brain mapping evidence from a neurosurgical perspective. *J Anat*. (2014) 225(2):132–51. doi: 10.1111/joa.12204

24. Sarubbo S, De Benedictis A, Merler S, Mandonnet E, Balbi S, Granieri E, et al. Towards a functional atlas of human white matter: functional atlas of white matter. *Hum Brain Mapp*. (2015) 36(8):3117–36. doi: 10.1002/hbm.22832

25. Fults DW. Commentary: occurrence, risk factors, and consequences of postoperative ischemia after glioma resection: a retrospective study. *Neurosurgery*. (2023) 92(1):e5–6. doi: 10.1227/neu.0000000000002173

26. Mandonnet E, Duffau H. An attempt to conceptualize the individual onco-functional balance: why a standardized treatment is an illusion for diffuse low-grade glioma patients. *Crit Rev Oncol Hematol*. (2018) 122:83–91. doi: 10.1016/j.critrevonc.2017.12.008

27. Roberts DW, Strohbehn JW, Hatch JF, Murray W, Kettenberger H. A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. *J Neurosurg*. (1986) 65(4):545–9. doi: 10.3171/jns.1986.65.4.0545

28. Chidambaram S, Stifano V, Demetres M, Teyssandier M, Palumbo MC, Redaelli A, et al. Applications of augmented reality in the neurosurgical operating room: a systematic review of the literature. *J Clin Neurosci*. (2021) 91:43–61. doi: 10.1016/j.jocn.2021.06.032

29. Cho J, Rahimpour S, Cutler A, Goodwin CR, Lad SP, Codd P. Enhancing reality: a systematic review of augmented reality in neuronavigation and education. *World Neurosurg*. (2020) 139:186–95. doi: 10.1016/j.wneu.2020.04.043

30. Léger É, Drouin S, Collins DL, Popa T, Kersten-Oertel M. Quantifying attention shifts in augmented reality image-guided neurosurgery. *Healthc Technol Lett*. (2017) 4 (5):188–92. doi: 10.1049/htl.2017.0062

31. Haouchine N, Juvekar P, Wells WM III, Cotin S, Golby A, Frisken S, et al. Deformation aware augmented reality for craniotomy using 3D/2D non-rigid registration of cortical vessels. In: Martel AL, Abolmaesumi P, Stoyanov D, editors. *Medical image computing and computer assisted intervention—mICCAI 2020*. Vol. 12264. Lecture Notes in Computer Science. Lima: Springer International Publishing (2020). p. 735–44. doi: 10.1007/978-3-030-59719-1_71

32. Haouchine N, Juvekar P, Nercessian M, Wells W, Golby A, Frisken S. Pose estimation and non-rigid registration for augmented reality during neurosurgery. *IEEE Trans Biomed Eng*. (2022) 69(4):1310–7. doi: 10.1109/TBME.2021.3113841

33. Bayer S, Maier A, Ostermeier M, Fahrig R. Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery. *Int J Biomed Imaging*. (2017) 2017:1–18. doi: 10.1155/2017/6028645

34. Gerard IJ, Kersten-Oertel M, Petrecca K, Sirhan D, Hall JA, Collins DL. Brain shift in neuronavigation of brain tumors: a review. *Med Image Anal*. (2017) 35:403–20. doi: 10.1016/j.media.2016.08.007

35. Orringer DA, Golby A, Jolesz F. Neuronavigation in the surgical management of brain tumors: current and future trends. *Expert Rev Med Devices*. (2012) 9(5):491–500. doi: 10.1586/erd.12.42

36. Furht B, editor. *Handbook of augmented reality*. New York: Springer (2011). doi: 10.1007/978-1-4614-0064-6

37. Contreras López WO, Navarro PA, Crispin S. Intraoperative clinical application of augmented reality in neurosurgery: a systematic review. *Clin Neurol Neurosurg*. (2019) 177:6–11. doi: 10.1016/j.clineuro.2018.11.018

38. Guha D, Alotaibi NM, Nguyen N, Gupta S, McFaul C, Yang VXD. Augmented reality in neurosurgery: a review of current concepts and emerging applications. *Can J Neurol Sci*. (2017) 44(3):235–45. doi: 10.1017/cjn.2016.443

39. Lee C, Wong GKC. Virtual reality and augmented reality in the management of intracranial tumors: a review. *J Clin Neurosci*. (2019) 62:14–20. doi: 10.1016/j.jocn.2018.12.036

40. Meola A, Cutolo F, Carbone M, Cagnazzo F, Ferrari M, Ferrari V. Augmented reality in neurosurgery: a systematic review. *Neurosurg Rev*. (2017) 40(4):537–48. doi: 10.1007/s10143-016-0732-9

41. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. (2021) 10(1):89. doi: 10.1186/s13643-021-01626-4

42. Foster ED, Deardorff A. Open science framework (OSF). *JMLA*. (2017) 105(2). doi: 10.5195/jmla.2017.88

43. Amir-Behghadami M, Janati A. Population, intervention, comparison, outcomes and study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. *Emerg Med J*. (2020) 37(6):387. doi: 10.1136/emermed-2020-209567

44. R: A Language and Environment for Statistical Computing. (2018). Available at: https://www.R-project.org/

45. Rutka JT. Editorial. Classes of evidence in neurosurgery. *JNS*. (2016) 126 (6):1747–8. doi: 10.3171/2016.4.JNS161067

46. Archip N, Clatz O, Whalen S, Kacher D, Fedorov A, Kot A, et al. Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. *NeuroImage*. (2007) 35(2):609–24. doi: 10.1016/j.neuroimage.2006.11.060

47. Carl B, Bopp M, Saß B, Pojskic M, Nimsky C. Augmented reality in intradural spinal tumor surgery. *Acta Neurochir*. (2019) 161(10):2181–93. doi: 10.1007/s00701-019-04005-0

48. Chen JG, Han KW, Zhang DF, Li ZX, Li YM, Hou LJ. Presurgical planning for supratentorial lesions with free slicer software and sina app. *World Neurosurg*. (2017) 106:193–7. doi: 10.1016/j.wneu.2017.06.146

49. De Mauro A, Raczkowsky J, Halatsch ME, Wörn H. Mixed reality neurosurgical microscope for training and intra-operative purposes. In: Shumaker R, editor. *Virtual and mixed reality*. Vol. 5622. Lecture Notes in Computer Science. San Diego, CA: Springer Berlin Heidelberg (2009). p. 542–9. doi: 10.1007/978-3-642-02771-0_60

50. Fick T, van Doormaal JAM, Hoving EW, Regli L, van Doormaal TPC. Holographic patient tracking after bed movement for augmented reality neuronavigation using a head-mounted display. *Acta Neurochir*. (2021) 163 (4):879–84. doi: 10.1007/s00701-021-04707-4

51. Finger T, Schaumann A, Schulz M, Thomale UW. Augmented reality in intraventricular neuroendoscopy. *Acta Neurochir*. (2017) 159(6):1033–41. doi: 10.1007/s00701-017-3152-x

52. Ghimire P, Lavrador JP, Baig Mirza A, Pereira N, Keeble H, Borri M, et al. Intraoperative mapping of pre-central motor cortex and subcortex: a proposal for supplemental cortical and novel subcortical maps to Penfield's motor homunculus. *Brain Struct Funct*. (2021) 226(5):1601–11. doi: 10.1007/s00429-021-02274-z

53. Hou Y, Ma L, Zhu R, Chen X, Zhang J. A low-cost iPhone-assisted augmented reality solution for the localization of intracranial lesions. *PLoS ONE*. (2016) 11(7): e0159185. doi: 10.1371/journal.pone.0159185

54. Inoue D, Cho B, Mori M, Kikkawa Y, Amano T, Nakamizo A, et al. Preliminary study on the clinical application of augmented reality neuronavigation. *J Neurol Surg A Cent Eur Neurosurg*. (2013) 74(2):071–6. doi: 10.1055/s-0032-1333415

55. Iseki H, Muragaki Y, Naemura K, Hayashi M, Hori T, Takakura K. *Clinical application of augmented reality in neurosurgical field. Proceedings Computer Graphics International 2003. IEEE Comput. Soc* (2003). p. 44–9. doi: 10.1109/CGI.2003.1214446

56. Koike T, Tanaka S, Kin T, Suzuki Y, Takayanagi S, Takami H, et al. Accurate preoperative identification of motor speech area as termination of arcuate Fasciculus depicted by Q-ball imaging tractography. *World Neurosurg*. (2022) 164: e764–71. doi: 10.1016/j.wneu.2022.05.041

57. Koike T, Kin T, Tanaka S, Takeda Y, Uchikawa H, Shiode T, et al. Development of innovative neurosurgical operation support method using mixed-reality computer graphics. *World Neurosurgery: X*. (2021) 11:100102. doi: 10.1016/j.wnsx.2021.100102

58. Koike T, Kin T, Tanaka S, Sato K, Uchida T, Takeda Y, et al. Development of a new image-guided neuronavigation system: mixed-reality projection mapping is accurate and feasible. *Oper Neurosurg*. (2021) 21(6):549–57. doi: 10.1093/ons/opab353

59. Liao H, Inomata T, Sakuma I, Dohi T. Surgical navigation of integral videography image overlay for open MRI-guided glioma surgery. In: Yang GZ, Jiang T, Shen D, Gu L, Yang J, editors. *Medical imaging and augmented reality*. Vol. 4091. Lecture Notes in Computer Science. Shanghai: Springer Berlin Heidelberg (2006). p. 187–94. doi: 10.1007/11812715_24

60. Liu Z, Li X, Lu J. Application evaluation of mixed-reality holographic imaging technology in the surgical treatment of spinal cord glioma. *Pak J Med Sci*. (2022) 38(7):1958. doi: 10.12669/pjms.38.7.4777

61. Luzzi S, Giotta Lucifero A, Baldoncini M, Del Maestro M, Galzio R. Postcentral gyrus high-grade glioma: maximal safe anatomic resection guided by augmented reality with fiber tractography and fluorescein. *World Neurosurg*. (2022) 159:108. doi: 10.1016/j.wneu.2021.12.072

62. Luzzi S, Giotta Lucifero A, Martinelli A, Maestro MD, Savioli G, Simoncelli A, et al. Supratentorial high-grade gliomas: maximal safe anatomical resection guided by augmented reality high-definition fiber tractography and fluorescein. *Neurosurg Focus*. (2021) 51(2):E5. doi: 10.3171/2021.5.FOCUS21185

63. Mascitelli JR, Schlachter L, Chartrain AG, Oemke H, Gilligan J, Costa AB, et al. Navigation-linked heads-up display in intracranial surgery: early experience. *Operative Surg*. (2018) 15(2):184–93. doi: 10.1093/ons/opx205

64. Satoh M, Nakajima T, Yamaguchi T, Watanabe E, Kawai K. Evaluation of augmented-reality based navigation for brain tumor surgery. *J Clin Neurosci*. (2021) 94:305–14. doi: 10.1016/j.jocn.2021.10.033

65. Satoh M, Nakajima T, Yamaguchi T, Watanabe E, Kawai K. Application of augmented reality to stereotactic biopsy. *Neurol Med Chir(Tokyo)*. (2019) 59 (11):444–7. doi: 10.2176/nmc.tn.2019-0128

66. Chen JG, Han KW, Zhang DF, Li ZX, Li YM, Hou LJ. Impact of virtual and augmented reality based on intraoperative magnetic resonance imaging and functional neuronavigation in glioma surgery involving eloquent areas. *World Neurosurg*. (2016) 96:375–82. doi: 10.1016/j.wneu.2016.07.107

67. Zhou Z, Yang Z, Jiang S, Zhuo J, Zhu T, Ma S. Augmented reality surgical navigation system based on the spatial drift compensation method for glioma resection surgery. *Med Phys*. (2022) 49(6):mp.15650. doi: 10.1002/mp.15650

68. Aschke M, Wirtz CR, Raczkowsky J, Worn H, Kunze S. *Augmented reality in operating microscopes for neurosurgical interventions. First international IEEE EMBS conference on neural engineering, 2003. Conference proceedings*. IEEE (2003). p. 652–5. doi: 10.1109/CNE.2003.1196913

69. Hanna MG, Ahmed I, Nine J, Prajapati S, Pantanowitz L. Augmented reality technology using microsoft HoloLens in anatomic pathology. *Arch Pathol Lab Med*. (2018) 142(5):638–44. doi: 10.5858/arpa.2017-0189-OA

70. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol*. (2004) 57(6):543–50. doi: 10.1016/j.jclinepi.2003.12.010

71. Probst P, Zaschke S, Heger P, Harnoss JC, Hüttner FJ, Mihaljevic AL, et al. Evidence-based recommendations for blinding in surgical trials. *Langenbecks Arch Surg*. (2019) 404(3):273–84. doi: 10.1007/s00423-019-01761-6

72. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery*. (1995) 118(3):459–67. doi: 10.1016/S0039-6060(05)80359-9

73. Takroni R, Sharma S, Reddy K, Zagzoog N, Aljoghaiman M, Alotaibi M, et al. Randomized controlled trials in neurosurgery. *Surg Neurol Int*. (2022) 13:379. doi: 10.25259/SNI_1032_2021

74. Probst P, Grummich K, Harnoss JC, Hüttner FJ, Jensen K, Braun S, et al. Placebo-controlled trials in surgery: a systematic review and meta-analysis. *Medicine (Baltimore)*. (2016) 95(17):e3516. doi: 10.1097/MD.0000000000003516

75. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM. Evaluation and stages of surgical innovations. *Lancet*. (2009) 374(9695):1089–96. doi: 10.1016/S0140-6736(09)61083-7

76. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. (2009) 374(9695):1105–12. doi: 10.1016/S0140-6736(09)61116-8

77. Hirst A, Philippou Y, Blazeby J, Campbell B, Campbell M, Feinberg J, et al. No surgical innovation without evaluation: evolution and further development of the

IDEAL framework and recommendations. *Ann Surg*. (2019) 269(2):211–20. doi: 10.1097/SLA.0000000000002794

78. Fick T, van Doormaal JAM, Hoving EW, Willems PWA, van Doormaal TPC. Current accuracy of augmented reality neuronavigation systems: systematic review and meta-analysis. *World Neurosurg*. (2021) 146:179–88. doi: 10.1016/j.wneu.2020.11.029

79. Tonutti M, Gras G, Yang GZ. A machine learning approach for real-time modelling of tissue deformation in image-guided neurosurgery. *Artif Intell Med*. (2017) 80:39–47. doi: 10.1016/j.artmed.2017.07.004

80. Gerard IJ, Kersten-Oertel M, Drouin S, Hall JA, Petrecca K, De Nigris D, et al. Combining intraoperative ultrasound brain shift correction and augmented reality visualizations: a pilot study of eight cases. *J Med Imag*. (2018) 5(2):1. doi: 10.1117/1.JMI.5.2.021210

81. Gerard IJ, Kersten-Oertel M, Hall JA, Sirhan D, Collins DL. Brain shift in neuronavigation of brain tumors: an updated review of intra-operative ultrasound applications. *Front Oncol*. (2021) 10:618837. doi: 10.3389/fonc.2020.618837

82. Tang KS, Cheng DL, Mi E, Greenberg PB. Augmented reality in medical education: a systematic review. *Can Med Ed J*. (2019) 11(1):e81. doi: 10.36834/cmej.61705

frontiers | Frontiers in Surgery

TYPE Case Report
PUBLISHED 22 August 2023
DOI 10.3389/fsurg.2023.1227473

OPEN ACCESS

EDITED BY
Mohamad Bydon,
Mayo Clinic, United States

REVIEWED BY
Toki Saito,
The University of Tokyo Hospital, Japan
Antonio Ferretti,
University of Studies G. d'Annunzio Chieti and Pescara, Italy

*CORRESPONDENCE
Elisa Colombo
✉ Elisa.colombo@usz.ch

RECEIVED 24 May 2023
ACCEPTED 21 July 2023
PUBLISHED 22 August 2023

CITATION
Colombo E, Bektas D, Regli L and van Doormaal T (2023) Case report: Impact of mixed reality on anatomical understanding and surgical planning in a complex fourth ventricular tumor extending to the lamina quadrigemina.
Front. Surg. 10:1227473.
doi: 10.3389/fsurg.2023.1227473

Case report: Impact of mixed reality on anatomical understanding and surgical planning in a complex fourth ventricular tumor extending to the lamina quadrigemina

Elisa Colombo*, Delal Bektas, Luca Regli and Tristan van Doormaal

Department of Neurosurgery and Clinical Neurocenter, University Hospital Zurich, Zurich, Switzerland

Background and importance: Tumors of the fourth ventricle account for 1%–5% of all intracranial neoplastic lesions and present with different configurations and anatomical challenges. Microsurgery represents the primary therapeutic strategy for the majority of fourth ventricular tumors, and adequate anatomical understanding and visualization are paramount to surgical planning and success. The authors present the case of a young patient with a complex fourth ventricular tumor, whose surgery was successfully planned using a novel mixed reality (MxR) system.
Case description: We present a case of a 31-year-old woman with a lesion extending from the fourth ventricle to the lamina quadrigemina and causing symptomatic hydrocephalus occlusus. Through the combined use of routine 2D images and an interactive 3D anatomical model, an interhemispheric transtentorial approach was used to remove 98% of the lesion with successful functional outcomes.
Conclusions: The application of advanced 3D visualization with a novel MxR system to the surgical planning of a complex fourth ventricular lesion proved relevant in designing the best surgical approach and trajectory to better identify potential intraoperative challenges and rehearse the patient-specific anatomy. The present case report endorses the implementation of advanced 3D visualization in routine perioperative practice.

KEYWORDS

mixed reality, 3D visualization, fourth ventricle, tumors of the IV ventricle, posterior fossa microneurosurgery

Introduction

To date, the primary therapeutic strategy for tumors of the fourth ventricle is microsurgical removal (1–3). Lesions of the fourth ventricle may present with a wide range of anatomical variations and are in close proximity with extremely eloquent structures in the context of the anatomical complexity of the posterior fossa. Therefore, surgical planning is fundamental to tailor the microsurgical approach for complex fourth ventricular lesions, and it may not be optimal if planned based on two-dimensional (2D) images only. The application of three-dimensional (3D) visualization using mixed reality (MxR) may offer a better anatomical understanding, and it has an emerging application in neurosurgery (4). Furthermore, diverse MxR systems have been developed, and initial

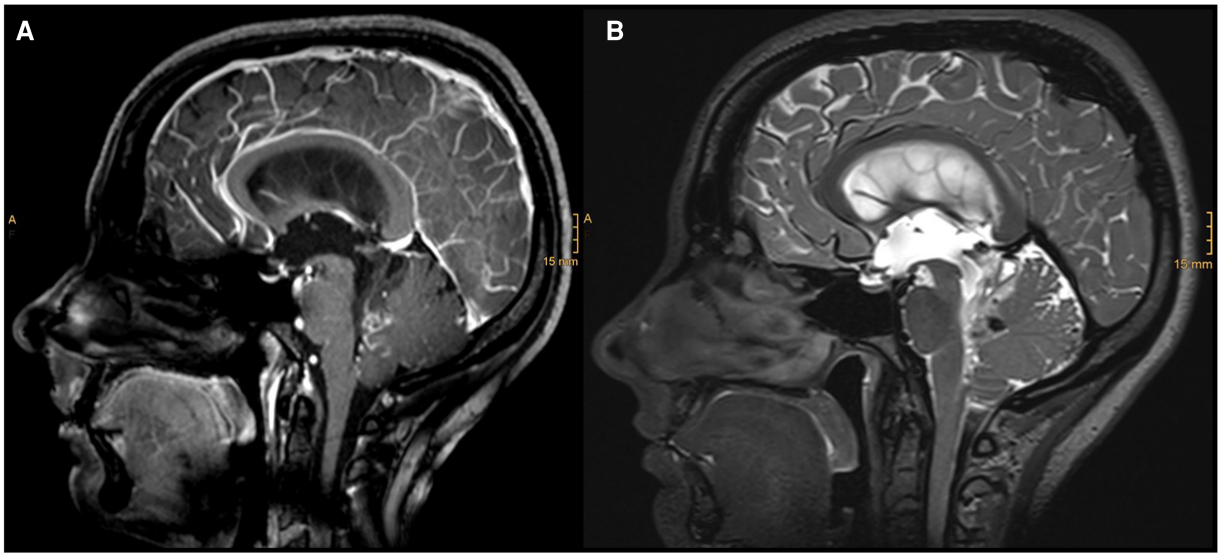


FIGURE 1
Diagnostic MRI study documenting a partially contrasted lesion of the IV ventricle with cystic components and infiltration of the vermis: (A) sagittal T1 with contrast; (B) sagittal T2.

experience on their implementation in the perioperative phase has been gathered (5–7). The advantages of patient-specific interactive 3D anatomical models may be multiple and range over the design of the optimal surgical approach and intraoperative corridor, a realistic and comprehensive understanding of the close anatomical structures, and the pre-operative rehearsal and anticipation of the intraoperative challenges (8, 9).

The authors describe a successful surgical approach for a complex fourth ventricular tumor extending until the lamina quadrigemina and infiltrating the basal veins designed using a novel mixed reality system.

Case description

A 31-year-old woman presented with a 5-day history of double vision in all directions, holocephalic headaches, and nausea. Radiologic workup showed a hyperintense lesion extending from the fourth ventricle to the lamina quadrigemina and causing a hydrocephalus occlusus (Figure 1). An external ventricular drainage (EVD) was immediately inserted, and 1 day after, the resection of the ventricular mass was scheduled.

Surgical strategy and patient positioning were discussed based on routinely used 2D DICOM (Digital Imaging and Communications in Medicine) images and further analyzed using 3D MxR technology.

Mixed reality system

The novel MxR system utilized for holographic rendering and 3D visualization consisted of different components (Figure 2). The primary component was a CE-certified cloud environment (Lumi, Augmedit bv, Naarden, Netherlands) providing a direct MxR output using a validated expanding mesh algorithm (Disior, Helsinki, Finland) (10, 11). The Lumi cloud environment had associated applications for MxR glasses, which represented the second component of the system and the main tool for 3D visualization of the hologram (Hololens 2, Microsoft Corporation, Redmond, WA, USA).

Hologram rendering

2D DICOM images from an external institution were loaded in a completely anonymized form on the novel CE-certified cloud environment (Lumi by Augmedit bv, Utrecht, Netherlands) starting from the acquisition of an axial T1 MPRAGE with contrast medium (Siemens Skyra 3 T MRI, TR = 1,800 ms, 168 slices, slice thickness = 1.0 mm, acquisition matrix = 256 × 232, FOV: 90.62, pixel spacing: 0.48828125). The magnetic resonance study was automatically segmented on the platform to highlight the skin, skull, brain, ventricular system, and tumor. Following the automatic segmentation, the tumor was further refined by means of manual segmentation on 3D Slicer (3D Slicer image computing platform | 3D Slicer) (12).

Hologram interactive 3D visualization

Through the MxR glasses, the primary user could visualize the hologram in 3D and interact with it and project it onto the real anatomy of the patient, manually matching it using her head to review and plan surgical positioning and the approach (Figure 3).

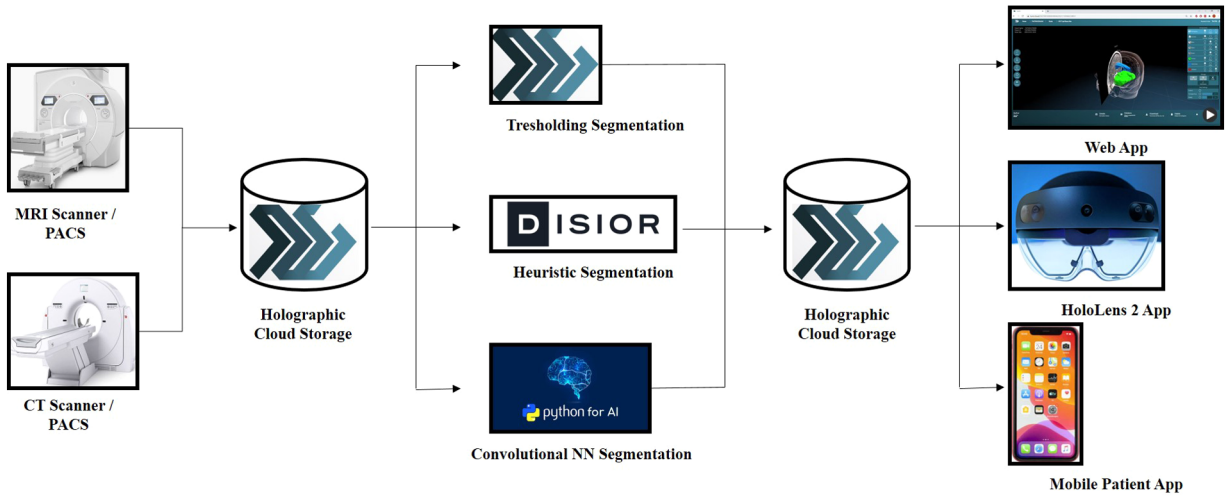


FIGURE 2
Schematic overview of the MxR system in all its components, starting from the imaging source, passing through the cloud-based storage and the strategies used for holographic rendering and ending with the visualization methods of the MxR output.

Surgical strategy

To manage the increased intracranial pressure caused by an occluded hydrocephalus due to the mass effect, a right frontal EVD was inserted a day prior to the removal of the tumor.

Based on the combined analysis of 2D images and, most importantly, the 3D model, the patient underwent a parietal craniotomy with a subsequent interhemispheric transtentorial approach to the fourth ventricle. The MxR-3D visualization of the venous anatomy and the relationship of the tumor with the ventricular system was decisive in choosing a right parietal craniotomy to release maximally the right parietal vein frontally. This approach granted a straight trajectory posteriorly from the right parietal vein and right in the axis of the tumor. The 3D holographic visualization greatly helped in planning and rehearsing a surgical trajectory that could avoid the risk of injury of the basal cerebral veins. Furthermore, surgical planning with the 3D hologram endorsed the need to open the tentorium parallel to the straight sinus to achieve better intraoperative vision and control of the tumor (Figure 4). Patient positioning was designed according to the MxR planning, and the patient was positioned three-quarters prone with the head secured in a skull clamp slightly elevated and tilted toward the floor. A linear occipito–parietal incision along the midline was the approach chosen.

To confirm gross–total resection of the tumor, an intraoperative MRI was performed, revealing 98% removal of the lesion with a small remnant at the tectal level, which was too adherent to the great vein of Galen to be safely removed.

Discussion

Fourth ventricular tumors account for 1%–5% of all intracranial neoplastic lesions and represent two-third of the lesions of the ventricular system (1, 2). To date, the classical and oldest approach to removing tumors of the fourth ventricle is the median suboccipital approach, which allows access to lesions of the cerebellar hemispheres, vermis, and medulla (1, 13). This approach is beneficial for extended anatomical exposure, enabling good surgical orientation and intraoperative manipulation. However, it is potentially associated with more tissue injury, the risk of sinus damage, and postoperative cerebrospinal fluid leakage (14). Another relevant surgical approach to lesions of the fourth ventricle is the telovelar approach, which exploits the natural space of the cerebellomedullary fissure to expose the tela choroidea and the inferior medullary velum (14, 15). The telovelar approach provides good access to the lateral recess, but exposing the posterolateral and the superolateral recesses is challenging and may require the removal of the cerebellar tonsils (14). The median suboccipital and telovelar approaches may be performed in a standard fashion or slightly modified depending on the anatomical presentation and extension of a fourth ventricular tumor. In order to tailor the surgical approach in a lesion-specific manner, an accurate pre-operative study of the anatomy of the lesions and their relationship with the surrounding structures is fundamental. In this regard, implementing augmented reality and modern 3D visualization techniques in neurosurgery has been qualitatively beneficial (16). Furthermore, applying MxR to surgical planning of complex cases gives the benefit of a non-threatening environment for less experienced surgeons, who can rehearse the procedure, foresee intraoperative difficulties, and increase surgical confidence (4).

The authors reported a complex case for which the use of MxR with an interactive patient-specific 3D anatomical model in adjunct with routine 2D images proved advantageous in gathering a better anatomical understanding and customizing the surgical approach and the intraoperative trajectory. Furthermore, surgical planning with MxR allowed a distinct patient positioning and a more realistic pre-operative analysis of the intraoperative surgical risks and challenges.

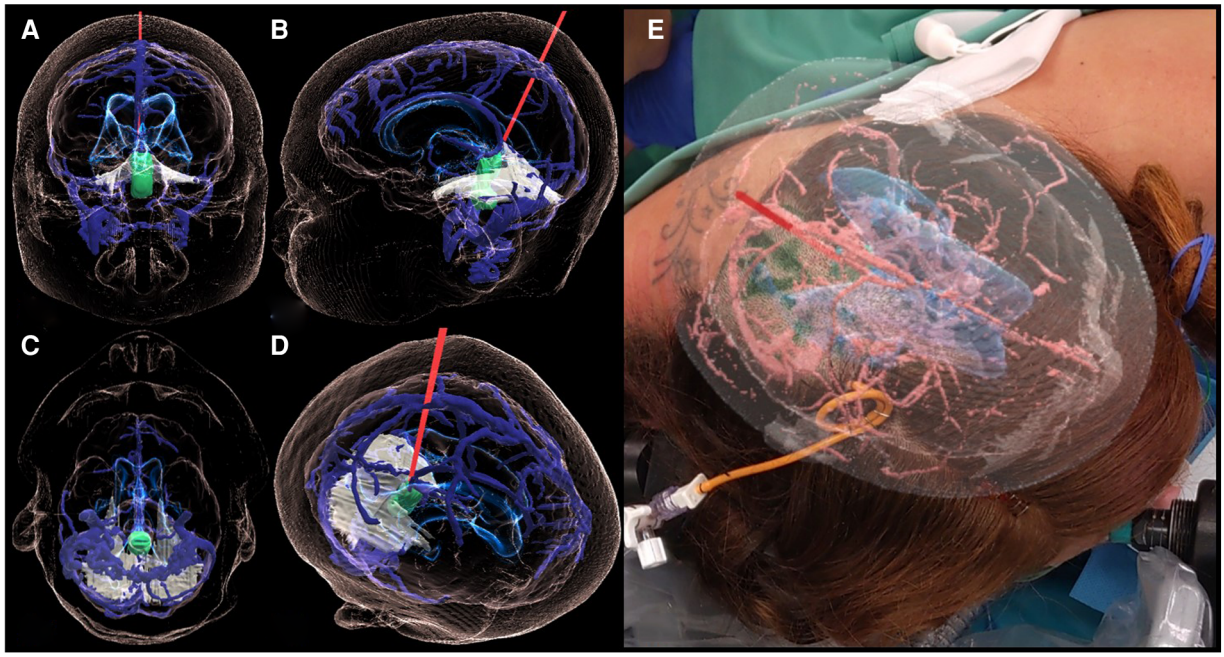


FIGURE 3
3D holographic rendering of the IV-ventricular lesion (green) and the intraoperative trajectory (red line) on the coronal (A), sagittal (B), axial (C), and a free (D) plane. Case report-specific application of MxR in the intraoperative environment showing the view of the operator wearing the MxR glasses and interacting with the hologram while matching it with the patient's head (E).

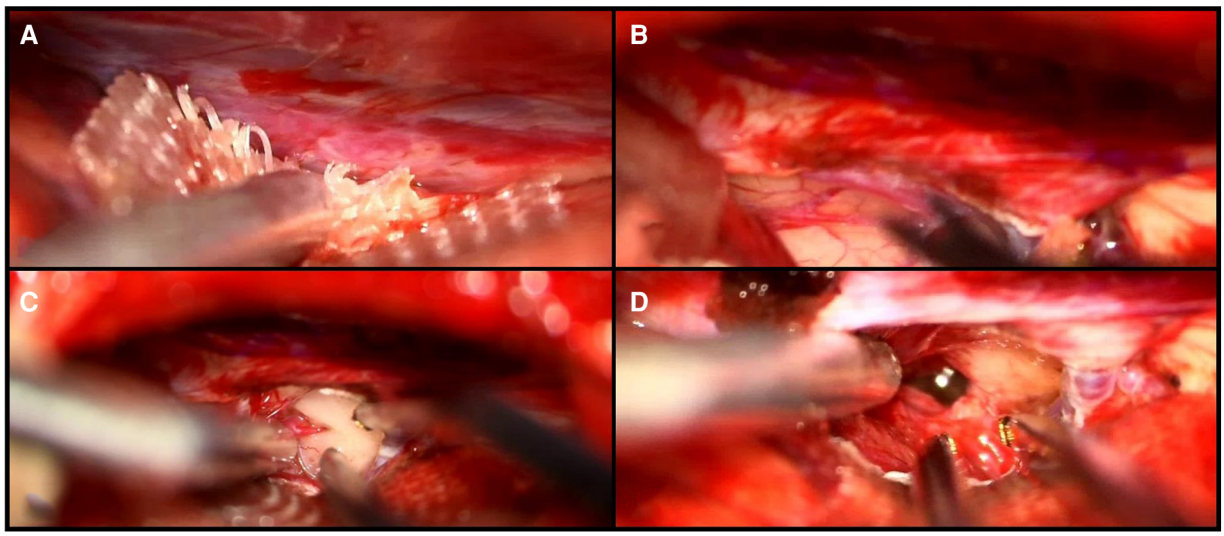


FIGURE 4
Intraoperative visualization of the falcotentorial margin (A), the small opening of the tentorium (B), the presentation of the tumor (C), and the IV ventricle after resection of the lesion (D).

Conclusion

The application of advanced 3D visualization with a novel MxR system to the surgical planning of a complex fourth ventricular lesion proved relevant to designing the best surgical approach and trajectory to better identify potential intraoperative challenges and rehearse the patient-specific anatomy. The present case report endorses the implementation of advanced 3D visualization in routine perioperative practice.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent for the publication of any potentially identifiable images or data included in this article was obtained from the individual(s).

Author contributions

EC and TD have made a substantial contribution to the concept and design of the article. LR and DB have helped with the acquisition and interpretation of data for the article. EC drafted the article. All authors contributed to the article and approved the submitted version.

Funding

This work is part of "SURGENT" under the auspices of University Medicine Zurich/Hochschulmedizin Zürich, SURGENT project (UZH—Hochschulmedizin Zürich).

Conflict of interest

TD is a co-founder and the CMO of Augmedit bv, an augmented reality company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsurg.2023.1227473/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Authors provided the possibility to visualize the hologram with a smartphone by scanning the QRCode provided below. The hologram is protected by a PIN: 715-899.

References

1. Sufianov R, Pitskhelauri D, Bykanov A. Fourth ventricle tumors: a review of series treated with microsurgical technique. *Front Surg.* (2022) 9:915253. doi: 10.3389/fsurg.2022.915253

2. Cohen AR. *Surgical disorders of the fourth ventricle*. Cambridge, MA: Blackwell Science (1996). p. 94–143.

3. Bertalanffy H, Krayenbuhl N. Ventricular tumors. In: Winn HR, editor. *Youmans neurological surgery*. 6th ed. New York: Elsevier (2011). p. 1537.

4. Zhang C, Gao H, Liu Z, Huang H. The potential value of mixed reality in neurosurgery. *J Craniofac Surg.* (2021) 32(3):940–3. doi: 10.1097/SCS.0000000000007317

5. Incekara F, Smits M, Dirven C, Vincent A. Clinical feasibility of a wearable mixed-reality device in neurosurgery. *World Neurosurg.* (2018) 118:e422–7. doi: 10.1016/j.wneu.2018.06.208

6. Cartucho J, Shapira D, Ashrafian H, Giannarou S. Multimodal mixed reality visualisation for intraoperative surgical guidance. *Int J Comput Assist Radiol Surg.* (2020) 15(5):819–26. doi: 10.1007/s11548-020-02165-4

7. Chiacchiaretta P, Perrucci MG, Caulo M, Navarra R, Baldiraghi G, Rolandi D, et al. A dedicated tool for presurgical mapping of brain tumors and mixed-reality navigation during neurosurgery. *J Digit Imaging.* (2022) 35(3):704–13. doi: 10.1007/s10278-022-00609-8

8. Jean WC, Yang Y, Srivastava A, Tai AX, Herur-Raman A, Kim HJ, et al. Study of comparative surgical exposure to the petroclival region using patient-specific, petroclival meningioma virtual reality models. *Neurosurg Focus.* (2021) 51(2):E13. doi: 10.3171/2021.5.FOCUS201036

9. Zawy Alsofy S, Sakellaropoulou I, Nakamura M, Ewelt C, Salma A, Lewitz M, et al. Impact of virtual reality in arterial anatomy detection and surgical planning in patients with unruptured anterior communicating artery aneurysms. *Brain Sci.* (2020) 10:963. doi: 10.3390/brainsci10120963

10. van Doormaal JAM, Fick T, Ali M, Köllen M, van der Kuijp V, van Doormaal TPC. Fully automatic adaptive meshing based segmentation of the ventricular system for augmented reality visualization and navigation. *World Neurosurg.* (2021) 156:e9–24. doi: 10.1016/j.wneu.2021.07.099

11. Fick T, Doormaal JAM van, Tosic L, van Zoest RJ, Meulstee JW, Hoving EW, et al. Fully automatic brain tumor segmentation for 3D evaluation in augmented reality. *Neurosurg Focus.* (2021) 51(2):E14. doi: 10.3171/2021.5.FOCUS21200

12. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging.* (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001

13. Yaşargil MG. *Microneurosurgery in 4 volumes IVB*. Stuttgart, NY: Thieme (1996). p. 63–4.

14. Mussi AC, Matushita H, Andrade FG, Rhoton AL. Surgical approaches to IV ventricle—anatomical study. *Childs Nerv Syst.* (2015) 31:1807–14. doi: 10.1007/s00381-015-2809-0

15. Tamriover N, Ulm AJ, Rhoton AL Jr, Yasuda A. Comparison of the transvermian and telovelar approaches to the fourth ventricle. *J Neurosurg.* (2004) 101:484–98. doi: 10.3171/jns.2004.101.3.0484

16. Mikhail M, Mithani K, Ibrahim GM. Presurgical and intraoperative augmented reality in neuro–oncologic surgery: clinical experiences and limitations. *World Neurosurg.* (2019) 128:268–76. doi: 10.1016/j.wneu.2019.04.256

frontiers | Frontiers in Surgery

TYPE Review
PUBLISHED 24 August 2023
DOI 10.3389/fsurg.2023.1241923

OPEN ACCESS

EDITED BY
Mohammad Mofatteh,
Queen's University Belfast, United Kingdom

REVIEWED BY
Bipin Chaurasia,
Neurosurgery Clinic, Nepal
Marco Mammi,
S. Croce and Carle Cuneo Hospital, Italy
Michael Young,
Beth Israel Deaconess Medical Center and Harvard Medical School, United States

*CORRESPONDENCE
Alireza Zali
✉ dr_alirezazali@yahoo.com

RECEIVED 17 June 2023
ACCEPTED 11 August 2023
PUBLISHED 24 August 2023

CITATION
Kazemzadeh K, Akhlaghdoust M and Zali A (2023) Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery.
Front. Surg. 10:1241923.
doi: 10.3389/fsurg.2023.1241923

Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery

Kimia Kazemzadeh[1,2], Meisam Akhlaghdoust[2,3,4] and Alireza Zali[2,3,4]*

[1]Students' Scientific Research Center, Tehran University of Medical Sciences, Tehran, Iran, [2]Network of Neurosurgery and Artificial Intelligence (NONAI), Universal Scientific Education and Research Network (USERN), Tehran, Iran, [3]Functional Neurosurgery Research Center, Shohada Tajrish Comprehensive Neurosurgical Center of Excellence, Shahid Beheshti University of Medical Sciences, Tehran, Iran, [4]USERN Office, Functional Neurosurgery Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Neurosurgical practitioners undergo extensive and prolonged training to acquire diverse technical proficiencies, while neurosurgical procedures necessitate a substantial amount of pre-, post-, and intraoperative clinical data acquisition, making decisions, attention, and convalescence. The past decade witnessed an appreciable escalation in the significance of artificial intelligence (AI) in neurosurgery. AI holds significant potential in neurosurgery as it supplements the abilities of neurosurgeons to offer optimal interventional and non-interventional care to patients by improving prognostic and diagnostic outcomes in clinical therapy and assisting neurosurgeons in making decisions while surgical interventions to enhance patient outcomes. Other technologies including augmented reality, robotics, and virtual reality can assist and promote neurosurgical methods as well. Moreover, they play a significant role in generating, processing, as well as storing experimental and clinical data. Also, the usage of these technologies in neurosurgery is able to curtail the number of costs linked with surgical care and extend high-quality health care to a wider populace. This narrative review aims to integrate the results of articles that elucidate the role of the aforementioned technologies in neurosurgery.

KEYWORDS

neurosurgery, artificial intelligence, augmented reality, robotics, virtual reality

Introduction

In contemporary society, artificial intelligence (AI) is widely perceived as an integral facet of human existence and has assumed a substantial function in the realm of medicine, encompassing domains such as diagnosis, prognosis, and treatment in recent decades. AI, in essence, represents the emulation of human cognitive faculties by machines, particularly computer systems, and was initially conceptualized in 1950. The emergence of deep learning (DL) and machine learning (ML) has provided a newfound opportunity to leverage personalized medicine and has concomitantly augmented the utilization of AI in medical procedures (1, 2).

Other technologies can be used in medicine as well. For instance, the field of Robotics is characterized by rapid progression, which is concurrently accompanied by advancements in AI and ML, ultimately leading to a metamorphosis of the medical practice (3). Augmented reality (AR) technology serves to enhance the physical world by rendering visible data that would otherwise be imperceptible to the human eye. In comparison to its virtual reality (VR) counterpart, AR technology boasts superior flexibility, albeit with a caveat of incomplete immersion on the part of the patient and physician. VR technology, on the other hand,

entails complete submersion into a virtual environment facilitated by specialized equipment. In this scenario, the patient or physician is afforded the most comprehensive visualization attainable, only restricted by the boundaries of the virtual world (4).

Neurosurgery is an arduous vocation that demands a plethora of skills and attributes from its practitioners. To achieve success in this field, neurosurgeons must undergo extensive training, exhibit an appropriate degree of manual dexterity, possess acceptable hand-eye coordination, effectively engage in decision-making processes, show compassion, communicate well with patients and their families, and work well within a team (5). The efficacy and outcomes of surgical procedures are partially contingent upon the proficiency of the operating surgeon, leading to variations in patient experiences and results across different settings. While successful surgeries have the potential to produce advantageous outcomes for patients, errors can yield unfavorable consequences and, at times, even harmful effects (6). For example, a notable proportion of medical inaccuracies that occur in neurosurgery are technical in nature, and pertain to the surgical procedures themselves, which can be obviated. This underscores the significance of practical measures intended to enhance the positive result of neurosurgical interventions, and diminish related inaccuracies, with the ultimate goal of delivering optimal healthcare to patients. Recent technological advancements have narrowed the divide between machines and humans, and have empowered computers to emulate, and surpass, innate human intelligence, thus resulting in the emergence of AI as well robotics, VR, and AR (7, 8).

In this study, we aimed to review the role of AI, VR, robotics, and AR in neurosurgery and clarify the promising perspective of neurosurgery with the help of the aforementioned technologies.

Artificial intelligence and neurosurgery

The utilization of computer systems to stimulate critical thinking and intelligent behavior was originally expounded upon by Turing in 1950 (9). Six years later, McCarthy provided an explanation of artificial intelligence, outlining it as the engineering and science of generating intelligent machines (10, 11). As time progressed, the development of AI through the use of more intricate algorithms resulted in performance that more closely resembled that of the human brain. Within the field of medicine, two subfields of AI, namely DL and ML, have emerged with significant roles. In ML, pattern identification is utilized to enable the analysis of specific situations, allowing for subsequent learning and the application of acquired data to future same scenarios. Also, this tool can be leveraged in the context of individualized patient care and clinical decision-making. DL, on the other hand, represents an advanced form of ML that operates more closely to the human brain. Algorithms are employed to establish an artificial neural network (ANN) that is able to make decisions and learn autonomously (12–14). In the past five decades, both DL and ML have played a noticeable part in the advancement of AI in the field of medicine. The utilization of predictive models has facilitated medical diagnosis, prediction of therapeutic responses, and preventative medicine (15). The employment of AI has resulted in a reduction of errors and costs of care, and has provided valuable context for patient care, thus resulting in several benefits (16).

AI is capable of enhancing the precision of treatment and diagnosis in the field of neurosurgery, while also providing neurosurgeons with timely and effective tools for pre-, post-, and intraoperative care. Neurosurgeons benefit from AI's ability to detect subtle malformations and abnormalities from clinical data as well as neuroradiological images that may elude even highly-trained human eyes. DL, a type of ML, utilizes neural networks with multiple layers of learning algorithms (17).

In the pre-operative phase of neurosurgical procedures, artificial intelligence (AI) can provide valuable assistance to surgeons by aiding in the diagnostic process, selecting appropriate patients for treatment, and guiding patients towards informed decisions (18). During the intra-operative phase, the technology of AI significantly improve the surgical performance of neurosurgeons as well as help to minimize the occurrence of errors in their procedures. In the postoperative phase, AI is utilized to accurately predict patient's prognosis, identify potential complications that may arise after surgery, and track pertinent data that is used to enhance the quality of aftercare and patient recovery. By leveraging the predictive capabilities of AI in the postoperative phase, pre-operative planning can be optimized to facilitate better patient care and decrease overall related costs. For instance, machine learning techniques can be employed to classify, regress, and cluster large data sets, thereby enabling the identification of risk factors and the prediction of surgical complications including cardiac and wound-related issues, as well as mortality rates among patients undergoing cervical discectomy as well as posterior lumbar spine fusion procedures (19, 20).

The utilization of ubiquitous and high-resolution radiological imaging in combination with electrophysiological data has become the preferred methods for providing neurosurgeons with unparalleled and noninvasive access to intracranial regions. In the field of neurosurgical medicine, effective decision-making requires the careful study, retention, analysis, and interpretation of a large quantity of complex and dynamic data. Typically, neurosurgeons rely on their clinical expertise and empirical evidence to formulate decisions and predict prognoses (21, 22).

The potential of AI in predicting the disease progression has been demonstrated through the use of DL algorithms trained on magnetic resonance imaging (MRI) data from a large, multi-institutional dataset. This approach has shown promise in replacing the need for invasive tissue sampling in predicting the progression of glioma in a non-invasive manner. The application of ML in this context has the potential to enhance the capitalization of existing data (23, 24).

In the context of temporal lobe epilepsy (TLE), as the most prevalent surgically remediable and pharmacoresistant type of epilepsy among adults, the performance of artificial intelligence (AI) has been found to surpass that of physicians. Specifically, AI demonstrated a 95.8% success rate in lateralising the influenced brain hemisphere, as opposed to the 66.7% demonstrated by

physicians, when utilising functional MRI data (25). This outcome is of particular significance, as an uncertain localisation of the epileptogenic zone is able to pose a noteworthy challenge in terms of allocating patients who are eligble to proper surgeries. Therefore, the utilization of AI in this context may have the potential to greatly enhance patient outcomes (26).

There exist additional instances in which artificial intelligence (AI) was utilized for the categorization and diagnosis of neurosurgical issues without the aid of radiological input. Specifically, AI exhibited a substantially heightened accuracy in discerning between single cells vs. multiunit spike clusters from electroencephalography recordings of twelve epilepsy patients who necessitated the implantation of chronic intracranial depth electrodes (27, 28). Due to its ability to concurrently utilize multiple variables, a capability that surpasses that of a human operator, AI can take into account numerous factors when planning treatment. As such, a study was conducted, which involved the creation of an artificial neural network, comprising of eleven clinical inputs, in order to train the algorithm for the survival rate prediction of patients with traumatic brain injuries (TBI). In addition, the performance of ML in terms of accuracy and sensitivity was superior to that of neurosurgeons and neurosurgery residents, and it was also more specific (29).

However, it is nothworthy that AI, ML, and DL are not imbued with any mystical properties. Rather, they represent a set of advanced statistical algorithms and mathematical models (which frequently depend on recursive functions) that can now be readily incorporated into everyday applications owing to the augmentation of computational capabilities. In continue, we will disscuss some examples of how AI can help in pre-, post-, and intraoperative care:

- Pre-operative phase: during the pre-operative phase of neurosurgery, AI has the potential to provide aid to surgeons in diagnosing the condition, the determination of patients for the appropriate treatment, and the facilitation of informed decision-making by patients (18). AI algorithms have been employed for automated neoplasm segmentation, localization of epileptogenic zones, identification of suitable candidates for epileptic surgery, prognostication of symptomatic cerebral vasospasm following aneurysmal subarachnoid hemorrhage, as well as estimation of tissue damage post-acute ischemic stroke (8). For instance, the categorization of tumor and epilepsy can be subjective, thereby leading to disparities in the decision-making among neurosurgeons. Upon preparing a robust outline and framework, algorithms utilizing AI can mitigate the subjective interpretation of the data and consequently diagnose medical conditions necessitating neurosurgical procedures (30, 31).
- Intra-operative phase: during the intra-operative phase of neurosurgical procedures, AI has the potential to amplify the surgeons' performance and mitigate some errors that are commonly encountered during neurosurgical procedures (31).

  The current traditional approach to performing intraoperative tissue biopsy involves transporting the tissue to a laboratory, processing it, and preparing specimens with the assistance of skilled laboratory professionals before pathologists interpret the results. This process has been in use for over a century and is both time-consuming and resource-intensive. However, there have been recent developments in the utilization of AI technology during the intraoperative phase of neurosurgery. For instance, Hollon et al. designed a label-free optical imaging workflow that can predict diagnosis of tumours in approxiamately real-time automatically. The tumor diagnosis techniques are able to predict the tumour diagnosis in less than 150 s, which is significantly faster than conventional methods that can take up to 30 min. Furthermore, their overall accuracy rate of 95% is marginally better than regular histology workflow, which has an accuracy rate of 94% (32–34).
- Post-operative phase: given that patients may necessitate multiple visits to different geographic locations like inpatient wards, outpatient clinics, pharmacies, intensive care units, emergency departments as well as laboratories, telemedicine possesses the capacity to curtail unnecessary travel for both patients and healthcare professionals (35, 36). The implementation of telemedicine services is held in high regard by both healthcare providers and patients and has the potential to enhance patient outcomes in the postoperative phase, particularly in regions with limited geographic access. The majority of patients welcomed postoperative videoconferencing which was found to be as effective and safe as in-person clinic visits for those who had elective neurosurgery (37).

Robotics and neurosurgery

Robotics, a fast-moving discipline, is transforming neurosurgery practice with advances in machine learning and artificial intelligence. Utilizing robotics in neurosurgery can efficiently omit mechanistic errors, decrease operation time, and prepare more extended respective margins using minimal-access operation. In this way, minimal complications and great surgical results will be achieved (3). Interestingly, it was reported that the first use of robotics in operation was a neurosurgical biopsy. The Unimation PUMA (Programmable Universal Machine for Assembly) 200 robot was utilized in a 52-year-old man to position a needle guided by a CT scan in a stereotactic biopsy of an intracerebral lesion (38). Then, the aforementioned robot was used as an assistant to retract delicate neural structures while resecting low-grade thalamic tumors among children (39). NeuroMate robot was the first FDA-approved device specifically generated for neurosurgical use (3).

Current available robotic systems used in surgery have three subtypes: master-slave, semi-active, and active (40). Active robotic systems can work autonomously and perform preprogrammed tasks. However, master-salve systems depend on surgeon input and lack preprogramming. Semi-active ones are hybrid in which surgeon inputs complement preprogrammed elements of the system (41). Improved visualization for surgeons, greater precision, as well as a decrease in fatigue are some benefits of using robotic systems in surgery (42). Regarding

limitations, there are some concerns about cost, hardware maintenance, and sterilization (43).

Generally, robotic systems can be used in neurosurgery for procedures with restricted operative spaces. Anatomical localization, surgeon's hand stabilization, placement of pedicle screws in spinal procedures, and plans to access deep brain targets are some robot applications in neurosurgery (43–45). Pathfinder, SpineAssist, Renaissance, Neuromate, and NeuroArm are common robotic systems utilized in neurosurgery (43, 46). Robotic assistant is more common among other surgeons, but specific aspects of neurosurgery including the technical and microsurgical nature of procedures as well as the history of its innovation in stereotaxy help it for being well incorporated with robotic assistance (45).

In 2022, Singh et al. claimed that the usage of robotic systems in neurosurgery is in its infancy yet. Almost 30 per 100 neurosurgical departments use robotic cranial methods and 40 per 100 departments use robotic spinal methods. While examining the possible application of robotic systems in neurosurgery, 13 clinical trials seemed to be applicable, and none of them were completed (47). Various indications for robotic usage during neurosurgery are identified. For instance, multiple studies claimed that screw placement assisting by robots during spinal surgery is accurate and safe (48–50) and cause less radiation exposure as well as fewer facet joint violations while screw placement in comparison to traditional surgery methods (51, 52). Three different systematic reviews and meta-analyses reported that the usage of robotic systems can lead to a high accuracy compared with conventional free-hand strategies (53–55).

In continue, we elucidate some examples of robotic systems used in neurosurgery in detail:

- The telesurgical robot: In this particular variety of robot, surgeons exercise remote control over the robot's actions. The NeuroArm, hailing from the University of Calgary in Canada, displays remarkable potential. It constitutes an MRI-compatible robotic arm that emulates a surgeon's manual gestures. It harnesses piezoelectric motors and boasts of eight degrees of freedom (DOF). This technology has undergone continuous development, with bespoke microsurgical instruments (equipped with force-sensing as well as force-calibration features) recently incorporated into the robotic arm's arsenal. Encouraging preliminary experiments conducted on rats have paved the way for its subsequent deployment on human subjects. Additionaly, It is the first robot to furnish the neurosurgeon with tactile feedback while simultaneously being operated remotely from a workstation located outside the operating room. Reports indicate that it has already been employed in over 1,000 neurosurgical methods, such as MRI-guided tumor biopsies, hematoma evacuations, and microsurgical dissection (43, 56, 57).
- The supervisory surgeon-controlled robot: the robotic system supports surgeons in executing accurate procedures. The PUMA robots, introduced in the 1980s, have become the most prevalent neurosurgical robots to date. Additional robots, such as the Pathfinder and Minerva robots, have been subsequently developed. These robots mainly undertake stereotactic duties, without or with a frame, and have advanced from directing biopsy needles as well as depth electrodes to inserting and devising pedicle screws in the spine (3).
- Handheld shared/controlled systems: The collaboration between surgeons and robots occurs at the site where they jointly dissect and manipulate the structures of brain through instruments. This allows the precise robot actions to complement the manual dexterity and manipulative skills of neurosurgeons, resulting in a synergy of capabilities. It can be likened to the optimal combination of two distinct worlds. The Steady Hand System, developed at John Hopkins University, is a representative instance of the few systems currently in development. This instrument, which is held by both the surgeon and the robot, permits finer dissection and eliminates tremor and muscle fatigue. Other devices, such as the Evolution 1, can be controlled for endoscopic procedures. The NeuRobot, developed at Shinshu University in Matsumoto, Japan, is a remotely operated device that comprises an endoscope equipped with twin tissue forceps, which can assist in tumour resection (3).

Augmented and virtual reality in neurosurgery

Virtual reality

Virtual reality (VR) is a process that entails the user's immersion in a system obscuring the natural world as well as generating a virtual realm for users' experience. VR can be classified as either immersive or non-immersive, based on whether the virtual world is generating as a powerful substitute for the real world or virtual environment, respectively (58).

Although the concept of VR was utilized for panoramic viewing as early as the eighteenth century, it was not until 1929 that the first VR simulator, specifically a flight simulator, was invented. However, the term "Virtual Reality" was not coined until 1987 (59). The development of VR technology can be traced back to the innovative contributions of Tom Furness, an electrical engineer, who was affiliated with the United States Air Force (60). Furness's contributions were groundbreaking and earned him the moniker of "The Godfather of Virtual Reality" (61). The introduction of a VR system in the field of medicine was pioneered by Robert Mann in the field of orthopedics for the first time. Subsequently, the head-mounted device (HMD) was introduced in the 1980s. Although VR had been utilized for the arachnophobia treatment in 1998, this marked the first reported use of the technology in pathology treatment (62, 63). It is noteworthy, however, that the first recorded use of VR in the therapy of neurosurgical disorders is a recent occurrence. For the first time, Clarke utilized the NeuroTouch neurosurgical simulator to excise a Left frontal meningioma in 2009 (see **Figure 1**) (64).

Throughout the literature, the terminology of Virtual Reality has been utilized interchangeably to encompass AR and Mixed Reality (MR). Anyway, it is crucial to note that MR, AR,

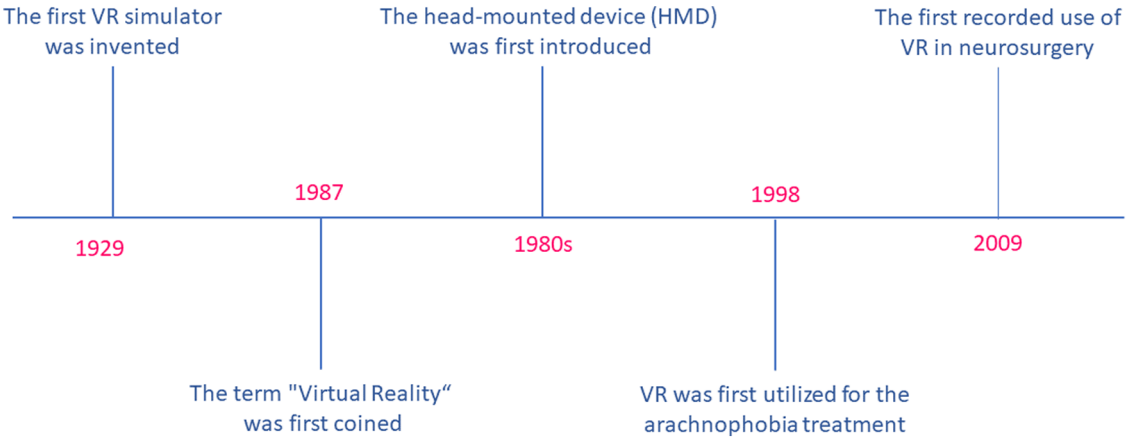


FIGURE 1
Virtual reality-related timeline.

and VR are fundamentally distinct technologies. VR specifically pertains to computer-generated three-dimensional (3D) immersive environments, while AR involves the projection of computer-developed images onto real-world images. On the other hand, MR entails the projection of virtual objects into the physical world, where the objects demonstrate spatial awareness and responsiveness (65). The early usage of VR technology yielded adverse effects such as temporarily impaired vision, a lack of sense of presence, vomiting, and nausea. However, these limitations were primarily attributed to the technical constraints of the VR technology available during that time, similar to the human eye's inability to focus in-depth on a 3D-rendered image (65–67). Similarly, high-definition 4 K 3D exoscopes have also been reported to produce similar limitations during surgical procedures (68, 69).

Cerebrovascular and neuro-interventional surgeries heavily rely on advanced neuroimaging strategies for operative prognostication and decision-making. The clinical use of VR enhances the diagnostic efficacy and accuracy of the aforementioned techniques (70). Hybrid angio-suites enable neurointerventionists to create an immersive VR model according to patient-specific anatomy, which improves crisis resource management, training, and procedural skills (71). VR technology has been of great utility due to its specific metric-based performance assessment which is outside the angio suite as well as its ability to conduct complex neuro interventions, such as mechanical thrombectomy, accompanied by the similar set of principles as in live patients (72). Surgeons can also benefit from gain access to a VR-based patient-specific model for better planning management or diagnosing strategies, along with planning complex hybrid or combined procedures requiring a combination of conventional and interventional surgical methods (73–75).

A primary obstacle encountered by individuals training to become neurosurgeons is performing procedures with bimanual dexterity within a narrow corridor that is bound by intricate and essential non-resilient bones and neurovascular structures. Neuronavigation is highly relied upon by those in training for better planning their localization and approach. However, it is not an appropriate means of advancing spatial reasoning abilities, also an excessive dependence on neuronavigation can hinder the skills development (58). Additionally, as the operation progresses, the brain shifts gradually, rendering the preoperative imaging used in the navigation system less accurate and useful. To circumvent this issue, an intraoperative brain imaging system (IBIS) was created that recognized any discrepancies between preoperative imaging and intraoperative ultrasound. Through the use of IBIS, intraoperative stimulation is altered in real-time, as well as inaccuracies are updated using AR (76, 77). The utilization of VR in training and simulation has proven to be a superior alternative in reducing operative stress duration and cognitive load, as well as enhancing efficacy for novice neurosurgeons, according to a study (78). There is a diverse range of available VR tools for neurosurgical training and education, including a multifunction head-mounted display (HMD) such as Microsoft HoloLens and Google Glass, in addition to haptic feedback tools such as Procedicus Vascular Interventional System Trainer (VIST), Immersive touch, NeuroVR, and synthetic tissue simulators like SynDaver, Creaplast, Thomas Jefferson University Durotomy Repair Module, and iDU optics 3D-printed models. Additionally, there are VOSTARS (video and optical see-through AR surgical systems) HMD-based surgical navigation platforms, as well as operating planning devices such as Dextroscope, Surgical Theatre, Synaptive Medical, and VPI Reveal. The use of computer simulation and VR has extended to various fields, including pilot training, medicine, and military, as a means to alleviate potential dangers by preparing a virtual simulator as well as haptic and visual feedback (79–81). Physics-based simulators pose a challenge due to their high computational demands and requisition of resources, both in terms of computing skills and software, to prepare haptic and visual feedback, along with formal trainings in 3D immersive simulation (79, 82, 83). Amongst the various displays in virtual reality (VR), HMDs offer the greatest engagement, with other

displays such as Google Glass featuring an OLED or LED display with a high refresh rate of 120 Hz as well as latency time of approximately 20 milliseconds (84). Additionally, VR plays a crucial role in tele-proctoring, facilitating the training of surgeons on complex techniques and procedures, independent on their geographical location (85). Also, immersive technologies have a profound effect on global virtual connections, enabling middle- and low-income countries to enhance their potential applications particularly while ongoing pandemics, like COVID-19. In future outpatient neurosurgery consultations, telemedicine is expected to have a crucial role as it facilitates the interaction between the surgeon and patients in a "merged reality" space, thereby enabling manual, visual, and verba interactions between them. While utilizing VR technology as an educational tool for neurosurgeons, specific quality control standards must be followed, including appropriate sound quality, high-resolution images, internet speed, high processing power, visual and haptic feedback, tissue fidelity, and organ structure. The main benefits of using VR technology in the field of neurosurgery training over animal and human models are its non-invasive nature, low cost, limitless repetition ability, as well as the extensive diversity and variety in cases which can be simulated. By the way, the ever-present concern is the realism and resolution of the VR technology (86–89). The employment of VR environments presents an opportunity to accurately gauge the performance of surgeons, evaluate their proficiency, and monitor their progression during training. In addition, the implementation of AR HMD visualization has been shown to elicit greater levels of enthusiasm and enjoyment in the learning process, particularly among younger surgeons (90–92).

Augumented reality

AR is a novel technological advancement that overlays 3-D virtual text or objects onto tangible objects (93). Divergent from VR, which generates a wholly fabricated environment, AR presents both virtual and tangible objects, thereby producing a semi-immersive experience for users. Giglioli et al. claimed that AR amplifies user perception of reality by integrating virtual content into the tangible world and displaying it simultaneously and in real-time. Additionally, they elucidate that AR encompasses an array of tools and methods that supplement physical reality with additional information (94). The implementation of AR technology in healthcare has been adopted by the field of neurosurgery at an early stage. This particular medical specialty depends highly on imagery for the purposes of preoperative planning and intraoperative neuronavigation. The present neuronavigation system lonely projects 2-D images (coronal, sagittal, and axial) on a computer screen, as explained by Pandya et al. In order to successfully navigate the 2D images into a 3D format, the surgeon must engage in a mental transformation and be able to project the visualized data onto the patient's view. However, this task creates a significant interruption in the surgical workflow as the neurosurgeon must frequently switch between the computer screen and the surgical field (70, 95).

The American Brain Tumor Association has reported that in 2013, the Central Brain Tumor Registry of the United States approximated 69,720 novel cases of primary brain tumors. Johns Hopkins Medicine suggests that the primary objective of operation for metastatic brain tumors is to remove and debulk the whole tumor during simultaneously preserving neurological function (96). Currently, the use of image-guided neurosurgery (IGNS) plays a critical role in achieving maximal brain tumor resection. Deng et al. have elucidated that IGNS utilizes patients'preoperative images to track the tumor's position against the preoperative images while surgery. Nevertheless, they have postulated that by using IGNS, surgeons must switch views between the surgical field and the computer screen every time he/she desires to control his/her relative position on the preoperative images and patient's brain (97). The division of the preoperative images on the computer screen into three distinct images (coronal, axial, and sagittal) necessitates the surgeon's mental amalgamation of these images to create a singular three-dimensional composite image. It has been posited that the repeated alteration of perspectives while surgery hinders the surgeon's workflow. Currently, numerous AR system prototypes are undergoing testing, specially for brain tumors management. Inoue et al. employed an AR system prototype together with a web camera in order to superimpose the brain tumor images in the patient's dura and skull. Similarly, Deng et al. utilized a wireless tablet computer AR neuronavigation system for operative planning and the execution of two cases in China (95, 97, 98).

Abe and colleagues conducted an experimental study on a virtual protractor with an AR system, known as VIPAR, for percutaneous vertebroplasty. The study involved the use of 5 patients and 40 spine phantoms. VIPAR was developed to provide real-time visualization of the vertebroplasty needle trajectory in a 3-D space while the procedure. Generally, percutaneous vertebroplasty is a minimally invasive procedure that is aimed at treating fractured spinal vertebrae that cause loss of function and pain. This procedure involves injecting medical grade cement into the fractured vertebra, and it significantly depends on the utilization of C-arm fluoroscopy in order to guide neurosurgeons in controlling the needle trajectory. Also, Abe et al. emphasized that while percutaneous vertebroplasty is generally considered a safe and almost easy procedure, incorrect needle placement can lead to cement leakage and neurovascular injury. Johns Hopkins Medicine has proffered that there are various risk factors which are related to vertebroplasty, such as rib or other surrounding bone fractures, hemorrhaging,, as well as cement leakage outside the bone. Upon conducting 40 spine phantom trials, Abe et al. discovered that the error of the insertion angle of the vertebroplasty needle while procedure was highly improved in comparison to present modalities. Furthermore, in these 5 VIPAR assisted percutaneous vertebroplasty procedures conducted in the clinical trial, there was a complete success rate, with no spinal pedicle breach or leakage of cement (99, 100).

Limitations

In the field of health care and medicine, AI has made significant progress. In the future, doctors and robots may collaborate to improve patient care. Nevertheless, patients may find it challenging to place their trust in a robot when it comes to surgical procedures, and it is often recommended that a neurosurgeon retain ultimate control. Traditional neurosurgeons typically dissuade the use of aforementioned technologies including AI during neurosurgical interventions. Conversely, an excessive reliance on AI may deter surgeons, particularly neurosurgeons, from mastering the necessary surgical skills (31, 101). For instance, AI necessitates an extensive dataset for its operation, thereby presenting the challenge of generating a plethora of clinically practical algorithms. This entails the storage of large-scale data, allowing for easy accessibility to abusers, thereby jeopardizing patient privacy. Numerous ethical concerns arise in this realm. Although the recording of patient data remains controversial, in the event of a misdiagnosis due to AI, moral and legal quandaries require prompt attention (102). The "black box dilemma" emerges, where both consumers and users lack comprehension concerning how the computer produces outcomes, ultimately hindering transparency in AI systems (103). One must also acknowledge that, regardless of how advanced AI becomes, it lacks human consciousness and the capacity to make conscientious and informed decisions (104).

It is highly recommended to certify and verify AI-based systems with a view to ensuring the safety of patients. Moreover, it is imperative to minimize the instances of AI system failures on patients. An additional challenge that looms ahead is the annotation of targets, given that the identification of anatomical structures can be a daunting task for even the neurosurgeons. In order to address this challenge, AI needs to be trained to recognize such intricate anatomy, in conjunction with other cutting-edge technologies, thereby enhancing accuracy in dealing with difficult targets. However, the fact remains that the bulk of data in a training set is dominated by standard cases, which makes cases with anatomical abnormalities a worrisome challenge for the future. In the context of endovascular procedures, AI is constrained by the lack of haptic feedback, which limits its potential for usage (102, 105). Nonetheless, AI can be extensively leveraged in surgeries for the elderly, but it is still incumbent upon the clinician to provide the necessary endorsement. Doctors must therefore acquire a working knowledge of computer science in order to effectively analyze and optimize the data and AI systems at their disposal (106, 107).

Conclusion

The field of AI in corporations with VR, AR, and robotics is an interdisciplinary area located at the interface of medicine, neuroscience, and engineering. In the realm of neurosurgery, they possess the potential to optimize patient outcomes. In the pre-, intra-, and postoperative stages of neurosurgery, they have the ability to enhance surgeons' skill sets. The recent technological advances in AI, VR, AR, and robotics have made it possible for humans and machines to collaborate to improve healthcare delivery. This is achieved via image acquisition, processing and interpretation, patient allocation to appropriate surgeries, intra-operative improvements, postoperative follow-up, as well as facilitating access to high-quality healthcare. However, more investigations are required to better evaluate the limitations. Also, the possibility and accessibility of the wide use of these techniques must be evaluated.

Author contributions

KK: drafting the manuscript & reviewing the literature. MA: reviewing and editing the manuscript. AZ: supervision, validation. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Burns E. What is artificial intelligence (AI)?: techtarget. Available at: https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

2. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc*. (2020) 92(4):807–12. doi: 10.1016/j.gie.2020.06.040

3. Qureshi YA, Mohammadi B. Robotic oesophago-gastric cancer surgery. *Ann R Coll Surg Engl*. (2018) 100(6 Suppl):23–30.

4. cprime. How are augmented and virtual reality used in medicine?: cprime. Available at: https://www.cprime.com/resources/blog/how-are-augmented-and-virtual-reality-used-in-medicine/#:~:text=One%20of%20the%20most%20useful,various%20scenarios%20that%20might%20occur

5. Wise J. Life as a neurosurgeon. *Br Med J*. (2020) 368.

6. Kaptigau WM, Rosenfeld JV, Kevau I, Watters DA. The establishment and development of neurosurgery services in Papua New Guinea. *World J Surg*. (2016) 40:251–7. doi: 10.1007/s00268-015-3268-1

7. Rolston JD, Zygourakis CC, Han SJ, Lau CY, Berger MS, Parsa AT. Medical errors in neurosurgery. *Surg Neurol Int*. (2014) 5(Suppl 10):S435. doi: 10.4103/2152-7806.142777

8. Mofatteh M. Neurosurgery and artificial intelligence. *AIMS Neurosci*. (2021) 8 (4):477–95. doi: 10.3934/Neuroscience.2021025

9. Holmes J, Sacchi L, Bellazzi R. Artificial intelligence in medicine. *Ann R Coll Surg Engl*. (2004) 86:334–8.

10. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metab Clin Exp*. (2017) 69:S36–40. doi: 10.1016/j.metabol.2017.01.011

11. Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. (2019) 8(7):2328. doi: 10.4103/jfmpc.jfmpc_440_19

12. Sanne A, Hoogenboom UB, Michael B. Wallace. Artificial intelligence in gastroenterology. The current state of play and the potential. How will it affect our practice and when?: mayoclinic (2020). Available at: https://mayoclinic.pure.elsevier.com/en/publications/artificial-intelligence-in-gastroenterology-the-current-state-of

13. Le Berre C, Sandborn WJ, Aridhi S, Devignes M-D, Fournier L, Smail-Tabbone M, et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology*. (2020) 158(1):76–94. e2. doi: 10.1053/j.gastro.2019.08.058

14. Alexandra T. Greenhill BRE. A Primer of AI in Medicine (2019). Available at: https://www.tigejournal.org/article/S1096–2883(19)30081–6/fulltext

15. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence-assisted gastroenterology—promises and pitfalls. *Off J Am Coll Gastroenterol*. 2019;114(3):422–8. doi: 10.1038/s41395-018-0268-4

16. IBM. What is artificial intelligence in medicine? IBM. Available at: https://www.ibm.com/topics/artificial-intelligence-medicine

17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521(7553):436–44. doi: 10.1038/nature14539

18. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman ML, et al. Machine learning and neurosurgical outcome prediction: a systematic review. *World Neurosurg*. (2018) 109:476–86.e1. doi: 10.1016/j.wneu.2017.09.149

19. Arvind V, Kim JS, Oermann EK, Kaji D, Cho SK. Predicting surgical complications in adult patients undergoing anterior cervical discectomy and fusion using machine learning. *Neurospine*. (2018) 15(4):329. doi: 10.14245/ns.1836248.124

20. Kim JS, Merrill RK, Arvind V, Kaji D, Pasik SD, Nwachukwu CC, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine*. (2018) 43(12):853. doi: 10.1097/BRS.0000000000002442

21. Emblem KE, Nedregaard B, Hald JK, Nome T, Due-Tonnessen P, Bjornerud A. Automatic glioma characterization from dynamic susceptibility contrast imaging: brain tumor segmentation using knowledge-based fuzzy clustering. *J Magn Reson Imaging*. (2009) 30(1):1–10.

22. Emblem KE, Pinho MC, Zöllner FG, Due-Tonnessen P, Hald JK, Schad LR, et al. A generic support vector machine model for preoperative glioma survival associations. *Radiology*. (2015) 275(1):228–34. doi: 10.1148/radiol.14140770

23. Yu J, Shi Z, Lian Y, Li Z, Liu T, Gao Y, et al. Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. *Eur Radiol*. (2017) 27:3509–22. doi: 10.1007/s00330-016-4653-3

24. Chang K, Bai HX, Zhou H, Su C, Bi WL, Agbodza E, et al. Residual convolutional neural network for the determination of IDH status in low-and high-grade gliomas from MR ImagingNeural network for determination of IDH status in gliomas. *Clin Cancer Res*. (2018) 24(5):1073–81. doi: 10.1158/1078-0432.CCR-17-2236

25. Chiang S, Levin HS, Haneef Z. Computer-automated focus lateralization of temporal lobe epilepsy using fMRI. *J Magn Reson Imaging*. (2015) 41(6):1689–94. doi: 10.1002/jmri.24696

26. Berg AT, Vickrey BG, Langfitt JT, Sperling MR, Walczak TS, Shinnar S, et al. The multicenter study of epilepsy surgery: recruitment and selection for surgery. *Epilepsia*. (2003) 44(11):1425–33. doi: 10.1046/j.1528-1157.2003.24203.x

27. Sinha M, Kennedy CS, Ramundo ML. Artificial neural network predicts CT scan abnormalities in pediatric patients with closed head injury. *J Trauma Acute Care Surg*. (2001) 50(2):308–12. doi: 10.1097/00005373-200102000-00018

28. Tankus A, Yeshurun Y, Fried I. An automatic measure for classifying clusters of suspected spikes into single cells versus multiunits. *J Neural Eng*. (2009) 6(5):056001. doi: 10.1088/1741-2560/6/5/056001

29. Rughani AI, Dumont TM, Lu Z, Bongard J, Horgan MA, Penar PL, et al. Use of an artificial neural network to predict head injury outcome. *J Neurosurg*. (2010) 113 (3):585–90. doi: 10.3171/2009.11.JNS09857

30. Yamashita K, Yoshiura T, Arimura H, Mihara F, Noguchi T, Hiwatashi A, et al. Performance evaluation of radiologists with artificial neural network for differential diagnosis of intra-axial cerebral tumors on MR images. *Am J Neuroradiol*. (2008) 29(6):1153–8. doi: 10.3174/ajnr.A1037

31. Lüders H, Acharya J, Baumgartner C, Benbadis S, Bleasel A, Burgess R, et al. Semiological seizure classification. *Epilepsia*. (1998) 39(9):1006–13. doi: 10.1111/j.1528-1157.1998.tb01452.x

32. Hollon TC, Pandian B, Adapa AR, Urias E, Save AV, Khalsa SSS, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med*. (2020) 26(1):52–8. doi: 10.1038/s41591-019-0715-9

33. Novis DA, Zarbo RJ. Interinstitutional comparison of frozen section turnaround time. *Arch Pathol Lab Med*. (1997) 121(6):559.

34. Gal AA, Cagle PT. The 100-year anniversary of the description of the frozen section procedure. *JAMA*. (2005) 294(24):3135–7. doi: 10.1001/jama.294.24.3135

35. Bardram JE, Bossen C. Mobility work: the spatial dimension of collaboration at a hospital. *Comput Support Coop Work*. (2005) 14:131–60. doi: 10.1007/s10606-005-0989-y

36. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. *BMC Med Inform Decis Mak*. (2012) 12(1):1–31. doi: 10.1186/1472-6947-12-1

37. Reider-Demer M, Raja P, Martin N, Schwinger M, Babayan D. Prospective and retrospective study of videoconference telemedicine follow-up after elective neurosurgery: results of a pilot program. *Neurosurg Rev*. (2018) 41:497–501. doi: 10.1007/s10143-017-0878-0

38. Kwoh YS, Hou J, Jonckheere EA, Hayati S. A robot with improved absolute positioning accuracy for CT guided stereotactic brain surgery. *IEEE Trans Biomed Eng*. (1988) 35(2):153–60. doi: 10.1109/10.1354

39. Drake JM, Joy M, Goldenberg A, Kreindler D. Computer-and robot-assisted resection of thalamic astrocytomas in children. *Neurosurgery*. (1991) 29(1):27–33. doi: 10.1227/00006123-199107000-00005

40. Peters BS, Armijo PR, Krause C, Choudhury SA, Oleynikov D. Review of emerging surgical robotic technology. *Surg Endosc*. (2018) 32:1636–55. doi: 10.1007/s00464-018-6079-2

41. Lane T. A short history of robotic surgery. *Ann R Coll Surg Engl*. (2018) 100(6 Suppl):5–7. doi: 10.1308/rcsann.supp1.5

42. Trybula SJ, Oyon DE, Wolinsky J-P. Robotic tissue manipulation and resection in spine surgery. *Neurosurg Clin*. (2020) 31(1):121–9.

43. Mattei TA, Rodriguez AH, Sambhara D, Mendel E. Current state-of-the-art and future perspectives of robotic technology in neurosurgery. *Neurosurg Rev*. (2014) 37:357–66. doi: 10.1007/s10143-014-0540-z

44. Elswick CM, Strong MJ, Joseph JR, Saadeh Y, Oppenlander M, Park P. Robotic-assisted spinal surgery: current generation instrumentation and new applications. *Neurosurg Clin*. (2020) 31(1):103–10.

45. Wang MY, Goto T, Tessitore E, Introduction VA. Robotics in neurosurgery. *Neurosurg Focus*. (2017) 42(5):E1. doi: 10.3171/2017.2.FOCUS1783

46. Ahmed SI, Javed G, Mubeen B, Bareeqa SB, Rasheed H, Rehman A, et al. Robotics in neurosurgery: a literature review. *J Pak Med Assoc*. (2018) 68(2):258.

47. Singh R, Wang K, Qureshi MB, Rangel IC, Brown NJ, Shahrestani S, et al. Robotics in neurosurgery: current prevalence and future directions. *Surg Neurol Int*. (2022) 13(373).

48. Laratta JL, Shillingford JN, Lombardi JM, Alrabaa RG, Benkli B, Fischer C, et al. Accuracy of S2 alar-iliac screw placement under robotic guidance. *Spine Deform*. (2018) 6(2):130–6. doi: 10.1016/j.jspd.2017.08.009

49. Lee NJ, Khan A, Lombardi JM, Boddapati V, Park PJ, Mathew J, et al. The accuracy of robot-assisted S2 alar-iliac screw placement at two different healthcare centers. *J Spine Surg*. (2021) 7(3):326. doi: 10.21037/jss-21-14

50. Vardiman AB, Wallace DJ, Crawford NR, Riggleman JR, Ahrendtsen LA, Ledonio CG. Pedicle screw accuracy in clinical utilization of minimally invasive navigated robot-assisted spine surgery. *J Robot Surg*. (2020) 14:409–13. doi: 10.1007/s11701-019-00994-3

51. Gao S, Wei J, Li W, Zhang L, Cao C, Zhai J, et al. Accuracy of robot-assisted percutaneous pedicle screw placement under regional anesthesia: a retrospective cohort study. *Pain Res Manag*. (2021) 2021.

52. Linden GS, Birch CM, Hresko MT, Cook D, Hedequist DJ. Intraoperative use of robotics with navigation for pedicle screw placement in treatment of pediatric high-grade spondylolisthesis: a preliminary report. *J Pediatr Orthop*. (2021) 41(10):591–6. doi: 10.1097/BPO.0000000000001947

53. Peng Y-N, Tsai L-C, Hsu H-C, Kao C-H. Accuracy of robot-assisted versus conventional freehand pedicle screw placement in spine surgery: a systematic review and meta-analysis of randomized controlled trials. *Ann Transl Med*. (2020) 8(13).

54. Li H-M, Zhang R-J, Shen C-L. Accuracy of pedicle screw placement and clinical outcomes of robot-assisted technique versus conventional freehand technique in spine surgery from nine randomized controlled trials: a meta-analysis. *Spine*. (2020) 45(2): E111–9. doi: 10.1097/BRS.0000000000003193

55. Fan Y, Du JP, Liu JJ, Zhang JN, Qiao HH, Liu SC, et al. Accuracy of pedicle screw placement comparing robot-assisted technology and the free-hand with fluoroscopy-guided method in spine surgery: an updated meta-analysis. *Medicine (Baltimore)*. (2018) 97(22).

56. Pandya S, Motkoski JW, Serrano-Almeida C, Greer AD, Latour I, Sutherland GR. Advancing neurosurgery with image-guided robotics. *J Neurosurg*. (2009) 111 (6):1141–9. doi: 10.3171/2009.2.JNS081334

57. Sutherland GR, Maddahi Y, Gan LS, Lama S, Zareinia K. Robotics in the neurosurgical treatment of glioma. *Surg Neurol Int*. (2015) 6(Suppl 1):S1. doi: 10.4103/2152-7806.151321

58. Pelargos PE, Nagasawa DT, Lagman C, Tenn S, Demos JV, Lee SJ, et al. Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery. *J Clin Neurosci*. (2017) 35:1–4. doi: 10.1016/j.jocn.2016.09.002

59. Madhavan K, Kolcun JPG, Chieng LO, Wang MY. Augmented-reality integrated robotics in neurosurgery: are we there yet? *Neurosurg Focus*. (2017) 42(5):E3. doi: 10.3171/2017.2.FOCUS177

60. Mertz L. Virtual reality pioneer tom furness on the past, present, and future of VR in health care. *IEEE Pulse*. (2019) 10(3):9–11. doi: 10.1109/MPULS.2019.2911808

61. Fiani B, De Stefano F, Kondilis A, Covarrubias C, Reier L, Sarhadi K. Virtual reality in neurosurgery:"can you see it?"—a review of the current applications and future potential. *World Neurosurg*. (2020) 141:291–8.

62. Mohammad EB, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: a randomized control trial. *Palliat Support Care*. (2019) 17(1):29–34. doi: 10.1017/S1478951518000639

63. Garcia-Palacios A, Hoffman H, Carlin A, Furness TA III, Botella C. Virtual reality in the treatment of spider phobia: a controlled study. *Behav Res Ther*. (2002) 40(9):983–93. doi: 10.1016/S0005-7967(01)00068-7

64. Clarke DB, D'Arcy RC, Delorme S, Laroche D, Godin G, Hajra SG, et al. Virtual reality simulator: demonstrated use in neurosurgical oncology. *Surg Innov*. (2013) 20 (2):190–7. doi: 10.1177/1553350612451354

65. Hu H-Z, Feng X-B, Shao Z-W, Xie M, Xu S, Wu X-H, et al. Application and prospect of mixed reality technology in medical field. *Curr Med Sci*. (2019) 39:1–6. doi: 10.1007/s11596-019-1992-8

66. Edwards PJ, King AP, Maurer CR, De Cunha DA, Hawkes DJ, Hill DL, et al. Design and evaluation of a system for microscope-assisted guided interventions (MAGI). *IEEE Trans Med Imaging*. (2000) 19(11):1082–93. doi: 10.1109/42.896784

67. Langreth R. Virtual reality: head mounted distress. *Pop Sci*. (1994) 5:49.

68. Wierzbicka M, Szyfter W, Greczka G, Gawęcki W. Otosurgery with the high-definition three-dimensional (3D) exoscope: advantages and disadvantages. *J Clin Med*. (2021) 10(4):777. doi: 10.3390/jcm10040777

69. Montemurro N, Scerrati A, Ricciardi L, Trevisi G. The exoscope in neurosurgery: an overview of the current literature of intraoperative use in brain and spine surgery. *J Clin Med*. (2022) 11(1):223. doi: 10.3390/jcm11010223

70. Mitha AP, Almekhlafi MA, Janjua MJJ, Albuquerque FC, McDougall CG. Simulation and augmented reality in endovascular neurosurgery: lessons from aviation. *Neurosurgery*. (2013) 72(Suppl 1):A107–14. doi: 10.1227/NEU.0b013e31827981fd

71. Rudarakanchana N, Van Herzeele I, Desender L, Cheshire NJ. Virtual reality simulation for the optimization of endovascular procedures: current perspectives. *Vasc Health Risk Manag*. (2015) 11:195–202. doi: 10.2147/VHRM.S46194

72. Liebig T, Holtmannspötter M, Crossley R, Lindkvist J, Henn P, Lönn L, et al. Metric-based virtual reality simulation: a paradigm shift in training for mechanical thrombectomy in acute stroke. *Stroke*. (2018) 49(7):e239–42. doi: 10.1161/STROKEAHA.118.021089

73. Montemurro N, Condino S, Cattari N, D'Amato R, Ferrari V, Cutolo F. Augmented reality-assisted craniotomy for parasagittal and convexity en plaque meningiomas and custom-made cranio-plasty: a preliminary laboratory report. *Int J Environ Res Public Health*. (2021) 18(19):9955. doi: 10.3390/ijerph18199955

74. Fick T, van Doormaal J, Hoving E, Regli L, van Doormaal T. Holographic patient tracking after bed movement for augmented reality neuronavigation using a head-mounted display. *Acta Neurochir*. (2021) 163:879–84. doi: 10.1007/s00701-021-04707-4

75. Sengupta M, Gupta A, Khanna M, Krishnan UR, Chakrabarti D. Role of virtual reality in balance training in patients with spinal cord injury: a prospective comparative pre-post study. *Asian Spine J*. (2020) 14(1):51. doi: 10.31616/asj.2019.0013

76. Drouin S, Kochanowska A, Kersten-Oertel M, Gerard IJ, Zelmann R, De Nigris D, et al. IBIS: an OR ready open-source platform for image-guided neurosurgery. *Int J Comput Assist Radiol Surg*. (2017) 12:363–78. doi: 10.1007/s11548-016-1478-0

77. Lee C, Wong GKC. Virtual reality and augmented reality in the management of intracranial tumors: a review. *J Clin Neurosci*. (2019) 62:14–20. doi: 10.1016/j.jocn.2018.12.036

78. Weigl M, Stefan P, Abhari K, Wucherer P, Fallavollita P, Lazarovici M, et al. Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-scale simulated procedure. *Surg Endosc*. (2016) 30:559–66. doi: 10.1007/s00464-015-4239-1

79. Hooten KG, Lister JR, Lombard G, Lizdas DE, Lampotang S, Rajon DA, et al. Mixed reality ventriculostomy simulation: experience in neurosurgical residency. *Neurosurgery*. (2014) 10:576–81.

80. Lafage R, Bess S, Glassman S, Ames C, Burton D, Hart R, et al. Virtual modeling of postoperative alignment after adult spinal deformity surgery helps predict associations between compensatory spinopelvic alignment changes, overcorrection, and proximal junctional kyphosis. *Spine*. (2017) 42(19):E1119–25. doi: 10.1097/BRS.0000000000002116

81. Davies RC, Johansson G, Boschian K, Lindé A, Minör U, Sonesson B. A practical example using VR in the assessment of brain injury. *Int J Virtual Real*. (1999) 4 (1):1–7. doi: 10.20870/IJVR.1999.4.1.2662

82. Lobel DA, Elder JB, Schirmer CM, Bowyer MW, Rezai AR. A novel craniotomy simulator provides a validated method to enhance education in the management of traumatic brain injury. *Neurosurgery*. (2013) 73(Suppl 1):57–65. doi: 10.1227/NEU.0000000000000116

83. Ramaswamy A, Monsuez B, Tapus A, editors. *Saferobots: a model-driven approach for designing robotic software architectures. 2014 International Conference on Collaboration Technologies and Systems (CTS)* (2014); IEEE.

84. Lee B, Liu CY, Apuzzo ML. Quantum computing: a prime modality in neurosurgery's future. *World Neurosurg*. (2012) 78(5):404–8. doi: 10.1016/j.wneu.2012.07.013

85. Higginbotham G. Virtual connections: improving global neurosurgery through immersive technologies. *Front Surg*. (2021) 8(629963).

86. Montemurro N, Perrini P. Will COVID-19 change neurosurgical clinical practice? *Br J Neurosurg*. (2022) 36(1):117–8. doi: 10.1080/02688697.2020.1773399

87. Ponce BA, Brabston EW, Shin Z, Watson SL, Baker D, Winn D, et al. Telemedicine with mobile devices and augmented reality for early postoperative care. *Annu Int Conf IEEE Eng Med Biol Soc*. (2016) 2016:4411–4.

88. Majmundar N, Ducruet AF, Wilkinson DA, Catapano JS, Patel J, Baranoski JF, et al. Telemedicine for endovascular neurosurgery consultation during the COVID-19 era: patient satisfaction survey. *World Neurosurg*. (2022) 158:e577–82. doi: 10.1016/j.wneu.2021.11.023

89. Konakondla S, Fong R, Schirmer CM. Simulation training in neurosurgery: advances in education and practice. *Adv Med Educ Pract*. (2017) 8:465–73. doi: 10.2147/AMEP.S113565

90. Montemurro N, Herbet G, Duffau H. Right cortical and axonal structures eliciting ocular deviation during electrical stimulation mapping in awake patients. *Brain Topogr*. (2016) 29(4):561–71. doi: 10.1007/s10548-016-0490-6

91. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. *JMIR Serious Games*. (2021) 9(3):e29080. doi: 10.2196/29080

92. Chan J, Pangal DJ, Cardinal T, Kugener G, Zhu Y, Roshannai A, et al. A systematic review of virtual reality for the assessment of technical skills in neurosurgery. *Neurosurg Focus*. (2021) 51(2):E15. doi: 10.3171/2021.5.FOCUS21210

93. Shuhaiber JH. Augmented reality in surgery. *Arch Surg*. (2004) 139(2):170–4. doi: 10.1001/archsurg.139.2.170

94. Giglioli IA C, Pallavicini F, Pedroli E, Serino S, Riva G. Augmented reality: a brand new challenge for the assessment and treatment of psychological disorders. *Comput Math Methods Med*. (2015) 2015:862942.

95. Pandya A, Siadat MR, Auner G. Design, implementation and accuracy of a prototype for medical augmented reality. *Comput Aided Surg*. (2005) 10(1):23–35. doi: 10.3109/10929080500221626

96. Tagaytayan R, Kelemen A, Sik-Lanyi C. Augmented reality in neurosurgery. *Arch Med Sci*. (2018) 14(3):572–8. doi: 10.5114/aoms.2016.58690

97. Deng W, Li F, Wang M, Song Z. Easy-to-use augmented reality neuronavigation using a wireless tablet PC. *Stereotact Funct Neurosurg*. (2014) 92(1):17–24. doi: 10.1159/000354816

98. Inoue D, Cho B, Mori M, Kikkawa Y, Amano T, Nakamizo A, et al. Preliminary study on the clinical application of augmented reality neuronavigation. *J Neurol Surg A Cent Eur Neurosurg*. (2013) 74(2):71–6. doi: 10.1055/s-0032-1333415

99. Abe Y, Sato S, Kato K, Hyakumachi T, Yanagibashi Y, Ito M, et al. A novel 3D guidance system using augmented reality for percutaneous vertebroplasty: technical note. *J Neurosurg Spine*. (2013) 19(4):492–501. doi: 10.3171/2013.7.SPINE12917

100. medicine jh. Vertebroplasty: johns hopkins medicine. Available at: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/vertebroplasty

101. Palmisciano P, Jamjoom AA, Taylor D, Stoyanov D, Marcus HJ. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. *World Neurosurg*. (2020) 138:e627–33. doi: 10.1016/j.wneu.2020.03.029

102. Bonaci T, Calo R, Chizeck HJ, editors. *App stores for the brain: privacy & security in brain-computer interfaces. 2014 IEEE International Symposium on Ethics in Science, Technology and Engineering* (2014); IEEE.

103. Verma V, Simone CB, Krishnan S, Lin SH, Yang J, Hahn SM. The rise of radiomics and implications for oncologic management. *J Natl Cancer Inst*. (2017) 109(7):djx055. doi: 10.1093/jnci/djx055

104. Danilov G, Shifrin M, Kotik K, Ishankulov T, Orlov YN, Kulikov A, et al. Artificial intelligence technologies in neurosurgery: a systematic literature review using topic modeling. Part II: research objectives and perspectives. *Современные технологии в медицине*. (2020) 12(6 (eng)):111–8.

105. Bravo J, Wali AR, Hirshman BR, Gopesh T, Steinberg JA, Yan B, et al. Robotics and artificial intelligence in endovascular neurosurgery. *Cureus*. (2022) 14(3).

106. Buchlak QD, Esmaili N, Leveque J-C, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev*. (2020) 43:1235–53. doi: 10.1007/s10143-019-01163-8

107. Joshi RS, Lau D, Ames CP. Artificial intelligence for adult spinal deformity: current state and future directions. *Spine J*. (2021) 21(10):1626–34. doi: 10.1016/j.spinee.2021.04.019

frontiers | Frontiers in Human Neuroscience

TYPE Review
PUBLISHED 07 September 2023
DOI 10.3389/fnhum.2023.1254417

OPEN ACCESS

EDITED BY
Mohammed Ali Alvi,
University Health Network (UHN), Canada

REVIEWED BY
Keng Siang Lee,
King's College Hospital NHS Foundation Trust, United Kingdom
Arun Karumattu Manattu,
University of Nebraska Omaha, United States

*CORRESPONDENCE
Reza Dashti
✉ Reza.Dashti@stonybrookmedicine.edu

†These authors have contributed equally to this work and share first authorship

RECEIVED 07 July 2023
ACCEPTED 23 August 2023
PUBLISHED 07 September 2023

CITATION
Gilotra K, Swarna S, Mani R, Basem J and Dashti R (2023) Role of artificial intelligence and machine learning in the diagnosis of cerebrovascular disease.
Front. Hum. Neurosci. 17:1254417.
doi: 10.3389/fnhum.2023.1254417

Role of artificial intelligence and machine learning in the diagnosis of cerebrovascular disease

Kevin Gilotra†, Sujith Swarna†, Racheed Mani, Jade Basem and Reza Dashti*

Dashti Lab, Department of Neurological Surgery, Stony Brook University Hospital, Stony Brook, NY, United States

Introduction: Cerebrovascular diseases are known to cause significant morbidity and mortality to the general population. In patients with cerebrovascular disease, prompt clinical evaluation and radiographic interpretation are both essential in optimizing clinical management and in triaging patients for critical and potentially life-saving neurosurgical interventions. With recent advancements in the domains of artificial intelligence (AI) and machine learning (ML), many AI and ML algorithms have been developed to further optimize the diagnosis and subsequent management of cerebrovascular disease. Despite such advances, further studies are needed to substantively evaluate both the diagnostic accuracy and feasibility of these techniques for their application in clinical practice. This review aims to analyze the current use of AI and MI algorithms in the diagnosis of, and clinical decision making for cerebrovascular disease, and to discuss both the feasibility and future applications of utilizing such algorithms.

Methods: We review the use of AI and ML algorithms to assist clinicians in the diagnosis and management of ischemic stroke, hemorrhagic stroke, intracranial aneurysms, and arteriovenous malformations (AVMs). After identifying the most widely used algorithms, we provide a detailed analysis of the accuracy and effectiveness of these algorithms in practice.

Results: The incorporation of AI and ML algorithms for cerebrovascular patients has demonstrated improvements in time to detection of intracranial pathologies such as intracerebral hemorrhage (ICH) and infarcts. For ischemic and hemorrhagic strokes, commercial AI software platforms such as RapidAI and *Viz.*AI have bene implemented into routine clinical practice at many stroke centers to expedite the detection of infarcts and ICH, respectively. Such algorithms and neural networks have also been analyzed for use in prognostication for such cerebrovascular pathologies. These include predicting outcomes for ischemic stroke patients, hematoma expansion, risk of aneurysm rupture, bleeding of AVMs, and in predicting outcomes following interventions such as risk of occlusion for various endovascular devices. Preliminary analyses have yielded promising sensitivities when AI and ML are used in concert with imaging modalities and a multidisciplinary team of health care providers.

Conclusion: The implementation of AI and ML algorithms to supplement clinical practice has conferred a high degree of accuracy, efficiency, and expedited detection in the clinical and radiographic evaluation and management of ischemic and hemorrhagic strokes, AVMs, and aneurysms. Such algorithms have been explored for further purposes of prognostication for these conditions, with promising preliminary results. Further studies should evaluate the longitudinal implementation of such techniques into hospital networks and residency programs

to supplement clinical practice, and the extent to which these techniques improve patient care and clinical outcomes in the long-term.

KEYWORDS

artificial intelligence, machine learning, deep learning, cerebrovascular, ischemic stroke, hemorrhagic stroke, aneurysm, arteriovenous malformation

1. Introduction

Cerebrovascular disease encompasses a wide range of pathologies that can confer a high risk of potentially life-threatening sequelae; hence, timely diagnosis and treatment is essential in preventing subsequent severe neurological deterioration (Santana Baskar et al., 2021). This requires a large team of clinicians and support staff to effectively work up and manage these patients, to enable them to receive the highest quality of care. On initial presentation outside of the hospital, emergency medical staff must quickly recognize symptoms of cerebrovascular disease, safely transfer the patient to the hospital, and obtain stroke imaging as early as possible (Santana Baskar et al., 2021). This is in tandem with timely clinical evaluation immediately on admission to determine the patient's neurological status and overall clinical picture, while also determining medical management, such as whether a patient is a candidate for medical thrombolytic therapy even before further intervention (Cumbler, 2015). From there, the radiology technologists and neuroradiologists work together to capture and interpret the appropriate imaging from which clinicians can hone in both on the critical diagnosis and in decision making for intervention/s (Cumbler, 2015; Green et al., 2021).

In recent years, with advances in technology and advanced machinery, health care has been incrementally augmented by the use of such software and technology to aid in diagnosis and decision-making for various medical conditions (Shuaib et al., 2020). Artificial intelligence (AI) is actively being implemented into many fields in medicine, and recent advancements in AI algorithms and machinery for diagnosing and treating cerebrovascular disease have the potential to revolutionize patient care (Soun et al., 2021). In order for AI to be further incorporated into the standard of care for cerebrovascular disease patients, many years of active collaboration between AI algorithm engineers and physicians are needed. In light of the current AI revolution in medicine, it is increasingly essential for health care professionals treating cerebrovascular patients to familiarize themselves with the applications of these innovations to their own field. By enhancing their knowledge, clinicians will find themselves more prepared when AI inevitably becomes an inherent part of future clinical practice (Lanzagorta-Ortega et al., 2022). In this review, we seek to provide an overview and evaluation of AI technologies applied to the field of cerebrovascular disease in the diagnosis of primary and secondary (lesional) hemorrhagic stroke, and ischemic stroke (IS). We performed a generalized review of the current literature by identifying articles that assessed the most updated AI and ML techniques. This was done by selecting the most relevant articles in the current cerebrovascular disease literature cited on the PubMed and Web of Science databases. In doing so, we aimed to summarize the clinical relevance of AI in cerebrovascular disease in simple terms so that practicing clinicians can gain a better appreciation of its current and potential future applications for cerebrovascular patients.

1.1. Terminology

Since this review requires clinicians to have a reasonable understanding of the fundamental concepts of AI, a discussion of relevant terminology would be beneficial.

AI fundamentally refers to the ability of a machine to solve tasks in a way that simulates human intelligence (Moor, 2006). Machine learning (ML), a subset of AI, utilizes large data sets to train computers to iteratively generate a model based on recognizing rules or patterns in data (Choi et al., 2020). After training, the model is then tested with real-life data (testing data) to assess its accuracy. A schematic of this process is shown in Figure 1. In ischemic stroke, ML algorithms can be developed to demonstrate the presence of infarct, the total area of infarcted brain tissue, and the occluded vessel in question. Such an algorithm may subsequently be tested on a group of non-contrast computer tomography scans (NCCTs), with results then compared to the interpretation of the NCCTs by a radiologist to ascertain the accuracy of the algorithm (Wang and Summers, 2012; Choi et al., 2020; Kim et al., 2023).

Models are often trained using three major classifications of learning: supervised, unsupervised, and reinforcement (Sarker, 2021). In supervised learning, the training data consists of pre-labeled information. With unsupervised learning, pre-existing information is absent which means the model must cluster a group of cases together based on similar characteristics and identify the relationships between said groups (Sarker, 2021). In contrast to unsupervised learning, reinforcement learning allows the model to analyze pre-existing data and determine the "correct" and "incorrect" answers for any given scenario.

Deep learning (DL), a further subclassification of ML, utilizes neural networks, a series of nodes or layers that are interconnected in such a way in order to simulate the human process of learning (Chen et al., 2022). The network begins with an input layer and ends with an output layer. In between both of these layers are a series of hidden layers, each with a given weight and bias (Jovel and Greiner, 2021). Each layer receives the input and assigns a weight to it. When the output exceeds a given bias/threshold, an activation function is applied and then fed forward to the next node. During the training phase, the model can adjust these weights and biases accordingly until it achieves the desired output (Jovel and Greiner, 2021). One such application may include software that performs a rapid analysis of a NCCT for a patient with acute ICH and then decides whether or not the neuroradiologist should immediately be notified to interpret the

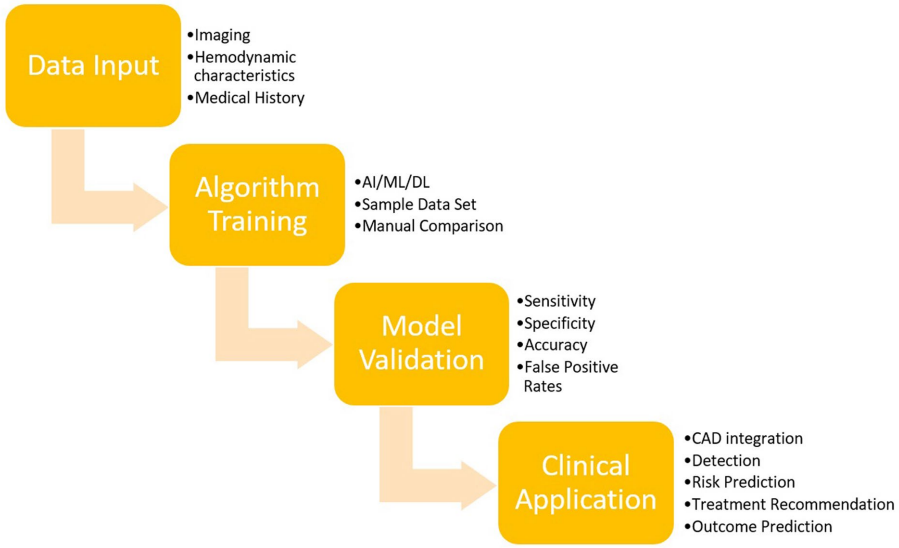


FIGURE 1
Overview of AI training and validation.

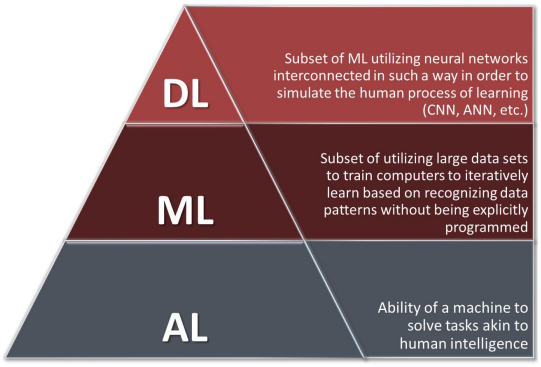


FIGURE 2
Schematic conceptual representation of AI, ML, and DL.

film or if the neurosurgery team should be notified to prepare an operating room. Ultimately, three major classifications of deep learning exist and are characterized primarily by the kind of data they use optimally. Artificial neural networks (ANNs) work best with numeric input, convolutional neural networks (CNNs) with visual input, and recurrent neural networks (RNNs) with time-series data (Banerjee et al., 2019). These concepts are outlined schematically in Figure 2.

2. Hemorrhagic stroke

2.1. Background

Hemorrhagic stroke, also referred to as intracerebral hemorrhage (ICH), has multiple subtypes based on anatomical location of the bleed, encompassing subarachnoid hemorrhage (SAH), intraventricular hemorrhage (IVH), intraparenchymal hemorrhage (IPH), epidural hematoma (EDH), and subdural hematoma (SDH) (Soun et al., 2021). ICH carries a significant morbidity and mortality that is steadily increasing in prevalence worldwide, with fatality ranging anywhere from 30 to 65% of all incidents (Rymer, 2011; Vangen-Lønne et al., 2017). Therefore, a timely diagnosis and neurosurgical intervention (if warranted), are essential for ICH patients to improve clinical outcomes (Forman et al., 2020). This is of the utmost importance when neuroimaging suggests the presence of hematoma expansion (HE), defined as an increase in ICH volume by greater than 33% from the initial collection of hemorrhage (Forman et al., 2020; Wang et al., 2021). The diagnosis of ICH can often be made through an initial NCCT of the head. Characterization of the ICH is subsequently performed through CT Angiograms (CTA), Computed Tomography Perfusion (CTP), Digital Subtraction Angiography (DSA), and Magnetic Resonance Imaging (MRI) to ultimately guide further management (Forman et al., 2020).

Neuroradiologists, neurosurgeons, and neurologists receive years of training to accurately diagnose ICH subtypes and to obtain volume measurements using the above imaging modalities. However, in the modern era, newly developed AI and ML algorithms can assist physicians to identify ICH and HE using standard imaging techniques (Wang et al., 2021). The utilization of these algorithms can both ease the interpretation of imaging and more accurately quantify hematoma volumes to improve diagnostic accuracy and guide decision-making as it pertains to acute surgical intervention or conservative and medical management (Guo et al., 2022). Many of these management decisions are dictated by whether the patient is likely to have a poor prognosis in the long-term.

HE is a pathological feature present in up to 40% of ICH patients, and is known to be one of many prognostic indicators of poor outcomes (Qureshi and Palesch, 2011). This makes it essential to accurately quantify hematoma volumes at admission and in serial imaging for ICH patients (Helal et al., 2019; Li et al., 2020). Prior history of anticoagulant/antiplatelet usage, higher baseline ICH volumes and lower admission GCS scores are the most notable risk factors for developing HE (Guo

et al., 2022). Furthermore, patients with HE subsequently are at greater risk of developing IVH and hydrocephalus (Yaghi et al., 2014). Despite the extensive literature surrounding ICH outcomes, HE is one of the few markers of outcome that can specifically be mitigated, and in some cases, prevented by neurosurgical intervention (Brouwers and Greenberg, 2013). This makes the identification of HE crucial during the care of ICH patients, as it effectively guides medical and surgical therapy while potentially preventing adverse outcomes (Aziz et al., 2015). The role of AI in these contexts pertains specifically in diagnosing subtypes of ICH, measuring ICH volumes, and in identifying HE and predictive signs of poor outcomes. Below, we discuss the current role of AI as well as how future technology can be implemented to improve neuroradiological care for ICH patients.

2.2. Standard techniques

One of the first techniques developed to evaluate ICH and measure hematoma volumes is manual segmentation (Nguyen et al., 2020). This entails the manual review of patients' head CTs by individual slices to calculate the hematoma volume (HV). Although manual segmentation is the gold standard per the current literature, given its high accuracy and error-proof methodology, it is often time consuming, particularly in ICH patients for whom timely diagnosis is essential. Moreover, with the increasing need for imaging in a growing population of ICH patients, efficient review of CT scans and other imaging modalities is critical. With this rising demand, the ABC/2 technique was developed. With this technique, "A" is defined as the length of the longest layer diameter, "B" defined as the perpendicular vertical line to A, and "C" is a by-product of layer thickness multiplied by bleeding layer number. The approximate value of HV is then estimated as the product of these variables divided by two. Many studies have demonstrated limited efficacy of the ABC/2 method, citing both overestimation and underestimation of HVs and perihematomal edema (PHE) (Wang et al., 2009; Chen et al., 2022; Hillal et al., 2022). As a result, ABC/2 was often unreliable at many level one stroke centers (Webb et al., 2015).

Since larger HVs and PHE are arguably the most important predictors of HE and poor prognosis in acute ICH, accurate detection and quantification of these variables are essential (Yaghi et al., 2014; Li et al., 2020). With an aging population and increasing incidence rates of ICH, there has been a growing demand for institutions to develop AI and ML algorithms that accurately detect PHE and calculate HVs. A select few of the most promising algorithms are reviewed below.

2.3. ML for qualitative detection of ICH

With the growth of AI as a prospective tool for ICH detection and classification, there have been many studies which have utilized and evaluated some of the algorithms and neural networks outlined above. The primary outcome reported in the literature is area under the curve (AUC), which is a marker for the predictive accuracy of the model on a scale of 0–1. One of the highest-powered studies was a retrospective study examining over 30,000 CT scans across India using deep learning algorithms in two datasets (Qure25k and CQ500) (Chilamkurthy et al., 2018). The authors utilized a natural language processing (NLP) algorithm to detect IPH, SDH, EDH, SAH, and IVH, as well as other pertinent findings such as calvarial fractures, mass effect, and midline shift. The authors noted that both datasets demonstrated strong degrees of accuracy per the consensus of independent radiologists in detecting both the actual hemorrhages as well as the other pertinent CT findings above with an AUC of 0.92 and 0.94 for the above datasets. Another study from 2021 assessed 25,000 CT scans using more advanced two sequence models with 2D CNN to classify subtypes of ICH; they reported AUCs greater than 0.98 for SAH, IVH, SDH, EDH, and IPH (Wang et al., 2021). Nishi et al. (2021) elucidated an algorithm solely designed for SAH detection with NCCTs and compared its detection results with five neurosurgeons and five general practitioners (Nishi et al., 2021). Across 135 patients with SAH, their algorithm demonstrated similar performance to the neurosurgeons and stronger performance than four of the other five physicians (Nishi et al., 2021). The findings for all of these studies above suggest that AI is capable of classifying ICH subtypes by evaluating simple neuroanatomy at a level that is on par with well-trained physicians,

Other 2D-CNNs utilized in the evaluation of ICH include GoogLeNet and AlexNet, which have been applied to the detection of basal ganglia hemorrhage (Desai et al., 2017). Desai et al. noted that both CNNs demonstrated a high degree of accuracy in detecting deep ICH (Desai et al., 2017). GoogLeNet, with a pretrained network, yielded the highest accuracy (sensitivity and specificity of 100, AUC = 1.0). The untrained AlexNet yielded a high but slightly reduced level of accuracy compared to GoogLeNet (sensitivity = 100%, specificity 80%, AUC = 0.95) (Desai et al., 2017). These results demonstrate the capacity for such algorithms to provide significant diagnostic value even for deep ICH.

Arbabshirani et al. also evaluated a fully 3D-CNN to not only detect ICH, but also to effectively triage CT scans to prioritize radiology worklists and expedite diagnosis times (Arbabshirani et al., 2018). This network was trained on over 37,000 studies and prospectively evaluated on 9,499 hitherto unseen studies. This model was able to successfully re-prioritize 94 studies from "routine" to "stat," and reduced the time to detection of ICH from 512 to 19 min. These findings demonstrate the further benefit not just in diagnosing ICH, but in expediting the process of diagnosis.

Another critical imaging modality in the detection of acute ICH is CTA. CTA allows clinicians to determine whether ICH is spontaneous, secondary to trauma, or from a pre-existing lesion such as a ruptured aneurysm or arteriovenous malformation (AVM) (Fu et al., 2023). Unlike NCCTs of the head which can be well-evaluated by most physicians, they require more subspecialist training to evaluate and often need more time to interpret in order to elicit an accurate diagnosis (Wada et al., 2007; Fu et al., 2023). One study evaluating a DL algorithm for 3,266 patients with CTAs of the head and neck showed a mean reduction time of 16.4 min when compared to trained radiologists with almost identical diagnostic accuracy (Fu et al., 2023). More research is needed to demonstrate the benefits of AI for interpretation of CTAs and other more complex imaging modalities such as MRIs/MRAs and DSAs in the context of ICH, as the current literature has primarily evaluated the use of AI in simple imaging techniques such as NCCTs (Hotta et al., 2014).

2.4. Novel machine learning techniques for quantitative hematoma evaluation

Most ML algorithms work by taking a pixel-wise approach and combining CT slice thickness to calculate HVs. Dhar et al. utilized a CNN model developed from U-Net architecture that demonstrated similar efficacy to manual segmentation techniques for calculating HVs and PHE (Dhar et al., 2020). Voxel-by-voxel overlap of hematoma segments was calculated using the Dice similarity coefficient (DSC), with a score of "0" suggesting no overlap at all and "1" suggesting maximum overlap between the manual segmentation method and the algorithm method. Across 224 CT scans in 124 patients, this study found that the UNet model had a DSC of 0.9 (IQR 0.85–0.93) for measuring HVs but only a DSC of 0.54 when measuring PHE. However, Ironside et al. (2020) used a fully automated segmentation algorithm with a goal of specifically calculating PHE in 400 patients with ICH. The automated algorithm was both faster and more accurate at detecting PHE (mean 480.5 ± 295.3 s/scan; $p < 0.0001$) and manual (mean 316.4 ± 168.8 s/scan; $p < 0.0001$) methods (Ironside et al., 2020).

Although both studies demonstrate strong efficacy of the algorithms developed, the conflicting results highlight the discrepancies that may exist between different ML algorithms when it comes to calculating PHE. Ironside et al. have followed up on their previously published results with the prospective QUANTUM study in 2022. Currently underway, this study intends to address relevant design considerations for an AI to accurately evaluate PHE in ICH patients (Ironside et al., 2022). When neurosurgeons use AI to help interpret NCCTs with acute hematomas, it is crucial that they are aware of how the AI was initially designed. This allows neurosurgeons to combine the appropriate mental resources from their clinical training with the expertise of the AI machinery to make faster and more accurate interpretations. While few subsequent studies in the literature have followed up on the evaluation of PHE, several studies have demonstrated the efficacy of using ML algorithms to calculate HVs. Yu et al. utilized "DR-UNet," a dimensional reduction analytical framework upgraded from the CNN UNet model used by Yu et al. (2022). Across 562 patients with a collective 13,825 CT scans, the DR-UNet model's hematoma volume calculations demonstrated a strongly positive correlation with the calculations made by a neuroradiologist using manual segmentation ($R^2 = 0.9979$, $p < 0.0001$). Few prior studies have demonstrated this level of statistical power with ML algorithms. The DR-UNet model also accurately evaluated 13 irregularly shaped hematomas that were shaped differently from the remainder of hematomas in the rest of the data set, thus demonstrating the potential for ML algorithms to be used even in evaluating atypical hematomas (Yu et al., 2022).

Such hematomas are often better evaluated using CTP imaging. CTP is an imaging technique that allows for the visualization and volumetric measurement of cortical infarcts well as adjacent penumbra. Wang et al. also developed an algorithm for CTP images that evaluated 49 ICH patients with concurrent interventricular hemorrhage (IVH), a feature known to be associated with poor outcomes in ICH patients (Wang et al., 2021). They found no statistically significant difference between CTP-based planimetry segmentation and their algorithm ($p = 0.614$), despite strong correlations between the two measurements ($r = 0.996$). However, the algorithm more accurately calculated volumes for the 56 ICH patients without IVH when compared to CTP-based planimetry ($r = 0.994$, $p < 0.001$). Due to the difficulty with calculating HVs with concurrent IVH, few studies have demonstrated algorithms that accurately calculate HVs in these patients (Wang et al., 2021).

Ultimately, the primary limitation of AI and ML algorithms in the current literature is their lack of efficacy in evaluating ICH in the context of IVH or infratentorial hemorrhages, both known to carry a worse prognosis than supratentorial ICH without IVH. Bleeding that extends to the ventricle can be associated with posthemorrhagic hydrocephalus, cerebral palsy, and permanent neurological deficits depending on which neural structures are compressed by the accumulation of ventricular blood (Wang et al., 2021). While the results from Wang et al. demonstrated some promising results, further studies are needed to substantiate the potential for utilizing such algorithms for ICH with concurrent IVH. ML algorithms are currently of limited use in patients with these types of ICH who are invariably more critically ill and, from a neuroanatomical standpoint, have more severe and complex disease. Neuroradiologists often use similar methodologies across each of the studies they read. Therefore, it is imperative to develop algorithms which can be consistently utilized for preoperative planning for any prospective patient with hemorrhagic stroke regardless of the further complications of IVH, hydrocephalus, or infratentorial bleeding.

In clinical practice, neuroradiologists often assess for other predictors of ICH severity on CT scans. For instance, "spot sign," defined as the presence of more than one focal enhancement within an acute hematoma, is a marker that suggests the presence of ongoing bleeding and often reliably predicts HE (Wada et al., 2007; Hillal et al., 2022). Another radiographic marker of severity is the "satellite sign," defined as the presence of a visible hemorrhage up to 20 mm away from the original hematoma. Other neuroradiological markers on CT scans assessed for by radiologists and neurosurgeons include "swirl sign," "black hole sign," "blend sign," and "island sign" (Selariu et al., 2012; Hillal et al., 2022). All of these signs are ultimately poor prognostic indicators suggesting that hematoma expansion is likely to occur, but each sign has its own intricacies, sensitivity, and specificity for such predictive values. The current literature has a select few studies which have compared the utilization of CNNs with a pre-existing deep learning model for accurately identifying the above predictive markers (Zhong et al., 2021). One such study was by Zhong et al., which demonstrated that a CNN model could identify "black hole" sign and "blend" sign when compared to a pre-existing deep learning model; however the algorithm failed to consistently identify "swirl" sign (Zhong et al., 2021). Therefore, further studies analyzing the use of such algorithms to consistently identify such radiographic signs are needed. If AI is able to successfully evaluate such radiographic signs with a high degree of accuracy, it could make the clinical training for neuroradiologists and neurosurgeons much more efficient by allowing them to allocate more mental resources to other important tasks needed for patient care.

Overall, results in the current literature have demonstrated the ability of ML algorithms to calculate HVs from CT scans (Kuo et al., 2019; Ma et al., 2022; Peng et al., 2022). A recent 2022 study by Tanioka et al. ventured even further, comparing multiple ML algorithms including XGBoost, random forests, support vector machines, and k-nearest neighbors (k-NN) (Tanioka et al., 2022). The authors suggested that the k-NN algorithm is the most superior given

its superior speed and accuracy relative to other algorithms (Tanioka et al., 2022). With improvement in technological advances, institutions are likely to continue developing more ML algorithms with better DSC scores, stronger correlations with gold standard techniques, and faster calculation times.

The utilization of such algorithms provides significant benefits for healthcare teams involved in ICH care. Neuroradiologists who first interpret the film can use the AI to better communicate with neurosurgeons the anatomical and radiographic signs relevant to the patient's clinical picture. This can ultimately allow for better preoperative planning of the respective neurosurgical intervention, which in the emergency setting, is essential for patients with severe disease. However, further studies are needed to further assess for and substantiate these potential benefits if applied to routine clinical practice.

3. Ischemic stroke

3.1. Background

Ischemic stroke (IS) is a devastating condition that annually affects nearly 795,000 people in the US alone and 11.6 million people globally (Saini et al., 2021; Tsao et al., 2022). As in ICH, earlier detection of IS can lead to more timely neurosurgical intervention and improvements in long-term morbidity and mortality. Currently, a number of imaging modalities exist that aid in both the detection as well as characterization of ischemic strokes. With the information gleaned from these studies, clinicians can estimate prognosis and make further decisions regarding both medical and surgical management. Significant advancements have been made toward utilizing artificial intelligence (AI) programs to rapidly and accurately identify abnormalities in many of these imaging studies and consequently assist in clinical decision making. The following discussion will explore some of these advancements and the potential role of AI in the diagnosis, prognostication, and management of ischemic stroke.

3.2. Detection of acute ischemic stroke

Central to the diagnosis of ischemic stroke is adequate access to accurate and rapid imaging. Although non-contrast computed tomography (NCCT) is often the initial study obtained for suspected stroke cases, its low sensitivity for detecting early ischemic stroke restricts its use primarily to the exclusion of ICH (Chalela et al., 2007). Compared to NCCT (sensitivity: 26%), MRI diffusion-weighted imaging (DWI) remains the most effective modality for the detection of early IS with a sensitivity of at least 83% in most studies (Chalela et al., 2007). Unfortunately, limitations in both accessibility and availability of MRI technology can delay crucial diagnosis. Moreover, a single MRI study takes upwards of 30 min to be performed and at most institutions, with additional delays in time associated with waiting for availability for a scanner in the setting of other emergent indications such as cauda equina syndrome, transporting the patient to and from the machine, and awaiting a confirmed read from an attending radiologist (Puhr-Westerheide et al., 2022). Although MRI is sensitive for IS, studies have shown mixed findings of long-term outcomes for patients with IS who received an MRI as their initial imaging study, which may be related, in part, to delays in imaging and subsequent interpretation (Kleindorfer et al., 2015). Therefore, AI may be of great value for reducing the delays in IS diagnosis seen across US hospitals.

The detection of subtle findings that can otherwise be missed by physicians encompasses the most promising aspects of AI (Nishio et al., 2020; Kaothanthong et al., 2022; Lu et al., 2022). To aid in the detection of "invisible" acute IS, Lu and colleagues developed a deep-learning model comprised of two deep CNNs. The first CNN was a localization model that used pre-labeled NCCT scans with both positive and negative findings to visually outline regions of interest suspicious for infarct. The resulting output was then fed-forward into the second CNN, a classification model that assigned a probability of acute IS to each study. After a training phase consisting of a subset of patients from Tongji Hospital (Institution A), the authors used the remaining patients helped internally validate the model and demonstrated reasonable performance as measured by sensitivity, specificity, accuracy, and AUC, a measure of the predictive power of a model (68.99, 98.22, 89.87, 83.61%, respectively). Even when applied to an external cohort consisting of patient data from The First Affiliated Hospital (Institution B), in which both patient demographics and image acquisition differed from that of Institution A, the model again demonstrated comparable performance in the aforementioned metrics (sensitivity: 62.99%; specificity: 89.65%; accuracy: 88.61%; AUC: 76.32%). This performance was superior to those of two experienced radiologists tested on the same collection of studies (AUC 76.32% vs. 64.01% vs. 64.39%, respectively). Furthermore, the performance of both radiologists improved with the assistance of the model (AUC 81.15 and 81.83%, respectively), suggesting the potential for a synergistic relationship between man and machine.

3.3. Characterization of ischemic stroke

After initial detection of IS, clinicians utilize a number of factors to help determine appropriate next steps in management including etiology, time of onset, presence of large vessel occlusion (LVO), core infarct volume, and size of penumbra. Early ischemic imaging findings changes, when present, can be pieced together and quantitatively assessed using the Alberta Stroke Program Early CT Score ASPECT score (Mokin et al., 2017). The evolution of the field of radiomics has allowed an unprecedented degree of speed and accuracy in the extraction of many of these features, especially compared to previously used segmentation methods. The following sections will highlight some of the ways in which machine learning algorithms have been employed to further enhance the characterization of acute ischemic stroke.

3.3.1. Etiology

In 1993, the multicenter Trial of Org in Acute Stroke Treatment (TOAST) clinical trial was completed to develop a classification system of acute IS based on primary etiology (Adams et al., 1993). This yielded five subcategories: large artery atherosclerosis, cardiogenic embolism, small vessel occlusion, IS due to other causes, and IS with unknown cause. Since then, further iterations of the TOAST classification system have been developed to allow for more accurate

subtyping; however, the most common causes of IS still belong to the first three subcategories (Wang et al., 2019). Due to the different natures in which these conditions evolve, the treatment strategies for each of these subtypes may also vary. Consequently, accurate and rapid subtyping is an important part of the diagnosis of IS.

To address this, Chen and colleagues developed a deep learning model to categorize causes of embolic IS due to either cardiogenic causes or large artery atherosclerosis (Chen et al., 2023). This model consisted of segmented CTA data that underwent feature extraction using both radiomics and CNN algorithms concurrently. The resulting output was then combined and fed-forward into a subtyping model consisting of nine separate classifiers to determine the most optimal outcome. The authors found the Adaboost algorithm to produce the highest performance (AUC: 0.9018, accuracy: 0.8929), suggesting that such models could potentially assist clinicians in making a timely diagnosis (Chen et al., 2023).

3.3.2. Time of onset

The time of onset since initial symptoms of a stroke is also an important factor to consider when determining treatment options. However, for many patients, this information may not be readily available. In fact, as many as 25% of ischemic strokes occur during sleep, often limiting available options (Lago et al., 1998; Fink et al., 2002; Serena et al., 2003). Previous research by Thomalla and colleagues has shown that diffusion-weighted imaging (DWI) in combination with T2-FLAIR imaging can actually provide insight into the approximate onset of stroke (Thomalla et al., 2011). They showed that the DWI-FLAIR mismatch, if detected, could suggest stroke onset within the prior 4–5 h. However, identifying this mismatch can be a challenging task even for experienced radiologists. Recent work has shown that deep learning could better facilitate this mismatch detection and achieve higher performance in most metrics (as measured by sensitivity, specificity, and accuracy) compared to human readers (Ho et al., 2017; Zhu et al., 2021; Polson et al., 2022).

ML has also shown promise in the field of metabolomics in determining stroke onset. Previous work has been performed to identify possible biomarkers of early stroke with varying degrees of success (Laskowitz et al., 2009; Montaner et al., 2011; Bustamante et al., 2017; Sidorov et al., 2019). Using a metabolomics-based machine learning framework called extreme gradient boost (XGBoost), Zhang and colleagues were able to identify multiple serum biomarkers in rats that were both predictive of the presence of stroke as well as the time of onset (Zhang et al., 2022). Together, these findings show the potential for ML in stroke management beyond the scope of imaging alone. Future areas of research should focus on how AI can be implemented in neurosurgical intervention specifically.

3.4. Core infarct volume segmentation

Accurate estimation of this core infarct volume through CTP is essential in guiding further management. However, among available CTP software programs, there is tremendous variability in the parameters used to define core infarct and the resulting perfusion maps (Fahmi et al., 2012; Koopman et al., 2019; Hoving et al., 2022). A number of ML algorithms have been trained using either CTP maps alone or some combination of these perfusion maps with other clinical data (Clerigues et al., 2019; Chen et al., 2020; Wang et al., 2020). The incorporation of these maps introduces inherent variability in the models themselves. Furthermore, each model must first be trained according to the individual CTP software program used prior to clinical application. To bypass these issues, de Vries and colleagues developed a U-Net like model, PerfU-Net, that utilized CTP source data without the use of any intermediate perfusion map (de Vries et al., 2023). The model also utilized the concept of symmetry awareness and skip connections, whereby a potentially infarcted hemisphere was compared to its healthy counterpart and the output from one layer became the input to every other subsequent layer in order to select for particularly salient elements of the image. While PerfU-Net did not perform at the level of some of the top models that did use CTP maps, it demonstrated comparable performance [Dice: 0.46, precision: 0.54, recall: 0.49, average volume difference (AVD) compared to ground truth using MRI-DWI: 12.74]. These findings suggest that AI can help standardize and relatively accurately estimate volumetric measurements of infarct.

3.5. Large vessel occlusion

Large vessel occlusion is a prominent cause of acute IS, which carries a disproportionately high level of morbidity and mortality (Smith et al., 2006, 2009; Malhotra et al., 2017). Consequently, rapid detection of LVO is of paramount importance in the initial workup of any stroke. AI has shown tremendous potential in this sphere, leading to the development of commercially available programs, such as and Rapid CTA and *Viz*.ai-LVO, that have reasonably high sensitivity and specificity (Murray et al., 2020; Karamchandani et al., 2023). Some recent work has even demonstrated a role for AI in detecting LVO using CTAs obtained in mobile stroke units (MSUs) (Czap et al., 2022). MSU CTAs exhibit limitations in their capacity to show ischemia in part due to the variability in image quality when compared to traditional scanners as well as an earlier than usual acquisition time after stroke onset. The model developed by Czap and colleagues was initially trained and tested using in-hospital CTAs, achieving an AUC of 0.84. The authors then tested the same model using CTAs obtained in MSUs and achieved a comparable performance with an AUC of 0.80 (Czap et al., 2022). Together, these results show that MSUs equipped with adequately trained machine learning algorithms can assist clinicians in achieving faster diagnoses in the acute setting. The value of this additional speed that AI offers cannot be understated in the acute setting where "time is brain."

3.6. Prognostication of ischemic stroke

In addition to facing the clock, clinicians are often challenged by patients' families who are concerned regarding the acute and long-term prognosis following their stroke. The answer to this question may vary depending on a number of factors including clinical features, radiological findings, and medical history as well as the treatment options available on presentation. The Alberta Stroke Program Computed Tomography Score (ASPECTS) was one of the first efforts to predict outcome following an MCA stroke as it allowed for the quantification of ischemic changes in the anterior circulation using NCCT (Barber et al., 2000). It was originally designed to

TABLE 1 Commercially utilized software applications utilizing AI in ischemic stroke and ICH.

Software	Application	Clinical use	Imaging modality
RapidAI	Rapid ICH	Detection and classification of ICH	CT
	Rapid CTA	Detection of LVO	CTA
	Rapid CTP	Detection of perfusion mismatch	CTP
	Rapid MR	Detection of perfusion mismatch	MRI/MRA
	RapidASPECTS	Facilitates ASPECTS grading for stroke	CT
Brainomix	e-ASPECTS	Facilitates ASPECTS grading for stroke	CT
	e-Blood	Detection and quantification of ICH	CT
	e-CTA	Detection of LVO, collateralization	CTA
	e-Mismatch	Detection of perfusion mismatch	CTA/CTP/MRA
Viz.ai	*Viz* LVO	Detection and triage of LVO	CTA
	Viz CTP	Detection of perfusion mismatch	CTP
	Viz ICH	Detection and triage of ICH	CT
Aidoc	Aidoc LVO	Detection and triage of LVO	CTA
	Aidoc CTP	Detection of perfusion mismatch	CTP
	Aidoc ICH	Detection and triage of ICH	CT
Avicenna.AI	CINA LVO	Detection and triage of LVO	CTA
	CINA ASPECTS	Facilitates ASPECTS grading for stroke	CT
	CINA ICH	Detection and triage of ICH	CT

identify those patients that would benefit from thrombolytic therapy. Since then, it has been used to decide which patients may benefit from mechanical thrombectomy (Yoo et al., 2014; Goyal et al., 2015; Powers et al., 2015).

Initially, the manual interpretation of ASPECTS was tedious and often showed variability due to a number of factors including human rater experience (Wardlaw et al., 2007; Menon et al., 2011). Today, the advent of commercially available AI software such as RapidAI's Rapid ASPECTS and Brainomix's e-ASPECTS has automated this process, resulting in improved inter-rater agreement (Nagel et al., 2017; Goebel et al., 2018; Shibata et al., 2022; Chan et al., 2023). Additionally, several studies have produced models that have been able to generate ASPECTS scores from NCCT in high concordance with ASPECTS scores derived from DWI (Albers et al., 2019; Kuang et al., 2019). These performances were superior to that achieved by experienced clinicians, thus suggesting that AI can play a future role in ischemic stroke management and not just diagnosis.

In addition to these programs, there are other ML models that have been developed to predict outcomes following IS. For example, AI has been used to predict the occurrence of specific events immediately following IS including malignant cerebral edema as well as hemorrhagic transformation and reinfarction following thrombectomy (Yu et al., 2018; Foroushani et al., 2020; Choi et al., 2021; Hoffman et al., 2023). Such measures may allow future clinicians to determine disease severity earlier in the management of IS patients. Multiple studies have demonstrated efforts to develop ML algorithms that can predict functional outcomes at various stages post-discharge (Forkert et al., 2015; Dragoş et al., 2023; Ozkara et al., 2023). Although the success of these studies is somewhat limited, it shows great promise for the future of IS care.

3.7. AI software implemented in practice in cerebrovascular accident patients

Through the iterative enhancements of AI demonstrated in the literature outlined above for CVA patients, there has been the recent implementation of commercial software platforms at various stroke centers utilizing AI for both ischemic and hemorrhagic stroke, some of which are highlighted in Table 1.

These software applications facilitate a systematic approach to suspected CVA patients, and can expedite the process of detecting CVA, and triaging who is eligible and may benefit from neurosurgical intervention. The practical implementation of AI software from a community hospital to a major stroke center is outlined in Figure 3.

4. AI in intracranial aneurysms

Intracranial aneurysms (IAs) have a prevalence of approximately 3.2% in the general adult population. While most aneurysms are found incidentally, there is always a risk of rupture that is associated with a 25% mortality rate during the first 24 h. Ruptured aneurysms are responsible for approximately 500,000 annual deaths worldwide, with the risk of rupture increased with larger aneurysms (Rinkel et al., 1998; Vlak et al., 2011; Morita et al., 2012; Jersey and Foster, 2023). Unfortunately, most patients with cerebral aneurysms are unaware of their condition since they remain asymptomatic until a rupture occurs and leads to long-term neurological deficits. Therefore, the accurate and timely diagnosis of aneurysms is essential for astute clinical decision-making regarding intervention and subsequent clinical outcomes. With respect to unruptured aneurysms, there is still debate on the specific guidelines regarding if and under what circumstances

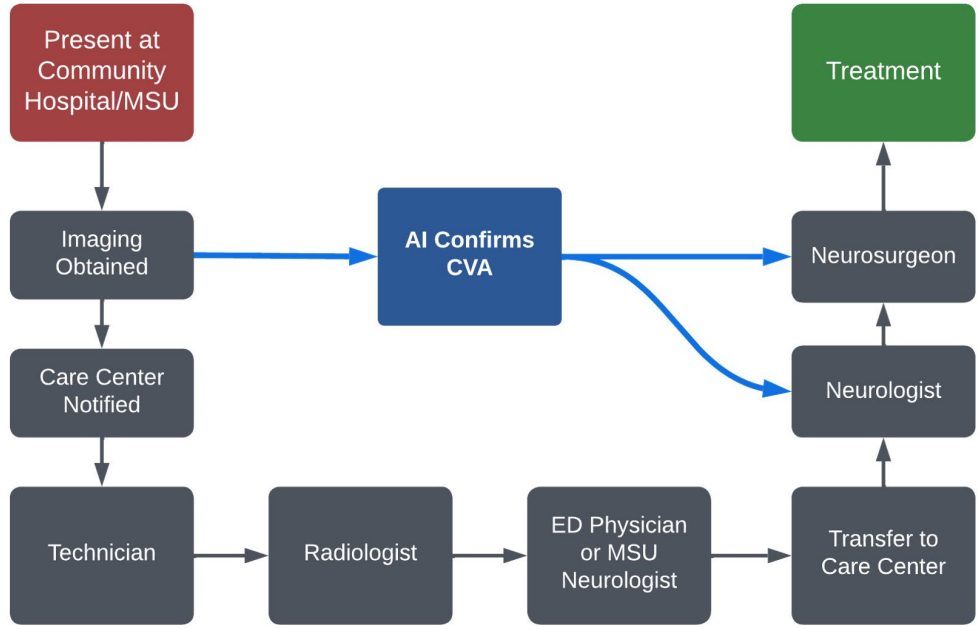


FIGURE 3
The use of AI software from an outside hospital to a major stroke center for a cerebrovascular patient requiring complex stroke intervention.

intervention is warranted, the prognostication of aneurysms, as well as the preferred form of neurosurgical intervention (Thompson et al., 2015). The introduction of AI, specifically DL algorithms, shows promise for improving present systems for the detection and prognostication of aneurysms.

4.1. The role of AI in aneurysm detection

Imaging modalities typically used for detection of aneurysms are similar to those used in the acute ICH setting, and they include DSA, CTA, and MRA. Each of these modalities can be operator-dependent and has varying degrees of accuracy in detecting subtle arterial lesions. Some of the potential diagnostic inaccuracies have been mitigated through the implementation of computer-aided diagnosis (CAD) systems. These are automated systems which, through the use of algorithms, can assist in the detection of abnormal imaging findings through the analysis of certain imaging features such as arterial wall defects (Arimura et al., 2004; Lauric et al., 2010; Yang et al., 2011; Hanaoka et al., 2019). While much of the earlier literature cited a high false positive rate of CAD, the recent incorporation of DL models, have allowed for greater adaptations and potential improvements to such systems. In theory, efficient DL models can facilitate greater sensitivity of this system in detecting IAs across various forms of imaging. AI can additionally process the dynamic nature of aneurysms to better understand its true size and morphology than simple diameter measurement can provide (Sahlein et al., 2022).

DSA is currently recommended as the first-line imaging modality due to its high sensitivity in detecting aneurysms of all sizes (Thompson et al., 2015). When utilized with AI, CNN classifiers have yielded promising results in retrospective studies, with the literature citing a sensitivity of 79–100% (Jerman et al., 2017; Duan et al., 2019; Hainc et al., 2020; Jin et al., 2020; Zeng et al., 2020; Ou et al., 2022). These studies have analyzed a wide array of algorithms, such as RAGS, RetinaNet, YOLOv3, ViDi, and UNet. Jerman et al. demonstrated a 100% sensitivity using a 7-layer 2D-CNN model (Jerman et al., 2017). Hainc et al. utilized a commercially available ML system on 2D images, and even while demonstrating a comparatively lower sensitivity (79%) of this algorithm, they demonstrated the feasibility of using AI, with the capacity for further improvement with the utilization of 3D images and manual ML training (Hainc et al., 2020). This can lead to lower error rates and earlier detection times in asymptomatic patients who are unbeknownst to their own risk of aneurysm rupture.

CTA is a less invasive imaging modality to diagnose aneurysms with a similarly high sensitivity and specificity rate. Its main limitations are the susceptibility to artifact and decreased sensitivity for smaller aneurysms compared to DSA (Thompson et al., 2015). Often, aneurysm detection is enhanced by vasculature rendering and bone subtraction. As a result, the utilization of ML in combination with CTA has a sensitivity that differs based on aneurysm characteristics, with many studies utilizing the HeadXNet model, UNet, and DeepMedic algorithms (Park et al., 2019; Shahzad et al., 2020; Shi et al., 2020; Alwalid et al., 2021; Bo et al., 2021; Pennig et al., 2021; Yang et al., 2021; Liu et al., 2023). Compared to other radiographic combinations, CTA had the most widespread sensitivities with high dependence on size and location. This was demonstrated by Shi et al., who analyzed the DaResUNet model, and detected a sensitivity ranging from 51.7 to 100% and from 60.6 to 100% for location (Shi et al., 2020). In spite of these limitations, other studies have demonstrated the overall efficacy of augmentation AI algorithms with clinicians and reliable segmentation (Park et al., 2019). The FDA-approved *Viz.*AI Aneurysm CNN had a sensitivity of 93.8% and accuracy of 94.0% for aneurysms

in their approved range of 4 mm or larger with a mean processing time of 114.7 s (Colasurdo et al., 2023). RAPID Aneurysm was one of the most promising accurate programs, demonstrating a sensitivity of 95% and accuracy of 99% across 51 patients (Heit et al., 2022). However, there remains a dearth of available data pertaining to CTA with ML in aneurysm diagnosis and management.

With respect to MRA, when using ML techniques with CAD, the sensitivities for aneurysm detection ranged from 70 to 100%, depending on technique and size (Nakao et al., 2018; Sichtermann et al., 2019; Ueda et al., 2019; Chen et al., 2020, 2023; Faron et al., 2020; Joo et al., 2020; Shimada et al., 2020; Sohn et al., 2021; Terasaki et al., 2021). Most of these models utilized validated ML algorithms UNet, DeepMedic, and ResNet with retrospective data. These include the 2D models of MRA and TOF-MRA, with higher sensitivities found for larger aneurysms and when the model was used in concert with human readers (Nakao et al., 2018; Sichtermann et al., 2019; Stember et al., 2019; Faron et al., 2020). Nakao et al. employed a "2.5D" approach by having the network input 2D-image representations of 3D objects through the combination of a CNN and MIP algorithm created to be incorporated with pre-existing CAD systems. This demonstrated similar outcomes to 3D network models (Chen et al., 2020; Faron et al., 2020; Joo et al., 2020; Sohn et al., 2021). Chen et al., using UNet, noted significant decreases in false positives (0.86 per case) compared to most other studies when running the algorithm after use of an automated vessel segmentation algorithm (Chen et al., 2020). The same authors conducted a 2023 multi-center follow-up study, in which they improved their Squeeze-and-Excitation (SE) 3D UNet network optimized with a five-fold cross-validation, which yielded an increase in 15.79% for patient-level sensitivity and further decrease in false positives (Chen et al., 2023). ML with CAD may be a suitable option for aneurysm detection in clinical scenarios where patient care is satisfied by noninvasive imaging alone.

Terasaki et al. have conducted one of the few multi-center studies evaluating ML models of 2D-CNN, 3D-CNN, and multidimensional CNN using TOF-MRA images with both 3.0 and 1.5T (Terasaki et al., 2021). The authors found the highest sensitivities rates of 82.1, 86.5, and 89.1% in the external tests with false positive rates of 5.9, 7.4, and 4.2%, respectively, validating this approach of AI and ML with MRA imaging (Terasaki et al., 2021) Of note, this study also displays the overall trend of high false positive rates of MRA compared to CTA and DSA. Teresaki et al.'s high false positive rate may lead to unnecessarily aggressive management being taken by clinicians if they solely rely on the AI to make their diagnosis. This can lead to a waste of scarce and expensive resources in the acute setting and may limit the trust that clinicians have in AI when managing patients with unruptured IAs. This is in stark contrast to Chen et al.'s findings whereby demonstrating fewer false positive rates may allow physicians to better trust these algorithms for guiding management. Future research on AI should attempt to mitigate the false positive rate as low as possible in order to allow for a more feasible AI implementation into the acute setting.

4.2. The role of AI in determining aneurysm rupture risk and prognostication

Upon detection of an unruptured IA, the next step is to determine whether surgical intervention is indicated. While no clear consensus hitherto exists, there are scoring systems such as the PHASES score, Ruptured Resemblance Score (RRS), and ELAPSS score which can guide clinical judgment (Greving et al., 2014; Thompson et al., 2015; Backes et al., 2017; Rajabzadeh-Oghaz et al., 2020). The PHASES score provides risks of IA rupture based on the following: population ethnicity, hypertension, age, size of aneurysm, earlier subarachnoid hemorrhage (SAH) from another IA, and site of IA. ELAPSS analyzes the location of the IA, patient age, size, and shape of the IA, and earlier SAH. The RRS assesses the similarity of a current IA to a cohort of previously ruptured IAs based on hemodynamic-morphological parameters such as wall shear stress. Current guidelines suggest that the greatest factors to consider for potential intervention are age, aneurysm location, aneurysm size, type of treatment center, and the available specialized intraoperative tools and techniques, such as intraoperative angiography. Non-sphericity index scores are evaluated through intraoperative angiography to objectively assess aneurysm characteristics, with higher scores correlating with more irregular shapes. The current literature shows that future growing aneurysms generally exhibited higher NSI scores (Hotta et al., 2014; Dhar et al., 2020).

One goal of ML in this setting is to predict the risk of rupture with incidental aneurysms. Various studies have demonstrated higher degrees of accuracy in predicting IA rupture with ML, particularly when compared to models using limited morphological features or the aforementioned scoring systems (Liu et al., 2018, 2019; Zhu et al., 2020; Bizjak et al., 2021; Sahlein et al., 2022; Li et al., 2023). Rapid Aneurysm, a newer tool from the RapidAI developers that have already implemented several tools into clinical practice, was recently used in a retrospective review of ruptured aneurysms that were treated conservatively (Sahlein et al., 2022). The authors found relative volumetric minimum enlargements of 6%, that were initially deemed undetectable by manual linear measurement (Sahlein et al., 2022).

Bizjak et al. demonstrated the use of models in predicting IA growth based on morphological features. The authors obtained morphological data for 44 IAs from CTA and MRA that subsequently utilized "deep shape" learning via the PointNet++ model to extract vascular surface meshes from the images to predict future aneurysm growth and rupture. This study demonstrated high sensitivity (0.96) and satisfactory accuracy (0.82) (Bizjak et al., 2021).

Furthermore, ML may also be particularly useful for smaller-sized aneurysms ($\leq$7 mm in diameter), for which the risk of rupture or growth is often harder for clinicians to predict (Ahn et al., 2021; Lee et al., 2021; Xiong et al., 2022). This has been demonstrated in a recent systematic review using meta-regression that found no significant predictors for small aneurysm growth or size (Lee et al., 2021). Therefore, the AI/ML granularity for aneurysm growth in this population is of particular note. Xiong et al. created an ML model in 1,400 patients in conjunction with a support vector machine (SVM) algorithm which utilizes supervised learning models with associated learning algorithms to analyze data for classification and regression analysis. The authors noted that SVM outperformed the PHASES score in predicting aneurysm rupture with an AUC of 0.817 and 0.893 in the internal and external validity cohorts, respectively. Through the use of ML, the authors concluded that maximum size, location, and irregular shape of the IAs were the major predictors of aneurysmal rupture (Xiong et al., 2022).

The highest predictive value for rupture risk in aneurysm patients was observed when algorithms include data beyond imaging characteristics, particularly when adding hemodynamic characteristics

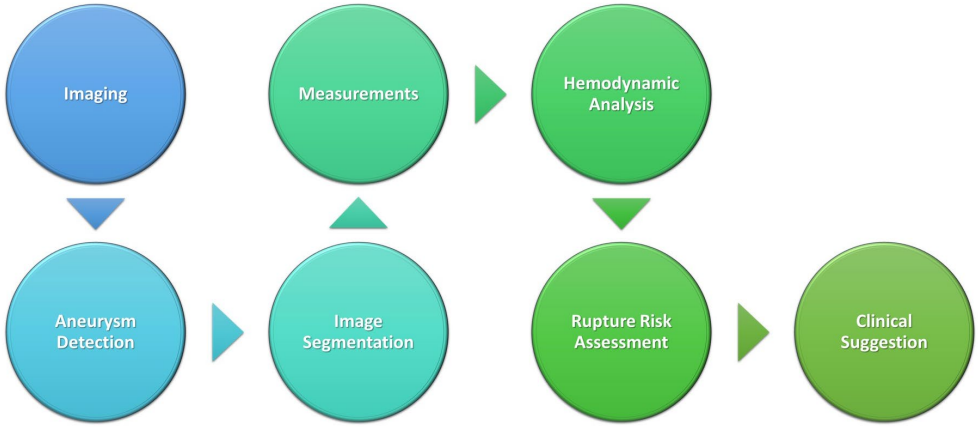


FIGURE 4
Schematic of AI process in aneurysm detection and clinical decision-making.

and clinical information (Chen et al., 2020, 2022; Detmer et al., 2020; Ou et al., 2022). Chen et al. was one of the first researchers to utilize hemodynamic characteristics in their modeling, finding them to be a more significant predictor than imaging in their retrospective study with 1,007 IA patients (Chen et al., 2020). Their tested algorithms (including random forest modeling, multilayer perceptron, and SVM) trained on two different hospital sets performed similarly to traditional logistic regression modeling. In 2022, researchers at Capital Medical University similarly compared different ML and deep learning algorithms both with and without hemodynamic features, noting greater accuracy for almost every measure when hemodynamic features were included (Chen et al., 2022). An example of this workflow is shown in Figure 4.

4.3. AI in predicting outcomes after IA intervention

ML has also been utilized not just in predicting clinical outcomes on presentation, but also in predicting the risk of occlusion after interventions with various endovascular devices. Various ML systems such as ElasticNet and UNet have been utilized in this way yielding sensitivity results ranging from 75 to 98% (Paliwal et al., 2018; Shiraz Bhurwani et al., 2020; Guédon et al., 2021; Jadhav et al., 2022). These studies generally utilized DSA imaging, but spanned multiple interventional methods including flow diverters, intrasaccular embolization devices, and pipeline embolization devices.

Paliwal et al. retrospectively analyzed 84 ICA sidewall aneurysms treated with flow diverters with ML algorithm parameters, including factors such as hemodynamics, morphology metrics, and morphometrics neck ratio (calculated on the 2D DSA images) (Paliwal et al., 2018). The most sensitive ML algorithm tested was the Guassian-SVM and Neural Network (with 90% accuracy). However, all 5 algorithms used in this study demonstrated significant improvement when using the "all-parameters" model that included 16 inputs compared to the" significant parameters" model which only included five inputs. Guedon et al. also studied ML algorithms in the context of flow diverters with the aim of utilizing AI to address the lack of scoring criteria for occlusion prediction after this treatment. The authors included 146 subjects with ElasticNet for feature selection and outcome prediction (occlusion or no occlusion). They demonstrated their DIANES score to have an 89% sensitivity and 81% accuracy (Guédon et al., 2021). Additionally, Guedon et al. and Paliwal et al. found that neck ratio may be a significant factor in occlusion 6 months after treatment (Paliwal et al., 2018; Guédon et al., 2021).

Jadhav et al. specifically focused on wide-neck bifurcation aneurysms treated with an intrasaccular device defining occlusion outcomes according to the Raymod-Roy Occlusion Classifications, which classifies aneurysm occlusions into three categories (complete occlusion, residual neck, and residual aneurysm), rather than the binary outcomes hither to reported in prior literature (Jadhav et al., 2022). The authors then created different "feature sets" that each contained different input combinations and analyzed each set with different algorithms. Random forest modeling used with a feature set combining clinical and imaging features displayed the highest accuracy of 75.3% and sensitivity of 91.8%. The authors additionally developed a neural network segmentation algorithm similar to UNet to automate 2D and 3D image characteristic calculations which performed similarly to the manual computations.

The long-term outcomes of patients with IAs is an important topic that can be addressed by advancing technology as well. Bhurwani et al.'s research defined outcomes as occlusion or no occlusion at 6 months offering predictive postoperative values for patients treated with pipeline embolization devices (Shiraz Bhurwani et al., 2020). They utilized Keras with DSA imaging and angiographic parametric imaging to train a deep neural network, finding an average sensitivity of 0.92 but a specificity of only 0.57. Other studies found higher predictive values when separating full occlusions from partial occlusions, which may account for the wide range in sensitivities within the literature (Jadhav et al., 2022). The combination of imaging and clinical factors improves sensitivity as well, with Guedon et al. utilizing ElasticNet to create the DIANES grading scale with an 89% sensitivity and 81% accuracy (Guédon et al., 2021). Further investigations are needed to reach a clearer consensus on the effectiveness of ML in predicting occlusion outcomes.

Of note, a novel ML algorithm was presented in 2021 by Williams et al., with their proposal of the Aneurysm Occlusion Assistant to provide real time surgical guidance (Williams et al., 2021). Using the

open source softwares Keras, Tensorflow, and skLearn combined with angiographic parametric imaging and segmented DSA imaging, the authors demonstrated predictions on occlusion risk after device placement within 7 s, with reasonable prediction accuracy for occlusion at 6-months (0.84) This avenue of prognostication yields significant potential for cerebrovascular neurosurgeons and requires further investigation.

5. Arteriovenous malformations

AVMs are aberrant, dysfunctional connections between arteries and veins without an intervening capillary bed (Cockroft et al., 2012). These aberrant lesions consist of an intervening nidus with an intertwining of blood vessels that have a high propensity to bleed (Cockroft et al., 2012). Brain AVMs frequently present with rupture which can lead to loss of consciousness, permanent neurological deficits, and death. Non-ruptured AVMs can be detected incidentally, or present with headaches, progressive neurological deficit, or seizure.

5.1. Current literature on the role of AI in AVMs

AVMs appear to have a greater propensity to bleed in pediatric patients compared to adults (El-Ghanem et al., 2016). Saggi et al. utilized three ML algorithms, random forest models, gradient boosted decision trees, to predict the risk of hemorrhage for AVMs in 189 pediatric patients who presented with or without hemorrhage (Saggi et al., 2022). The ML algorithm discerned that smaller AVM sizes, left-sided AVMs, and the presence of a concurrent arterial aneurysm were all predictors of hemorrhage on presentation. When compared with a conventional regression approach, only the ML algorithm was able to pick up on these subtleties.

Few studies have demonstrated that AI can successfully predict hemorrhage secondary to AVM in adult patients. Across 1810 patients, Oermann et al. developed a 3D-surface ML algorithm model that accurately predicted the risk of adverse events in patients who received radiosurgery for AVM solely based off imaging findings (Oermann et al., 2016). Oermann et al. compared their algorithm to widely used scoring systems such as Spetzler-Martin grading scale (Spetzler and Martin, 1986), radiosurgery-based AVM score (RBAS) (Pollock and Flickinger, 2002), and the Virginia Radiosurgery AVM Scale (VRAS) (Starke et al., 2013). For all three comparisons of the algorithm, the AUC for the predictive capability ranged from 0.6 to 0.7 depending on the time point at which the comparison was made. Jiao et al. performed a similar study where AI-based indicators were used to predict the likelihood of postoperative motor deficits in patients who received AVM resection surgery (Jiao et al., 2023). When compared to the Spetzler-Martin grading scale, the highest AUC was observed in the logistic regression model of 0.88 (Jiao et al., 2023).

5.2. Limitations and future directions for ML algorithms in cerebrovascular neurosurgery

Although novel machine learning (ML) techniques show significant promise for cerebrovascular surgery, there are still many limitations that must be accounted for prior to further application in clinical practice. Of note, the majority of the studies analyzing ischemic stroke, ICH, AVM, and aneurysms were retrospective in nature with small sample sizes. Furthermore, much of the literature utilizes retrospective analyses, meaning that the actual, prospective implementation of these models is still pending. These mostly single center studies lead to small datasets on which the ML algorithms can train. The reported sensitivities and accuracies for these models are thus difficult to compare and generalize. Additionally, significant manual input and calculation was required to analyze the desired parameters. This minimizes the automaticity and time saving potential of AI as a clinical tool. There is a greater need for multi-center prospective studies to substantiate such AI algorithms prior to their full implementation into clinical practice. The lack of substantive studies evaluating AI is specifically notable for vascular lesions such as aneurysms and AVMs.

Moreover, when deep learning models are developed, they take into consideration how the institution's electronic medical record (EMR) system and picture archiving/communication systems (PACS) will display radiological films. A major drawback that applies to most of the studies noted in this review is that they are single-center studies. Therefore, it would be difficult to implement any of these ML algorithms at other institutions where the PACS software and CT imaging protocols for ICH patients may have subtle differences. Many critically ill ICH patients at larger academic centers are often transferred from smaller outside hospitals where the initial head CT was performed (Vahidy et al., 2016). A successful algorithm is one that can be implemented at both tertiary centers and smaller hospitals with differing software. Future studies should consider the flexibility of ML algorithms so that they can evaluate the algorithms' efficacy in samples across multiple institutions with different EMRs and PACS software.

There is also the ethical dilemma that comes with the use of an automated system in clinical decision making for cerebrovascular neurosurgeons. AI can be beneficial in creating more standardization between physician detection and decisions while also analyzing more variables or in identifying more minute vascular abnormalities the human eye can. However, the large rates of false positives shown in the literature are proof of principle which is two-fold. Firstly, of the need for further investigation and iterative improvement in these algorithms, and secondly, that these algorithms are a useful adjunct but are no substitute for the judgment of a clinician.

Nevertheless, AI provides a promising avenue to revolutionize the practice of modern neurosurgery, and further lines of inquiry are needed both to improve on what has been done, and to open up further avenues of use in cerebrovascular surgery.

6. Conclusion

Cerebrovascular disorders often carry a significant management morbidity and mortality, making timely diagnosis and intervention essential. The application of AI, particularly ML and DL, in the realm of cerebrovascular neurosurgery has facilitated a more expedited detection, triaging, and prognostication of cerebrovascular pathologies such as ischemic strokes, ICH, AVMs, and aneurysms. Moreover, the literature has demonstrated that having this technology has conferred improvements in timely detection of these pathologies

for physicians in training. Some of these AI software applications have already been implemented into clinical practice, particularly in the realms of ICH and ischemic stroke. As this technology continues to develop and be applied in increasingly innovative ways, it is important to remember that many of these models depend on a pre-determined gold standard to determine ground truth. As such, their maximum potential may be limited by the information immediately available to the algorithm. Clinicians must ultimately go beyond simply evaluating how certain machinery may be beneficial in the pre, intra, or post-operative setting. In order to ensure patient safety, physicians and their staff should have a thorough understanding of the methodology used to develop AI, its advantages and limitations in the clinical setting as well as the barriers to its implementation for specific patient populations. Future multicenter prospective studies are needed to further substantiate these algorithms for further application in clinical practice, particularly with vascular lesions such as aneurysms and AVMs. While in its relative nascency in clinical practice, AI technology provides significant promise as an adjunct for neurosurgeons in revolutionizing clinical decision-making and subsequent clinical outcomes in cerebrovascular surgery.

Author contributions

RD: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. KG: Conceptualization, Methodology, Writing – original draft. SS: Conceptualization, Methodology, Writing – original draft. RM: Conceptualization, Writing – original draft, Writing – review & editing. JB: Conceptualization, Methodology, Writing – original draft.

Acknowledgments

We would like to acknowledge all members of the Dashti Lab; faculty, residents, and medical students, past and present, for their efforts in this line of research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adams, H. P. J., Bendixen, B. H., Kappelle, L. J., Biller, J., Love, B. B., Gordon, D. L., et al. (1993). Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of org 10172 in acute stroke treatment. *Stroke* 24, 35–41. doi: 10.1161/01.STR.24.1.35

Ahn, J. H., Kim, H. C., Rhim, J. K., Park, J. J., Sigmund, D., Park, M. C., et al. (2021). Multi-view convolutional neural networks in rupture risk assessment of small, Unruptured intracranial aneurysms. *J. Pers. Med.* 11:239. doi: 10.3390/jpm11040239

Albers, G. W., Wald, M. J., Mlynash, M., Endres, J., Bammer, R., Straka, M., et al. (2019). Automated calculation of Alberta Stroke program early CT score: validation in patients with large hemispheric infarct. *Stroke* 50, 3277–3279. doi: 10.1161/STROKEAHA.119.026430

Alwalid, O., Long, X., Xie, M., Yang, J., Cen, C., Liu, H., et al. (2021). CT angiography-based Radiomics for classification of intracranial aneurysm rupture. *Front. Neurol.* 12:619864. doi: 10.3389/fneur.2021.619864

Arbabshirani, M. R., Fornwalt, B. K., Mongelluzzo, G. J., Suever, J. D., Geise, B. D., Patel, A. A., et al. (2018). Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *npj Digit. Med.* 1:9. doi: 10.1038/s41746-017-0015-z

Arimura, H., Li, Q., Korogi, Y., Hirai, T., Abe, H., Yamashita, Y., et al. (2004). Automated computerized scheme for detection of unruptured intracranial aneurysms in three-dimensional magnetic resonance angiography. *Acad. Radiol.* 11, 1093–1104. doi: 10.1016/j.acra.2004.07.011

Aziz, H., Branco, B. C., Braun, J., Hughes, J. D., Goshima, K. R., Trinidad-Hernandez, M., et al. (2015). The influence of do-not-resuscitate status on the outcomes of patients undergoing emergency vascular operations. *J. Vasc. Surg.* 61, 1538–1542. doi: 10.1016/j.jvs.2014.11.087

Backes, D., Rinkel, G. J. E., Greving, J. P., Velthuis, B. K., Murayama, Y., Takao, H., et al. (2017). ELAPSS score for prediction of risk of growth of unruptured intracranial aneurysms. *Neurology* 88, 1600–1606. doi: 10.1212/WNL.0000000000003865

Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., et al. (2019). Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. *Artif. Intell. Med.* 97, 79–88. doi: 10.1016/j.artmed.2018.11.004

Barber, P. A., Demchuk, A. M., Zhang, J., and Buchan, A. M. (2000). Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet* 355, 1670–1674. doi: 10.1016/S0140-6736(00)02237-6

Bizjak, Ž., Pernuš, F., and Špiclin, Ž. (2021). Deep shape features for predicting future intracranial aneurysm growth. *Front. Physiol.* 12:644349. doi: 10.3389/fphys.2021.803863

Bo, Z. H., Qiao, H., Tian, C., Guo, Y., Li, W., Liang, T., et al. (2021). Toward human intervention-free clinical diagnosis of intracranial aneurysm via deep neural network. *Patterns* 2:100197. doi: 10.1016/j.patter.2020.100197

Brouwers, H. B., and Greenberg, S. M. (2013). Hematoma expansion following acute intracerebral hemorrhage. *Cerebrovasc. Dis.* 35, 195–201. doi: 10.1159/000346599

Bustamante, A., López-Cancio, E., Pich, S., Penalba, A., Giralt, D., García-Berrocoso, T., et al. (2017). Blood biomarkers for the early diagnosis of stroke: the stroke-chip study. *Stroke* 48, 2419–2425. doi: 10.1161/STROKEAHA.117.017076

Chalela, J. A., Kidwell, C. S., Nentwich, L. M., Luby, M., Butman, J. A., Demchuk, A. M., et al. (2007). Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet* 369, 293–298. doi: 10.1016/S0140-6736(07)60151-2

Chan, N., Sibtarin, N., Booth, T., de Souza, P., Bibby, S., Mah, Y. H., et al. (2023). Machine-learning algorithm in acute stroke: real-world experience. *Clin. Radiol.* 78, e45–e51. doi: 10.1016/j.crad.2022.10.007

Chen, D., Bian, L., He, H.-Y., Li, Y.-D., Ma, C., and Mao, L.-G. (2022). Evaluation of traumatic subdural hematoma volume by using image segmentation assessment based on deep learning. *Comput. Math. Methods Med.* 2022:3830245. doi: 10.1155/2022/3830245

Chen, Y., Chen, J., Wei, D., Li, Y., and Zheng, Y., (Eds.) OctopusNet: a deep learning segmentation network for multi-modal medical images. Proceedings of the First International Workshop on Multiscale Multimodal Medical Imaging, MMMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China: Springer (2020).

Chen, Y., He, Y., Jiang, Z., Xie, Y., and Nie, S. (2023). Ischemic stroke subtyping method combining convolutional neural network and radiomics. *J. Xray Sci. Technol.* 31, 223–235. doi: 10.3233/XST-221284

Chen, G., Lu, M., Shi, Z., Xia, S., Ren, Y., Liu, Z., et al. (2020). Development and validation of machine learning prediction model based on computed tomography angiography-derived hemodynamics for rupture status of intracranial aneurysms: a Chinese multicenter study. *Eur. Radiol.* 30, 5170–5182. doi: 10.1007/s00330-020-06886-7

Chen, Y., Mancini, M., Zhu, X., and Akata, Z. (2022). Semi-supervised and unsupervised deep visual learning: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* doi: 10.1109/TPAMI.2022.3201576

Chen, R., Mo, X., Chen, Z., Feng, P., and Li, H. (2022). An integrated model combining machine learning and deep learning algorithms for classification of rupture status of IAs. *Front. Neurol.* 13:868395. doi: 10.3389/fneur.2022.1047876

Chen, G., Wei, X., Lei, H., Liqin, Y., Yuxin, L., Yakang, D., et al. (2020). Automated computer-assisted detection system for cerebral aneurysms in time-of-flight magnetic resonance angiography using fully convolutional network. *Biomed. Eng. Online* 19:38. doi: 10.1186/s12938-020-00770-7

Chen, G., Yifang, B., Jiajun, Z., Dongdong, W., Zhiyong, Z., Ruoyu, D., et al. (2023). Automated unruptured cerebral aneurysms detection in TOF MR angiography images using dual-channel SE-3D UNet: a multi-center research. *Eur. Radiol.* 33, 3532–3543. doi: 10.1007/s00330-022-09385-z

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., et al. (2018). Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* 392, 2388–2396. doi: 10.1016/S0140-6736(18)31645-3

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., and Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Transl. Vis. Sci. Technol.* 9:14. doi: 10.1167/tvst.9.2.14

Choi, J. M., Seo, S. Y., Kim, P. J., Kim, Y. S., Lee, S. H., Sohn, J. H., et al. (2021). Prediction of hemorrhagic transformation after ischemic Stroke using machine learning. *J. Pers. Med.* 11:863. doi: 10.3390/jpm11090863

Clerigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., and Lladó, X. (2019). Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Comput. Biol. Med.* 115:103487. doi: 10.1016/j.compbiomed.2019.103487

Cockroft, K. M., Jayaraman, M. V., Amin-Hanjani, S., Derdeyn, C. P., McDougall, C. G., and Wilson, J. A. (2012). A perfect storm: how a randomized trial of unruptured brain arteriovenous malformations' (ARUBA's) trial design challenges notions of external validity. *Stroke* 43, 1979–1981. doi: 10.1161/STROKEAHA.112.652032

Colasurdo, M., Shalev, D., Robledo, A., Vasandani, V., Luna, Z. A., Rao, A. S., et al. (2023). Validation of an automated machine learning algorithm for the detection and analysis of cerebral aneurysms. *J. Neurosurg.*, 1–8. doi: 10.3171/2023.1.JNS222304

Cumbler, E. (2015). In-hospital ischemic stroke. *Neurohospitalist* 5, 173–181. doi: 10.1177/1941874415588319

Czap, A. L., Bahr-Hosseini, M., Singh, N., Yamal, J. M., Nour, M., Parker, S., et al. (2022). Machine learning automated detection of large vessel occlusion from Mobile Stroke unit computed tomography angiography. *Stroke* 53, 1651–1656. doi: 10.1161/STROKEAHA.121.036091

de Vries, L., Emmer, B. J., Majoie, C. B., Marquering, H. A., and Gavves, E. (2023). PerfU-net: baseline infarct estimation from CT perfusion source data for acute ischemic stroke. *Med. Image Anal.*:102749. doi: 10.1016/j.media.2023.102749

Desai, V., Flanders, A., and Lakhani, P. (2017). Application of deep learning in neuroradiology: automated detection of basal ganglia hemorrhage using 2D-convolutional neural networks. *arXiv*. doi: 10.48550/arXiv.1710.03823

Detmer, F. J., Lückehe, D., Mut, F., Slawski, M., Hirsch, S., Bijlenga, P., et al. (2020). Comparison of statistical learning approaches for cerebral aneurysm rupture assessment. *Int. J. Comput. Assist. Radiol. Surg.* 15, 141–150. doi: 10.1007/s11548-019-02065-2

Dhar, R., Falcone, G. J., Chen, Y., Hamzehloo, A., Kirsch, E. P., Noche, R. B., et al. (2020). Deep learning for automated measurement of hemorrhage and Perihematomal edema in Supratentorial intracerebral hemorrhage. *Stroke* 51, 648–651. doi: 10.1161/STROKEAHA.119.027657

Dragoş, H. M., Stan, A., Pintican, R., Feier, D., Lebovici, A., Panaitescu, P., et al. (2023). MRI Radiomics and predictive models in assessing ischemic stroke outcome—a systematic review. *Diagnostics* 13:857. doi: 10.3390/diagnostics13050857

Duan, H., Huang, Y., Liu, L., Dai, H., Chen, L., and Zhou, L. (2019). Automatic detection on intracranial aneurysm from digital subtraction angiography with cascade convolutional neural networks. *Biomed. Eng. Online* 18:110. doi: 10.1186/s12938-019-0726-2

El-Ghanem, M., Kass-Hout, T., Kass-Hout, O., Alderazi, Y. J., Amuluru, K., Al-Mufti, F., et al. (2016). Arteriovenous malformations in the pediatric population: review of the existing literature. *Interv. Neurol.* 5, 218–225. doi: 10.1159/000447605

Fahmi, F., Marquering, H., Streekstra, G., Beenen, L., Velthuis, B., VanBavel, E., et al. (2012). Differences in CT perfusion summary maps for patients with acute ischemic stroke generated by 2 software packages. *Am. J. Neuroradiol.* 33, 2074–2080. doi: 10.3174/ajnr.A3110

Faron, A., Sichtermann, T., Teichert, N., Luetkens, J. A., Keulers, A., Nikoubashman, O., et al. (2020). Performance of a deep-learning neural network to detect intracranial aneurysms from 3D TOF-MRA compared to human readers. *Clin. Neuroradiol.* 30, 591–598. doi: 10.1007/s00062-019-00809-w

Fink, J. N., Kumar, S., Horkan, C., Linfante, I., Selim, M. H., Caplan, L. R., et al. (2002). The stroke patient who woke up: clinical and radiological features, including diffusion and perfusion MRI. *Stroke* 33, 988–993. doi: 10.1161/01.STR.0000014585.17714.67

Forkert, N. D., Verleger, T., Cheng, B., Thomalla, G., Hilgetag, C. C., and Fiehler, J. (2015). Multiclass support vector machine-based lesion mapping predicts functional outcome in ischemic stroke patients. *PLoS One* 10:e0129569. doi: 10.1371/journal.pone.0129569

Forman, R., Slota, K., Ahmad, F., Garg, R., John, S., Da Silva, I., et al. (2020). Intracerebral hemorrhage outcomes in the very elderly. *J. Stroke Cerebrovasc. Dis.* 29:104695. doi: 10.1016/j.jstrokecerebrovasdis.2020.104695

Foroushani, H. M., Hamzehloo, A., Kumar, A., Chen, Y., Heitsch, L., Slowik, A., et al. (2020). Quantitative serial CT imaging-derived features improve prediction of malignant cerebral edema after ischemic stroke. *Neurocrit. Care.* 33, 785–792. doi: 10.1007/s12028-020-01056-5

Fu, F., Shan, Y., Yang, G., Zheng, C., Zhang, M., Rong, D., et al. (2023). Deep learning for head and neck CT angiography: stenosis and plaque classification. *Radiology* 307:e220996. doi: 10.1148/radiol.220996

Goebel, J., Stenzel, E., Guberina, N., Wanke, I., Koehrmann, M., Kleinschnitz, C., et al. (2018). Automated ASPECT rating: comparison between the frontier ASPECT score software and the Brainomix software. *Neuroradiology* 60, 1267–1272. doi: 10.1007/s00234-018-2098-x

Goyal, M., Demchuk, A. M., Menon, B. K., Eesa, M., Rempel, J. L., Thornton, J., et al. (2015). Randomized assessment of rapid endovascular treatment of ischemic stroke. *N. Engl. J. Med.* 372, 1019–1030. doi: 10.1056/NEJMoa1414905

Green, T. L., McNair, N. D., Hinkle, J. L., Middleton, S., Miller, E. T., Perrin, S., et al. (2021). Care of the Patient with Acute Ischemic Stroke (Posthyperacute and prehospital discharge): update to 2009 comprehensive nursing care scientific statement: a scientific statement from the American Heart Association. *Stroke* 52, e179–e197. doi: 10.1161/STR.0000000000000357

Greving, J. P., Wermer, M. J., Brown, R. D. Jr., Morita, A., Juvela, S., Yonekura, M., et al. (2014). Development of the PHASES score for prediction of risk of rupture of intracranial aneurysms: a pooled analysis of six prospective cohort studies. *Lancet Neurol.* 13, 59–66. doi: 10.1016/S1474-4422(13)70263-1

Guédon, A., Thépenier, C., Shotar, E., Gabrieli, J., Mathon, B., Premat, K., et al. (2021). Predictive score for complete occlusion of intracranial aneurysms treated by flow-diverter stents using machine learning. *J. Neurointerv. Surg.* 13, 341–346. doi: 10.1136/neurintsurg-2020-016748

Guo, R., Zhang, R., Liu, R., Liu, Y., Li, H., Ma, L., et al. (2022). Machine learning-based approaches for prediction of patients' functional outcome and mortality after spontaneous intracerebral hemorrhage. *J. Pers. Med.* 12:112. doi: 10.3390/jpm12010112

Hainc, N., Mannil, M., Anagnostakou, V., Alkadhi, H., Blüthgen, C., Wacht, L., et al. (2020). Deep learning based detection of intracranial aneurysms on digital subtraction angiography: a feasibility study. *Neuroradiol. J.* 33, 311–317. doi: 10.1177/1971400920937647

Hanaoka, S., Nomura, Y., Takenaga, T., Murata, M., Nakao, T., Miki, S., et al. (2019). HoTPiG: a novel graph-based 3-D image feature set and its applications to computer-assisted detection of cerebral aneurysms and lung nodules. *Int. J. Comput. Assist. Radiol. Surg.* 14, 2095–2107. doi: 10.1007/s11548-019-01942-0

Heit, J. J., Honce, J. M., Yedavalli, V. S., Baccin, C. E., Tatit, R. T., Copeland, K., et al. (2022). RAPID aneurysm: artificial intelligence for unruptured cerebral aneurysm detection on CT angiography. *J. Stroke Cerebrovasc. Dis.* 31:106690. doi: 10.1016/j.jstrokecerebrovasdis.2022.106690

Helal, H. H. A. E., Bahnasy, W. S., Ghali, A. A., and Rabie, M. O. (2019). Early hematoma expansion in primary intracerebral hemorrhage: incidence and predictors. *Egypt. J. Neurol. Psychiat. Neurosurg.* 55:61. doi: 10.1186/s41983-019-0108-5

Hillal, A., Ullberg, T., Ramgren, B., and Wassélius, J. (2022). Computed tomography in acute intracerebral hemorrhage: neuroimaging predictors of hematoma expansion and outcome. *Insights Imaging* 13:180. doi: 10.1186/s13244-022-01309-1

Ho, K. C., Speier, W., El-Saden, S., and Arnold, C. W. (2017). Classifying acute ischemic Stroke onset time using deep imaging features. *AMIA Annu. Symp. Proc.* 2017, 892–901.

Hoffman, H., Wood, J. S., Cote, J. R., Jalal, M. S., Masoud, H. E., and Gould, G. C. (2023). Machine learning prediction of malignant middle cerebral artery infarction after mechanical thrombectomy for anterior circulation large vessel occlusion. *J. Stroke Cerebrovasc. Dis.* 32:106989. doi: 10.1016/j.jstrokecerebrovasdis.2023.106989

Hotta, K., Sorimachi, T., Osada, T., Baba, T., Inoue, G., Atsumi, H., et al. (2014). Risks and benefits of CT angiography in spontaneous intracerebral hemorrhage. *Acta Neurochir.* 156, 911–917. doi: 10.1007/s00701-014-2019-7

Hoving, J. W., Koopman, M. S., Tolhuisen, M. L., Voorst, H. V., Brehm, M., Berkhemer, O. A., et al. (2022). Accuracy of CT perfusion ischemic core volume and location estimation: a comparison between four ischemic core estimation approaches using syngo.Via. *PLoS One* 17:e0272276. doi: 10.1371/journal.pone.0272276

Ironside, N., Chen, C. J., Mutasa, S., Sim, J. L., Ding, D., Marfatiah, S., et al. (2020). Fully automated segmentation algorithm for perihematomal edema volumetry after spontaneous intracerebral hemorrhage. *Stroke* 51, 815–823. doi: 10.1161/STROKEAHA.119.026764

Ironside, N., Patrie, J., Ng, S., Ding, D., Rizvi, T., Kumar, J. S., et al. (2022). Quantification of hematoma and perihematomal edema volumes in intracerebral hemorrhage study: design considerations in an artificial intelligence validation (QUANTUM) study. *Clin. Trials* 19, 534–544. doi: 10.1177/17407745221105886

Jadhav, A., Kashyap, S., Bulu, H., Dholakia, R., Liu, A. Y., Syeda-Mahmood, T., et al. (2022). Towards automatic prediction of outcome in treatment of cerebral aneurysms. *AMIA Annu. Symp. Proc.* 2022, 570–579.

Jerman, T., Pernus, F., Likar, B., and Špiclin, Ž. (Eds). Aneurysm detection in 3D cerebral angiograms based on intra-vascular distance mapping and convolutional neural networks. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017); (2017), 18–21.

Jersey, A. M., and Foster, D. M.. *Cerebral aneurysm*. Treasure Island, FL: StatPearls Publishing (2023).

Jiao, Y., Zhang, J., Yang, X., Zhan, T., Wu, Z., Li, Y., et al. (2023). Artificial intelligence-assisted evaluation of the spatial relationship between brain arteriovenous malformations and the corticospinal tract to predict postsurgical motor defects. *AJNR Am. J. Neuroradiol.* 44, 17–25. doi: 10.3174/ajnr.A7735

Jin, H., Geng, J., Yin, Y., Hu, M., Yang, G., Xiang, S., et al. (2020). Fully automated intracranial aneurysm detection and segmentation from digital subtraction angiography series using an end-to-end spatiotemporal deep neural network. *J. Neurointerv. Surg.* 12, 1023–1027. doi: 10.1136/neurintsurg-2020-015824

Joo, B., Ahn, S. S., Yoon, P. H., Bae, S., Sohn, B., Lee, Y. E., et al. (2020). A deep learning algorithm may automate intracranial aneurysm detection on MR angiography with high diagnostic performance. *Eur. Radiol.* 30, 5785–5793. doi: 10.1007/s00330-020-06966-8

Jovel, J., and Greiner, R. (2021). An introduction to machine learning approaches for biomedical research. *Front. Med.* 8:771607. doi: 10.3389/fmed.2021.771607

Kaothanthong, N., Atsavasirilert, K., Sarampakhul, S., Chantangphol, P., Songsaeng, D., and Makhanov, S. (2022). Artificial intelligence for localization of the acute ischemic stroke by non-contrast computed tomography. *PLoS One* 17:e0277573. doi: 10.1371/journal.pone.0277573

Karamchandani, R. R., Helms, A. M., Satyanarayana, S., Yang, H., Clemente, J. D., Defilipp, G., et al. (2023). Automated detection of intracranial large vessel occlusions using Viz.ai software: experience in a large, integrated stroke network. *Brain Behav.* 13:e2808. doi: 10.1002/brb3.2808

Kim, B. J., Zhu, K., Qiu, W., Singh, N., McDonough, R., Cimflova, P., et al. (2023). Predicting DWI-FLAIR mismatch on NCCT: the role of artificial intelligence in hyperacute decision making. *Front. Neurol.* 14:1201223. doi: 10.3389/fneur.2023.1201223

Kleindorfer, D., Khoury, J., Alwell, K., Moomaw, C. J., Woo, D., Flaherty, M. L., et al. (2015). The impact of magnetic resonance imaging (MRI) on ischemic stroke detection and incidence: minimal impact within a population-based study. *BMC Neurol.* 15:175. doi: 10.1186/s12883-015-0421-2

Koopman, M. S., Berkhemer, O. A., Geuskens, R., Emmer, B. J., van Walderveen, M. A. A., Jenniskens, S. F. M., et al. (2019). Comparison of three commonly used CT perfusion software packages in patients with acute ischemic stroke. *J. Neurointerv. Surg.* 11, 1249–1256. doi: 10.1136/neurintsurg-2019-014822

Kuang, H., Najm, M., Chakraborty, D., Maraj, N., Sohn, S., Goyal, M., et al. (2019). Automated ASPECTS on noncontrast CT scans in patients with acute ischemic stroke using machine learning. *Am. J. Neuroradiol.* 40, 33–38. doi: 10.3174/ajnr.A5889

Kuo, W., Häne, C., Mukherjee, P., Malik, J., and Yuh, E. L. (2019). Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc. Natl. Acad. Sci. U. S. A.* 116, 22737–22745. doi: 10.1073/pnas.1908021116

Lago, A., Geffner, D., Tembl, J., Landete, L., Valero, C., and Baquero, M. (1998). Circadian variation in acute ischemic stroke: a hospital-based study. *Stroke* 29, 1873–1875. doi: 10.1161/01.STR.29.9.1873

Lanzagorta-Ortega, D., Carrillo-Pérez, D. L., and Carrillo-Esper, R. (2022). Artificial intelligence in medicine: present and future. *Gac. Med. Mex.* 158, 17–21. doi: 10.24875/GMM.M22000688

Laskowitz, D. T., Kasner, S. E., Saver, J., Remmel, K. S., and Jauch, E. C. (2009). Clinical usefulness of a biomarker-based diagnostic test for acute stroke: the biomarker rapid assessment in ischemic injury (BRAIN) study. *Stroke* 40, 77–85. doi: 10.1161/STROKEAHA.108.516377

Lauric, A., Miller, E., Frisken, S., and Malek, A. M. (2010). Automated detection of intracranial aneurysms based on parent vessel 3D analysis. *Med. Image Anal.* 14, 149–159. doi: 10.1016/j.media.2009.10.005

Lee, K. S., Zhang, J. J. Y., Alalade, A. F., Vine, R., Lanzino, G., Park, N., et al. (2021). Radiological surveillance of small unruptured intracranial aneurysms: a systematic review, meta-analysis, and meta-regression of 8428 aneurysms. *Neurosurg. Rev.* 44, 2013–2023. doi: 10.1007/s10143-020-01420-1

Li, P., Liu, Y., Zhou, J., Tu, S., Zhao, B., Wan, J., et al. (2023). A deep-learning method for the end-to-end prediction of intracranial aneurysm rupture risk. *Patterns* 4:100709. doi: 10.1016/j.patter.2023.100709

Li, Z., You, M., Long, C., Bi, R., Xu, H., He, Q., et al. (2020). Hematoma expansion in intracerebral hemorrhage: an update on prediction and treatment. *Front. Neurol.* 11:702. doi: 10.3389/fneur.2020.00702

Liu, J., Chen, Y., Lan, L., Lin, B., Chen, W., Wang, M., et al. (2018). Prediction of rupture risk in anterior communicating artery aneurysms with a feed-forward artificial neural network. *Eur. Radiol.* 28, 3268–3275. doi: 10.1007/s00330-017-5300-3

Liu, Q., Jiang, P., Jiang, Y., Ge, H., Li, S., Jin, H., et al. (2019). Prediction of aneurysm stability using a machine learning model based on PyRadiomics-derived morphological features. *Stroke* 50, 2314–2321. doi: 10.1161/STROKEAHA.119.025777

Liu, X., Mao, J., Sun, N., Yu, X., Chai, L., Tian, Y., et al. (2023). Deep learning for detection of intracranial aneurysms from computed tomography angiography images. *J. Digit. Imaging* 36, 114–123. doi: 10.1007/s10278-022-00698-5

Lu, J., Zhou, Y., Lv, W., Zhu, H., Tian, T., Yan, S., et al. (2022). Identification of early invisible acute ischemic stroke in non-contrast computed tomography using two-stage deep-learning model. *Theranostics* 12, 5564–5573. doi: 10.7150/thno.74125

Ma, C., Wang, L., Gao, C., Liu, D., Yang, K., Meng, Z., et al. (2022). Automatic and efficient prediction of hematoma expansion in patients with hypertensive intracerebral hemorrhage using deep learning based on CT images. *J. Pers. Med.* 12:779. doi: 10.3390/jpm12050779

Malhotra, K., Gornbein, J., and Saver, J. L. (2017). Ischemic strokes due to large-vessel occlusions contribute disproportionately to stroke-related dependence and death: a review. *Front. Neurol.* 8:651. doi: 10.3389/fneur.2017.00651

Menon, B. K., Puetz, V., Kochar, P., and Demchuk, A. M. (2011). ASPECTS and other neuroimaging scores in the triage and prediction of outcome in acute stroke patients. *Neuroimaging Clin.* 21, 407–423. doi: 10.1016/j.nic.2011.01.007

Mokin, M., Primiani, C. T., Siddiqui, A. H., and Turk, A. S. (2017). ASPECTS (Alberta Stroke program early CT score) measurement using Hounsfield unit values when selecting patients for Stroke Thrombectomy. *Stroke* 48, 1574–1579. doi: 10.1161/STROKEAHA.117.016745

Montaner, J., Mendioroz, M., Ribó, M., Delgado, P., Quintana, M., Penalba, A., et al. (2011). A panel of biomarkers including caspase-3 and D-dimer may differentiate acute stroke from stroke-mimicking conditions in the emergency department. *J. Intern. Med.* 270, 166–174. doi: 10.1111/j.1365-2796.2010.02329.x

Moor, J. (2006). The Dartmouth College artificial intelligence conference: the next fifty years. *AI Mag.* 27, 87–91.

Morita, A., Kirino, T., Hashi, K., Aoki, N., Fukuhara, S., Hashimoto, N., et al. (2012). The natural course of unruptured cerebral aneurysms in a Japanese cohort. *N. Engl. J. Med.* 366, 2474–2482. doi: 10.1056/NEJMoa1113260

Murray, N. M., Unberath, M., Hager, G. D., and Hui, F. K. (2020). Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J. Neurointerv. Surg.* 12, 156–164. doi: 10.1136/neurintsurg-2019-015135

Nagel, S., Sinha, D., Day, D., Reith, W., Chapot, R., Papanagiotou, P., et al. (2017). E-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed tomography scans of acute ischemic stroke patients. *Int. J. Stroke* 12, 615–622. doi: 10.1177/1747493016681020

Nakao, T., Hanaoka, S., Nomura, Y., Sato, I., Nemoto, M., Miki, S., et al. (2018). Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. *J. Magn. Reson. Imaging* 47, 948–953. doi: 10.1002/jmri.25842

Nguyen, V. N., Wallace, D., Ajmera, S., Akinduro, O., Smith, L. J., Giles, K., et al. (2020). Management of subdural hematohygromas in abusive head trauma. *Neurosurgery* 86, 281–287. doi: 10.1093/neuros/nyz076

Nishi, T., Yamashiro, S., Okumura, S., Takei, M., Tachibana, A., Akahori, S., et al. (2021). Artificial intelligence trained by deep learning can improve computed tomography diagnosis of nontraumatic subarachnoid hemorrhage by nonspecialists. *Neurol. Med. Chir.* 61, 652–660. doi: 10.2176/nmc.oa.2021-0124

Nishio, M., Koyasu, S., Noguchi, S., Kiguchi, T., Nakatsu, K., Akasaka, T., et al. (2020). Automatic detection of acute ischemic stroke using non-contrast computed tomography and two-stage deep learning model. *Comput. Methods Prog. Biomed.* 196:105711. doi: 10.1016/j.cmpb.2020.105711

Oermann, E. K., Rubinsteyn, A., Ding, D., Mascitelli, J., Starke, R. M., Bederson, J. B., et al. (2016). Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. *Sci. Rep.* 6:21161. doi: 10.1038/srep21161

Ou, C., Li, C., Qian, Y., Duan, C. Z., Si, W., Zhang, X., et al. (2022). Morphology-aware multi-source fusion-based intracranial aneurysms rupture prediction. *Eur. Radiol.* 32, 5633–5641. doi: 10.1007/s00330-022-08608-7

Ou, C., Qian, Y., Chong, W., Hou, X., Zhang, M., Zhang, X., et al. (2022). A deep learning-based automatic system for intracranial aneurysms diagnosis on three-dimensional digital subtraction angiographic images. *Med. Phys.* 49, 7038–7053. doi: 10.1002/mp.15846

Ozkara, B. B., Karabacak, M., Hamam, O., Wang, R., Kotha, A., Khalili, N., et al. (2023). Prediction of functional outcome in stroke patients with proximal middle cerebral artery occlusions using machine learning models. *J. Clin. Med.* 12:839. doi: 10.3390/jcm12030839

Paliwal, N., Jaiswal, P., Tutino, V. M., Shallwani, H., Davies, J. M., Siddiqui, A. H., et al. (2018). Outcome prediction of intracranial aneurysm treatment by flow diverters using machine learning. *Neurosurg. Focus.* 45:E7. doi: 10.3171/2018.8.FOCUS18332

Park, A., Chute, C., Rajpurkar, P., Lou, J., Ball, R. L., Shpanskaya, K., et al. (2019). Deep learning-assisted diagnosis of cerebral aneurysms using the HeadXNet model. *JAMA Netw. Open* 2:e195600. doi: 10.1001/jamanetworkopen.2019.5600

Peng, Q., Chen, X., Zhang, C., Li, W., Liu, J., Shi, T., et al. (2022). Deep learning-based computed tomography image segmentation and volume measurement of intracerebral hemorrhage. *Front. Neurosci.* 16:965680. doi: 10.3389/fnins.2022.965680

Pennig, L., Hoyer, U. C. I., Krauskopf, A., Shahzad, R., Jünger, S. T., Thiele, F., et al. (2021). Deep learning assistance increases the detection sensitivity of radiologists for

secondary intracranial aneurysms in subarachnoid hemorrhage. *Neuroradiology* 63, 1985–1994. doi: 10.1007/s00234-021-02697-9

Pollock, B. E., and Flickinger, J. C. (2002). A proposed radiosurgery-based grading system for arteriovenous malformations. *J. Neurosurg.* 96, 79–85. doi: 10.3171/jns.2002.96.1.0079

Polson, J. S., Zhang, H., Nael, K., Salamon, N., Yoo, B. Y., El-Saden, S., et al. (2022). Identifying acute ischemic stroke patients within the thrombolytic treatment window using deep learning. *J. Neuroimaging* 32, 1153–1160. doi: 10.1111/jon.13043

Powers, W. J., Derdeyn, C. P., Biller, J., Coffey, C. S., Hoh, B. L., Jauch, E. C., et al. (2015). 2015 American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early Management of Patients with Acute Ischemic Stroke Regarding Endovascular Treatment: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 46, 3020–3035. doi: 10.1161/STR.0000000000000074

Puhr-Westerheide, D., Froelich, M. F., Solyanik, O., Gresser, E., Reidler, P., Fabritius, M. P., et al. (2022). Cost-effectiveness of short-protocol emergency brain MRI after negative non-contrast CT for minor stroke detection. *Eur. Radiol.* 32, 1117–1126. doi: 10.1007/s00330-021-08222-z

Qureshi, A. I., and Palesch, Y. Y. (2011). Antihypertensive treatment of acute cerebral hemorrhage (ATACH) II: design, methods, and rationale. *Neurocrit. Care.* 15, 559–576. doi: 10.1007/s12028-011-9538-3

Rajabzadeh-Oghaz, H., Waqas, M., Veeturi, S. S., Vakharia, K., Tso, M. K., Snyder, K. V., et al. (2020). A data-driven model to identify high-risk aneurysms and guide management decisions: the rupture resemblance score. *J. Neurosurg.* 135, 9–16. doi: 10.3171/2020.5.JNS193264

Rinkel, G. J., Djibuti, M., Algra, A., and van Gijn, J. (1998). Prevalence and risk of rupture of intracranial aneurysms: a systematic review. *Stroke* 29, 251–256. doi: 10.1161/01.STR.29.1.251

Rymer, M. M. (2011). Hemorrhagic stroke: intracerebral hemorrhage. *Mo. Med.* 108, 50–54.

Saggi, S., Winkler, E. A., Ammanuel, S. G., Morshed, R. A., Garcia, J. H., Young, J. S., et al. (2022). Machine learning for predicting hemorrhage in pediatric patients with brain arteriovenous malformation. *J. Neurosurg. Pediatr.* 30, 203–209. doi: 10.3171/2022.4.PEDS21470

Sahlein, D. H., Gibson, D., Scott, J. A., DeNardo, A., Amuluru, K., Payner, T., et al. (2022). Artificial intelligence aneurysm measurement tool finds growth in all aneurysms that ruptured during conservative management. *J. Neurointerv. Surg.* 15, 766–770. doi: 10.1136/jnis-2022-019339

Saini, V., Guada, L., and Yavagal, D. R. (2021). Global epidemiology of Stroke and access to acute ischemic Stroke interventions. *Neurology* 97, S6–s16. doi: 10.1212/WNL.0000000000012781

Santana Baskar, P., Cordato, D., Wardman, D., and Bhaskar, S. (2021). In-hospital acute stroke workflow in acute stroke - systems-based approaches. *Acta Neurol. Scand.* 143, 111–120. doi: 10.1111/ane.13343

Sarker, I. H. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Comput. Sci.* 2:160. doi: 10.1007/s42979-021-00592-x

Selariu, E., Zia, E., Brizzi, M., and Abul-Kasim, K. (2012). Swirl sign in intracerebral haemorrhage: definition, prevalence, reliability and prognostic value. *BMC Neurol.* 12:109. doi: 10.1186/1471-2377-12-109

Serena, J., Dávalos, A., Segura, T., Mostacero, E., and Castillo, J. (2003). Stroke on awakening: looking for a more rational management. *Cerebrovasc. Dis.* 16, 128–133. doi: 10.1159/000070592

Shahzad, R., Pennig, L., Goertz, L., Thiele, F., Kabbasch, C., Schlamann, M., et al. (2020). Fully automated detection and segmentation of intracranial aneurysms in subarachnoid hemorrhage on CTA using deep learning. *Sci. Rep.* 10:21799. doi: 10.1038/s41598-020-78384-1

Shi, Z., Miao, C., Schoepf, U. J., Savage, R. H., Dargis, D. M., Pan, C., et al. (2020). A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. *Nat. Commun.* 11:6090. doi: 10.1038/s41467-020-19527-w

Shibata, S., Sakurai, K., Tachikawa, K., Ko, R., Hino, S., Fukano, T., et al. (2022). The utility of automated ASPECTS in acute ischemic stroke for intravenous recombinant tissue plasminogen activator (IV-rtPA) therapy. *Neurol. Int.* 14, 981–990. doi: 10.3390/neurolint14040077

Shimada, Y., Tanimoto, T., Nishimori, M., Choppin, A., Meir, A., Ozaki, A., et al. (2020). Incidental cerebral aneurysms detected by a computer-assisted detection system based on artificial intelligence: a case series. *Medicine* 99:e21518. doi: 10.1097/MD.0000000000021518

Shiraz Bhurwani, M. M., Waqas, M., Podgorsak, A. R., Williams, K. A., Davies, J. M., Snyder, K., et al. (2020). Feasibility study for use of angiographic parametric imaging and deep neural networks for intracranial aneurysm occlusion prediction. *J. Neurointerv. Surg.* 12, 714–719. doi: 10.1136/neurintsurg-2019-015544

Shuaib, A., Arian, H., and Shuaib, A. (2020). The increasing role of artificial intelligence in health care: will robots replace doctors in the future? *Int. J. Gen. Med.* 13, 891–896. doi: 10.2147/IJGM.S268093

Sichtermann, T., Faron, A., Sijben, R., Teichert, N., Freiherr, J., and Wiesmann, M. (2019). Deep learning-based detection of intracranial aneurysms in 3D TOF-MRA. *AJNR Am. J. Neuroradiol.* 40, 25–32. doi: 10.3174/ajnr.A5911

Sidorov, E., Sanghera, D. K., and Vanamala, J. K. P. (2019). Biomarker for ischemic stroke using metabolome: a clinician perspective. *J. Stroke* 21, 31–41. doi: 10.5853/jos.2018.03454

Smith, W. S., Lev, M. H., English, J. D., Camargo, E. C., Chou, M., Johnston, S. C., et al. (2009). Significance of large vessel intracranial occlusion causing acute ischemic stroke and TIA. *Stroke* 40, 3834–3840. doi: 10.1161/STROKEAHA.109.561787

Smith, W. S., Tsao, J. W., Billings, M. E., Johnston, S. C., Hemphill, J. C. 3rd, Bonovich, D. C., et al. (2006). Prognostic significance of angiographically confirmed large vessel intracranial occlusion in patients presenting with acute brain ischemia. *Neurocrit. Care* 4, 14–17. doi: 10.1385/NCC:4:1:014

Sohn, B., Park, K. Y., Choi, J., Koo, J. H., Han, K., Joo, B., et al. (2021). Deep learning-based software improves clinicians' detection sensitivity of aneurysms on brain TOF-MRA. *AJNR Am. J. Neuroradiol.* 42, 1769–1775. doi: 10.3174/ajnr.A7242

Soun, J. E., Chow, D. S., Nagamine, M., Takhtawala, R. S., Filippi, C. G., Yu, W., et al. (2021). Artificial intelligence and acute stroke imaging. *AJNR Am. J. Neuroradiol.* 42, 2–11. doi: 10.3174/ajnr.A6883

Spetzler, R. F., and Martin, N. A. (1986). A proposed grading system for arteriovenous malformations. *J. Neurosurg.* 65, 476–483. doi: 10.3171/jns.1986.65.4.0476

Starke, R. M., Yen, C. P., Ding, D., and Sheehan, J. P. (2013). A practical grading scale for predicting outcome after radiosurgery for arteriovenous malformations: analysis of 1012 treated patients. *J. Neurosurg.* 119, 981–987. doi: 10.3171/2013.5.JNS1311

Stember, J. N., Chang, P., Stember, D. M., Liu, M., Grinband, J., Filippi, C. G., et al. (2019). Convolutional neural networks for the detection and measurement of cerebral aneurysms on magnetic resonance angiography. *J. Digit. Imaging* 32, 808–815. doi: 10.1007/s10278-018-0162-z

Tanioka, S., Yago, T., Tanaka, K., Ishida, F., Kishimoto, T., Tsuda, K., et al. (2022). Machine learning prediction of hematoma expansion in acute intracerebral hemorrhage. *Sci. Rep.* 12:12452. doi: 10.1038/s41598-022-15400-6

Terasaki, Y., Yokota, H., Tashiro, K., Maejima, T., Takeuchi, T., Kurosawa, R., et al. (2021). Multidimensional deep learning reduces false-positives in the automated detection of cerebral aneurysms on time-of-flight magnetic resonance angiography: a multi-center study. *Front. Neurol.* 12:742126. doi: 10.3389/fneur.2021.742126

Thomalla, G., Cheng, B., Ebinger, M., Hao, Q., Tourdias, T., Wu, O., et al. (2011). DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4·5 h of symptom onset (PRE-FLAIR): a multicentre observational study. *Lancet Neurol.* 10, 978–986. doi: 10.1016/S1474-4422(11)70192-2

Thompson, B. G., Brown, R. D. Jr., Amin-Hanjani, S., Broderick, J. P., Cockroft, K. M., Connolly, E. S. Jr., et al. (2015). Guidelines for the Management of Patients with Unruptured Intracranial Aneurysms: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 46, 2368–2400. doi: 10.1161/STR.0000000000000070

Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., et al. (2022). Heart disease and stroke statistics-2022 update: a report from the American Heart Association. *Circulation* 145, e153–e639. doi: 10.1161/CIR.0000000000001052

Ueda, D., Yamamoto, A., Nishimori, M., Shimono, T., Doishita, S., Shimazaki, A., et al. (2019). Deep learning for MR angiography: automated detection of cerebral aneurysms. *Radiology* 290, 187–194. doi: 10.1148/radiol.2018180901

Vahidy, F., Nguyen, C., Albright, K. C., Boehme, A. K., Mir, O., Sands, K. A., et al. (2016). Transferring patients with intracerebral hemorrhage does not increase In-Hospital mortality. *PLoS One* 11:e0159174. doi: 10.1371/journal.pone.0159174

Vangen-Lønne, A. M., Wilsgaard, T., Johnsen, S. H., Løchen, M. L., Njølstad, I., and Mathiesen, E. B. (2017). Declining incidence of ischemic Stroke: what is the impact of changing risk factors? The Tromsø study 1995 to 2012. *Stroke* 48, 544–550. doi: 10.1161/STROKEAHA.116.014377

Vlak, M. H., Algra, A., Brandenburg, R., and Rinkel, G. J. (2011). Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. *Lancet Neurol.* 10, 626–636. doi: 10.1016/S1474-4422(11)70109-0

Wada, R., Aviv, R. I., Fox, A. J., Sahlas, D. J., Gladstone, D. J., Tomlinson, G., et al. (2007). CT angiography "spot sign" predicts hematoma expansion in acute intracerebral hemorrhage. *Stroke* 38, 1257–1262. doi: 10.1161/01.STR.0000259633.59404.f3

Wang, C. W., Juan, C. J., Liu, Y. J., Hsu, H. H., Liu, H. S., Chen, C. Y., et al. (2009). Volume-dependent overestimation of spontaneous intracerebral hematoma volume by the ABC/2 formula. *Acta Radiol.* 50, 306–311. doi: 10.1080/02841850802647039

Wang, Y.-J., Li, Z.-X., Gu, H.-Q., Zhai, Y., Jiang, Y., Zhao, X.-Q., et al. (2019). China stroke statistics 2019: a report from the National Center for Healthcare Quality Management in Neurological Diseases, China National Clinical Research Center for Neurological Diseases, the Chinese Stroke Association, National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention and Institute for Global Neuroscience and Stroke Collaborations. *Stroke Vasc. Neurol.* 5, 211–239. doi: 10.1136/svn-2020-000457

Wang, X., Shen, T., Yang, S., Lan, J., Xu, Y., Wang, M., et al. (2021). A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. *Neuroimage Clin.* 32:102785. doi: 10.1016/j.nicl.2021.102785

Wang, G., Song, T., Dong, Q., Cui, M., Huang, N., and Zhang, S. (2020). Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by

image synthesis and attention-based deep neural networks. *Med. Image Anal.* 65:101787. doi: 10.1016/j.media.2020.101787

Wang, T., Song, N., Liu, L., Zhu, Z., Chen, B., Yang, W., et al. (2021). Efficiency of a deep learning-based artificial intelligence diagnostic system in spontaneous intracerebral hemorrhage volume measurement. *BMC Med. Imaging* 21:125. doi: 10.1186/s12880-021-00657-6

Wang, S., and Summers, R. M. (2012). Machine learning and radiology. *Med. Image Anal.* 16, 933–951. doi: 10.1016/j.media.2012.02.005

Wardlaw, J. M., Farrall, A. J., Perry, D., von Kummer, R., Mielke, O., Moulin, T., et al. (2007). Factors influencing the detection of early CT signs of cerebral ischemia: an internet-based, international multiobserver study. *Stroke* 38, 1250–1256. doi: 10.1161/01.STR.0000259715.53166.25

Webb, A. J., Ullman, N. L., Morgan, T. C., Muschelli, J., Kornbluth, J., Awad, I. A., et al. (2015). Accuracy of the ABC/2 score for intracerebral hemorrhage: systematic review and analysis of MISTIE, CLEAR-IVH, and CLEAR III. *Stroke* 46, 2470–2476. doi: 10.1161/STROKEAHA.114.007343

Williams, K. A., Podgorsak, A. R., Bhurwani, M. M. S., Rava, R. A., Sommer, K. N., and Ionita, C. N. (2021). The aneurysm occlusion assistant, an AI platform for real time surgical guidance of intracranial aneurysms. *Proc. SPIE Int. Soc. Opt. Eng.* 11601, 153–160. doi: 10.1117/12.2581003

Xiong, W., Chen, T., Li, J., Xiang, L., Zhang, C., Xiang, L., et al. (2022). Interpretable machine learning model to predict rupture of small intracranial aneurysms and facilitate clinical decision. *Neurol. Sci.* 43, 6371–6379. doi: 10.1007/s10072-022-06351-x

Yaghi, S., Dibu, J., Achi, E., Patel, A., Samant, R., and Hinduja, A. (2014). Hematoma expansion in spontaneous intracerebral hemorrhage: predictors and outcome. *Int. J. Neurosci.* 124, 890–893. doi: 10.3109/00207454.2014.887716

Yang, X., Blezek, D. J., Cheng, L. T., Ryan, W. J., Kallmes, D. F., and Erickson, B. J. (2011). Computer-aided detection of intracranial aneurysms in MR angiography. *J. Digit. Imaging* 24, 86–95. doi: 10.1007/s10278-009-9254-0

Yang, J., Xie, M., Hu, C., Alwalid, O., Xu, Y., Liu, J., et al. (2021). Deep learning for detecting cerebral aneurysms with CT angiography. *Radiology* 298, 155–163. doi: 10.1148/radiol.2020192154

Yoo, A. J., Zaidat, O. O., Chaudhry, Z. A., Berkhemer, O. A., González, R. G., Goyal, M., et al. (2014). Impact of pretreatment noncontrast CT Alberta Stroke program early CT score on clinical outcome after intra-arterial stroke therapy. *Stroke* 45, 746–751. doi: 10.1161/STROKEAHA.113.004260

Yu, Y., Guo, D., Lou, M., Liebeskind, D., and Scalzo, F. (2018). Prediction of hemorrhagic transformation severity in acute Stroke from source perfusion MRI. *I.E.E.E. Trans. Biomed. Eng.* 65, 2058–2065. doi: 10.1109/TBME.2017.2783241

Yu, N., Yu, H., Li, H., Ma, N., Hu, C., and Wang, J. (2022). A robust deep learning segmentation method for hematoma volumetric detection in intracerebral hemorrhage. *Stroke* 53, 167–176. doi: 10.1161/STROKEAHA.120.032243

Zeng, Y., Liu, X., Xiao, N., Li, Y., Jiang, Y., Feng, J., et al. (2020). Automatic diagnosis based on spatial information fusion feature for intracranial aneurysm. *IEEE Trans. Med. Imaging* 39, 1448–1458. doi: 10.1109/TMI.2019.2951439

Zhang, Y., Zhu, D., Li, T., Wang, X., Zhao, L., Yang, X., et al. (2022). Detection of acute ischemic stroke and backtracking stroke onset time via machine learning analysis of metabolomics. *Biomed. Pharmacother.* 155:113641. doi: 10.1016/j.biopha.2022.113641

Zhong, J. W., Jin, Y. J., Song, Z. J., Lin, B., Lu, X. H., Chen, F., et al. (2021). Deep learning for automatically predicting early haematoma expansion in Chinese patients. *Stroke Vasc. Neurol.* 6, 610–614. doi: 10.1136/svn-2020-000647

Zhu, H., Jiang, L., Zhang, H., Luo, L., Chen, Y., and Chen, Y. (2021). An automatic machine learning approach for ischemic stroke onset time identification based on DWI and FLAIR imaging. *Neuroimage Clin.* 31:102744. doi: 10.1016/j.nicl.2021.102744

Zhu, W., Li, W., Tian, Z., Zhang, Y., Wang, K., Zhang, Y., et al. (2020). Stability assessment of intracranial aneurysms using machine learning based on clinical and morphological features. *Transl. Stroke Res.* 11, 1287–1295. doi: 10.1007/s12975-020-00811-2

Glossary

AI	Artificial intelligence
ML	Machine learning
DL	Deep learning
AVM	Arteriovenous malformation.
ICH	Intracerebral hemorrhage
IA	Intracranial aneurysm
CVA	Cerebrovascular accident
NCCT	Non-contrast computed tomography
ANN	Artificial neural network
CNN	Convolutional neural network
RNN	Recurrent neural network
IVH	Intraventricular hemorrhage
SAH	Subarachnoid hemorrhage
SDH	Subdural hematoma
EDH	Epidural hematoma
IPH	Intraparenchymal hemorrhage
CTA	Computed tomography angiogram
CTA	Computed tomography perfusion
DSA	Digital subtraction angiography
MRI	Magnetic resonance imaging
HE	Hematoma expansion
PHE	Perihematomal edema
SVM	Support vector machine
k-NN	k-nearest neighbor
DSC	Dice similarity coefficient
AVD	Average volume difference
AUC	Area under the curve
DWI	Diffusion-weighted imaging
TOAST	Trial of Org in Acute Stroke Treatment
FLAIR	Fluid-attenuated inversion recovery
ASPECTS	Alberta Stroke Program Computed Tomography Score
TOF	Time-of-flight
SE	Squeeze-and-excitation
RBAS	Radiosurgery-based AVM Score
VRAS	Virginia Radiosurgery AVM Scale
EMR	Electronic medical record
PACS	Picture archiving/communication systems

frontiers | Frontiers in Surgery

TYPE Original Research
PUBLISHED 15 September 2023
DOI 10.3389/fsurg.2023.1222859

OPEN ACCESS

EDITED BY
Philipp Taussky,
Harvard Medical School, United States

REVIEWED BY
Mirza Pojskic,
University Hospital of Giessen and Marburg, Germany,
Xiao Xiao,
Perception & AI Technologies Ltd., China

*CORRESPONDENCE
Aure Enkaoua
aure.enkaoua.17@ucl.ac.uk

†These authors share last authorship.

RECEIVED 15 May 2023
ACCEPTED 11 August 2023
PUBLISHED 15 September 2023

CITATION
Enkaoua A, Islam M, Ramalhinho J, Dowrick T, Booker J, Khan DZ, Marcus HJ and Clarkson MJ (2023) Image-guidance in endoscopic pituitary surgery: an in-silico study of errors involved in tracker-based techniques.
Front. Surg. 10:1222859.
doi: 10.3389/fsurg.2023.1222859

Image-guidance in endoscopic pituitary surgery: an in-silico study of errors involved in tracker-based techniques

Aure Enkaoua[1]*, Mobarakol Islam[1], João Ramalhinho[1], Thomas Dowrick[1], James Booker[1,2], Danyal Z. Khan[1,2], Hani J. Marcus[1,2†] and Matthew J. Clarkson[1†]

[1]Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK, [2]Division of Neurosurgery, National Hospital for Neurology and Neurosurgery, London, UK

Background: Endoscopic endonasal surgery is an established minimally invasive technique for resecting pituitary adenomas. However, understanding orientation and identifying critical neurovascular structures in this anatomically dense region can be challenging. In clinical practice, commercial navigation systems use a tracked pointer for guidance. Augmented Reality (AR) is an emerging technology used for surgical guidance. It can be tracker based or vision based, but neither is widely used in pituitary surgery.
Methods: This pre-clinical study aims to assess the accuracy of tracker-based navigation systems, including those that allow for AR. Two setups were used to conduct simulations: (1) the standard pointer setup, tracked by an infrared camera; and (2) the endoscope setup that allows for AR, using reflective markers on the end of the endoscope, tracked by infrared cameras. The error sources were estimated by calculating the Euclidean distance between a point's true location and the point's location after passing it through the noisy system. A phantom study was then conducted to verify the in-silico simulation results and show a working example of image-based navigation errors in current methodologies.
Results: The errors of the tracked pointer and tracked endoscope simulations were 1.7 and 2.5 mm respectively. The phantom study showed errors of 2.14 and 3.21 mm for the tracked pointer and tracked endoscope setups respectively.
Discussion: In pituitary surgery, precise neighboring structure identification is crucial for success. However, our simulations reveal that the errors of tracked approaches were too large to meet the fine error margins required for pituitary surgery. In order to achieve the required accuracy, we would need much more accurate tracking, better calibration and improved registration techniques.

KEYWORDS
augmented reality, pituitary surgery, computer-assisted surgery, tracking, neurosurgery

1. Introduction

The pituitary gland is situated within an exceptionally dense anatomical region, surrounded by critical neurovascular structures such as the optic nerves and internal carotid arteries (1). There is significant anatomical variation between patients, and pituitary tumours often distort this complex anatomy, making safe recognition and avoidance of critical structures difficult during surgery (2). The current gold standard

surgical technique for the resection of pituitary adenomas is through a transsphenoidal approach (3). The endoscopic endonasal transsphenoidal approach allows for excellent wide-angle visualisation, but almost always relies on a monocular endoscopic camera, resulting in limited depth perception, which can further impair appreciation of critical structures (4).

Surgical navigation systems are established adjuncts used to support intra-operative orientation and navigation. The most used navigation tool is a tracked pointer, where a set of reflective markers are placed at the top of the pointer. An infrared (IR) camera tracks the markers, allowing the location of the pointer's tip to be visualised on the pre-operative scan. This approach is cognitively demanding for the surgeon, as the surgeon must map the position of the pointer displayed on the pre-operative MRI scan onto the live endoscopic video. Moreover, it impacts the surgical workflow as the surgeon needs to repeatedly stop operating, remove their instruments, and place the probe into the operative field. Therefore, alternative techniques that remove the need for multiple displays and for manual placement of probes to regions of interest may allow for both a reduced cognitive load and an improved surgical workflow.

Augmented reality (AR) is an emerging display technology that allows structures of interest from a pre-operative MRI to be displayed directly onto the live endoscopic video. AR has already been used in several surgical procedures with varying success (5, 6). Within pituitary surgery, several research groups have previously reported the use of tracker-based AR (7, 8) but it has not been widely adopted in routine practice. A US survey that was conducted to investigate the use of intra-operative neuronavigation found that only 7% of cases used image guidance systems (9). Despite there being AR products for microscopic surgery such as the SyncAR (10), to the best of our knowledge there are no approved AR devices in endoscopic pituitary surgery. Emerging alternatives such as vision-based techniques have also been proposed although not yet widely used (11, 12).

The aim of this study was to assess the accuracy of tracker-based navigation systems, including those used for AR systems found in the research literature.

2. Materials and methods

2.1. Study overview

We adopted a simulation-based study methodology to assess whether, under standard conditions, tracker-based guidance is sufficiently accurate to allow for guidance during pituitary adenoma resection. Simulations were developed for two different setups- a standard surgical tracked pointer setup; and a tracked endoscope setup that enabled augmented reality. The pointer and endoscope both had reflective markers, tracked by an IR camera for localisation purposes. A system overview of these setups can be seen in **Figure 1** and will be described in detail in the following sections. We then conducted a phantom study to further validate our simulation results. The system setup can be seen in **Figure 2**.

In the navigation methods described, the shared objective is to present pre-operative imaging information to the surgeon in an intuitive fashion. This is considered as a geometrical problem where each radiological image and the physical operating theatre are described by specific coordinates which must be mapped to each other. This mapping is done via mathematical transformations and allows for information from one coordinate system to be displayed onto another. In the presented navigation methods, the IR camera can track the 3-dimensional locations of all the reflective markers. A unique grouping of reflective markers attached to a tool defines a local coordinate system. In this paper, a mathematical transform is defined as ${}^{B}T_{A}$ and maps a point from the coordinate system A to the coordinate system B. Transformations are assumed to be rigid and can therefore only be composed of rotations and translations.

In the following subsections, the study design is presented to define the chosen geometrical values used in the setup of the simulations. This is followed by the two different simulation setups. Finally, we describe how noise was simulated to investigate its effects on the system accuracy.

A phantom study was subsequently conducted to verify and demonstrate the results of the simulations and show a working example of the image-based navigation errors in pituitary surgery with current methodologies.

2.2. Simulations

2.2.1. Study design

For the purpose of the simulations, the layout of the tracker and tracked tools are defined based on realistic estimates of the physical layout in a typical pituitary surgery environment performed at a single academic neurosurgical centre.

The simulated endoscope is based on the geometry of the Karl Storz Hopkins telescope 7230 0° of length 180 mm. The camera projection is modelled using a pinhole model and calibrated using Zhang's camera calibration algorithm (13, 14). The pointer used in the simulation is the NDI pointer part number 8700340 with a length of 160 mm. A diagram of the components involved in each simulation can be seen in **Figure 1**. For both simulations, the patient is placed in front of the IR camera at a distance of 2 m. The patient is attached to a Mayfield clamp which has an NDI reference marker part number 8700339 attached to it. The distance between the patient and the reference is simplified and estimated to be 0.3 m superior to the pituitary gland.

Using the known dimensions and relative positions of the different coordinate systems, a mathematical transform can be found. With this transform, a virtual point in one coordinate system (e.g., MRI) can be converted to another coordinate system (e.g., live endoscopic video).

When the IR cameras are localising the position of the reflective markers, there is an associated error by which the IR camera localises the markers, referred to as the volumetric accuracy (σ) (15). The value of the volumetric accuracy σ by which an IR camera can locate markers can vary depending on the design of the

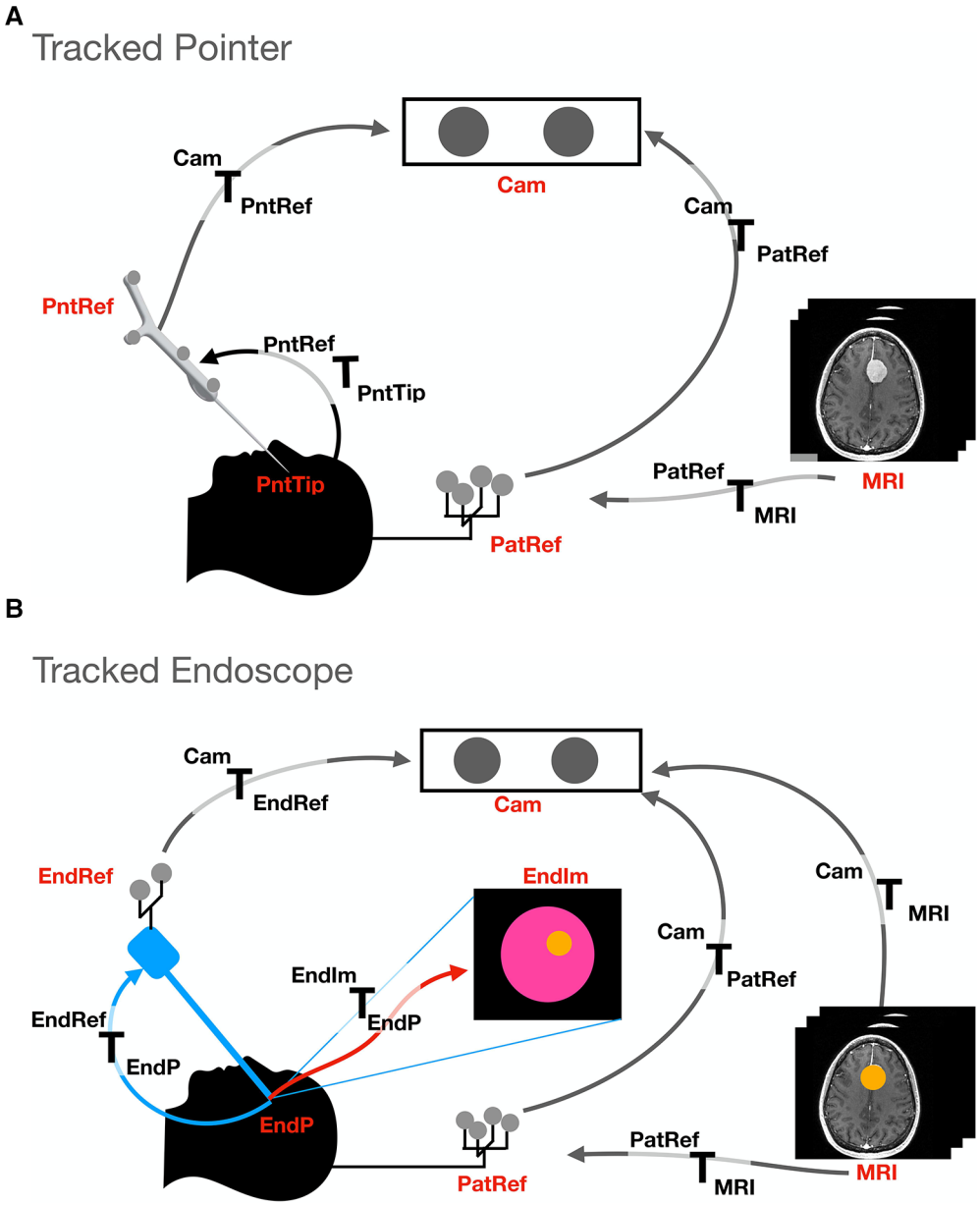


FIGURE 1
Transformations and setup involved in the two different tracking methods presented: (A). Tracked pointer setup, (B). Tracked endoscope setup. The abbreviations used stand for the different coordinate systems and are as follows- camera (Cam), pointer reference (PntRef), endoscope reference (EndRef), patient according to the mayfield clamp reference (PatRef), MRI, pointer's tip (PntTip), endoscope tip (EndP) and endoscope video frames (EndIm).

camera. Different models such as the NDI Polaris Vega and Vicra have σ values of 0.12 and 0.25 mm respectively.[1, 2, 3]

Moreover, most models have a specified working range, where outside the working range the error increases, and when outside the field of view, the device stops tracking the markers accurately. The volumetric precision quoted in this article is an error value of 0.2 mm as the average found by Koivukangas et al. (15). However, values between 0–0.5 mm were investigated and can be accessed in the **Supplementary material**. During this simulation, the volumetric accuracy was modelled by adding noise using a Gaussian distribution with varying standard deviation σ.

The effects of noise are then calculated quantitatively by obtaining the target registration error (TRE). TRE is defined as

[1]NDI polaris vega: https://www.ndigital.com/optical-measurement-technology/polaris-vega/
[2]NDI polaris vicra: https://www.ndigital.com/optical-measurement-technology/polaris-vicra/
[3]NDI website: https://www.ndigital.com

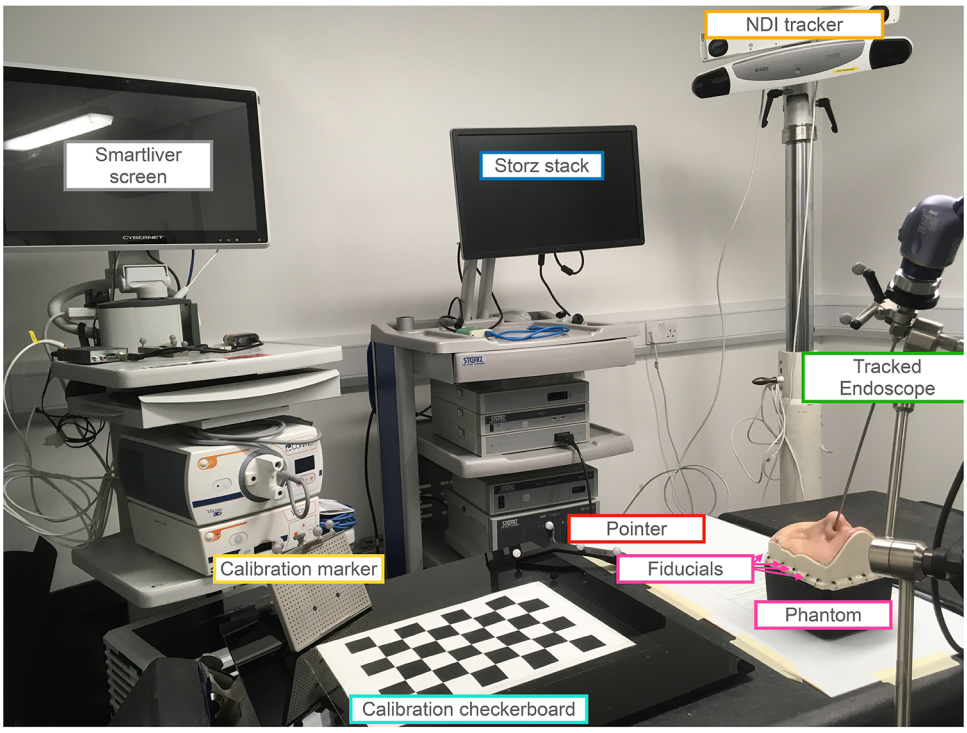


FIGURE 2
Setup for the phantom study involving both the tracked pointer and tracked endoscope setups. The endoscope is securely held in place by a clamp and positioned inside the nostril of the UpSurgeOn phantom. The phantom itself features 20 fiducials attached around its base, serving to align the CT scan and phantom to a common optical tracker space. An infrared NDI tracker is used to locate the position of the markers on the endoscope, pointer and calibration markers. The endoscope is calibrated for the AR display using the calibration checkerboard. The AR overlay is displayed on the Smartliver screen, allowing for real-time visualisation of the AR display. The Storz stack facilitates the connection of the endoscope camera and displays the endoscope's output, providing guidance on the precise placement of the pointer tip.

the Euclidean distance between a point's true location and the same point's location after adding noise to the measurements and propagating errors through the system (16). Therefore, in order to calculate errors of different systems, the TRE is calculated between the target point set as ground truth and the same point after being transformed with the noisy transforms. To provide a meaningful average, the process of adding noise as described above is repeated over 10,000 samples of noise.

2.2.2. Simulation types

2.2.2.1. Tracked pointer setup

The goal of the setup is to be able to locate the pointer's tip on the MRI scan to improve surgeons' orientation. A transform ${}^{MRI}T_{PntTip}$ is therefore obtained to convert a point from the pointer's tip coordinate system to the MRI coordinate system.

Figure 1A shows the different coordinate systems of the pointer setup. The transformations between the different coordinate systems of the simulation can be deduced since the relative positions and dimensions are known by design. For example, the transform between the pointer tip and the reference is a simple translation of the pointer's length in the y direction of the pointer's tip coordinate system (see Figure 3). This is performed for all coordinate systems-pointer reference (*PntRef*) to camera (*cam*), camera to patient reference (*PatRef*) and patient reference to *MRI*. Once individual transformations have been obtained, it is possible to multiply any point in the pointer's coordinate system by these transformations to obtain the same point in *MRI* coordinates.

2.2.2.2. Tracked endoscope setup

Unlike the tracked pointer, the goal of the endoscope setup is to have an AR display of the MRI scan in endoscope coordinates. A detailed map of the involved components can be seen in Figure 1B. In this case, it is necessary to obtain ${}^{EndIm}T_{MRI}$ that will transform a point from the MRI coordinate system to each endoscopic video frame (*EndIm*). The relative positions of all components are known, enabling a point to be converted from the coordinate system of the MRI to the patient reference (*PatRef*), from *PatRef* to camera (*cam*), from *cam* to endoscope reference (*EndRef*) and from *EndRef* to the endoscope tip (*EndP*). Finally, the point is projected from a 3D point in *EndP* coordinates onto the 2D endoscopic video coordinate system (*EndIm*).

2.2.3. Noise and analysis

In the two tracked setups, each transformation is prone to noise. In this section, we describe how we add realistic levels of

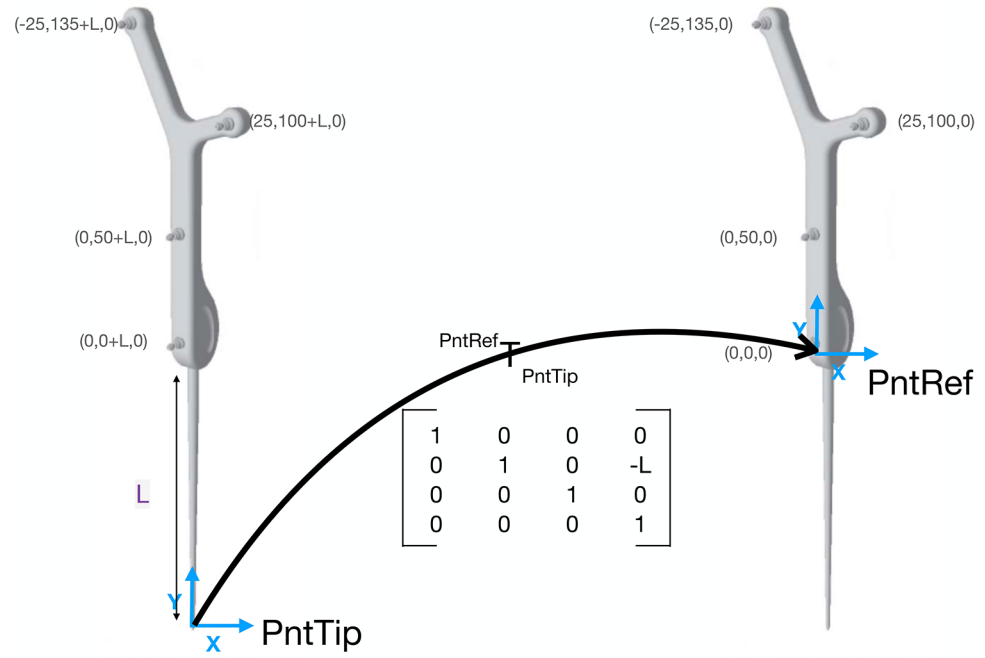


FIGURE 3
Example transformation of the tracked pointer. The coordinates of the pointer's markers are taken from the NDI documentation sheet. The transformation ${}^{PntRef}T_{PntTip}$ is a 4 × 4 transformation matrix. If any point in the *PntTip* coordinate system is multiplied by this matrix, the point will be now in *PntRef* coordinates. The transformation is a simple translation in the y direction as the only difference is the length of the pointer.

noise onto simulated measurements and investigate what its effects are on the system accuracy.

The following errors were associated with the tracked pointer approach:

2.2.3.1. Tracking noise

Since tracked tools were used in this method, there is a localisation error associated with the markers. This was modeled by adding Gaussian noise with standard deviation of 0.2 mm to the location of each marker, and recomputing the tracking transformation.

2.2.3.2. Tool length effect

The effects of different pointer lengths on the tracking noise were investigated. The longer the distance between the tool's end and the markers, the larger the associated tracking error will be as the errors magnify with increasing distance. The lengths investigated were between 100 to 300 mm in steps of 10 mm increments. The TRE associated with the tracking noise was generated for each of the lengths.

2.2.3.3. Surface based registration

The total system error will also be affected by the registration transformation between the patient's coordinate system and the MRI coordinate system. The effect of registration error was studied by adding noise to the rotation (Euler angles in °) and translation (displacement in mm) parameters of the registration transform.

In the endoscope simulation, the same investigations on tracking error, tool length and surface-based registration error as mentioned in the pointer simulation were performed. However, there were also some additional sources of error with this setup.

2.2.3.4. Hand-eye calibration error

The errors involved in the calibration of the transformation between the tracking marker on the endoscope and the camera coordinate system of the endoscope were simulated. The TRE was calculated after adding noise to the translations and Euler angles composing the camera hand-eye matrix.

2.2.3.5. AR

As the endoscope setup investigates an AR display, this error can also be expressed in camera space, measured in pixels. In this simulation, noise was added to all the transformations mentioned. Once noise was added, our interest point was projected from 3D to 2D using all the noisy transformations. The TRE was then calculated between the point transformed with the gold-standard transforms and the point transformed with the noisy transforms.

2.3. Phantom study

The setup of the phantom experiment can be seen in Figure 2. The phantom was the UpSurgeOn BrainBox TNS model, which is used and validated for simulation training of the transsphenoidal endonasal approach (17). A CT scan of the phantom was obtained with the Medtronic O-arm CT O2 Intraoperative Imaging System.[4]

[4]Medtronic website: https://www.medtronic.com/uk-en/index.html

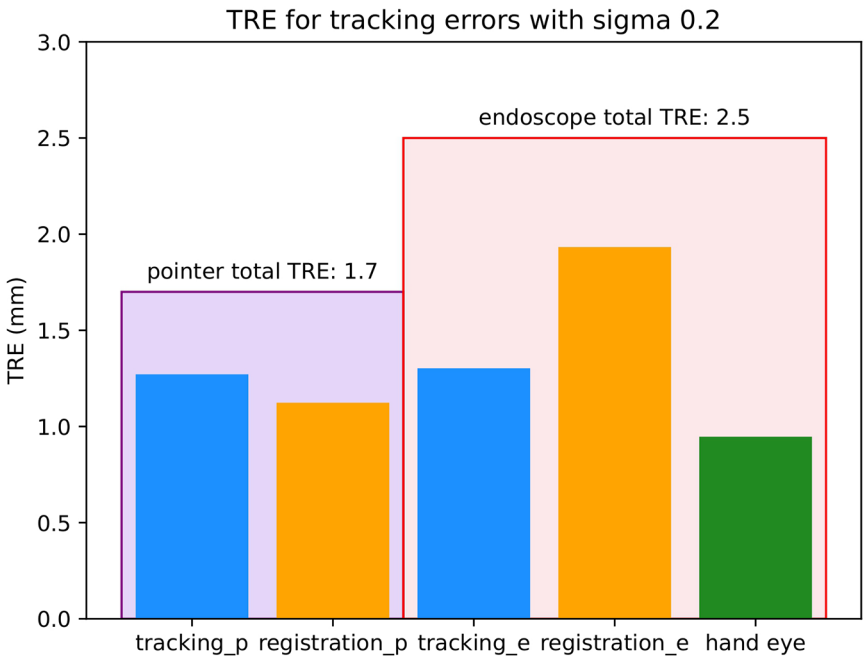


FIGURE 4
Target registration errors (TREs) involved for each of the simulations with σ 0.2 mm. All units are in mm. The large encompassing purple bar on the left represents the total TRE of the pointer simulation and the large encompassing red bar on the right represents the total TRE of the tracked endoscope simulation. Within each of these bars there are the different sources of error involved in that setup. The different sources of error are tracking (blue) which is the localisation error of the IR cameras, registration (orange) which is the surface-based registration error of MRI to patient coordinates and hand eye (green) which is the hand eye error of the endoscope.

The endoscope used was the 30 cm Storz model number 27020 AA.[5] The pointer was the same as in the simulations, of model number 8700340, the markers placed on the endoscope for tracking were the NDI reference marker part number 8700449 and the NDI polaris Vega[6] tracked the marker locations. The Smartliver software was used (18–20) to obtain the AR view and pointer locations.

In order to match the phantom and CT to the common optical tracker space, fiducials were attached along the base of the phantom box as seen in the setup (**Figure 2**) and their ground truth locations were obtained manually from the CT scan. The same locations were then sampled with the tracked pointer tip and could therefore be matched using point-based registration using the procrustes algorithm (21).

To replicate the pointer simulation and calculate the TRE, a screw was placed on the tumour as seen in **Figure 6** to represent a target location. The pointer was then passed through the nostrill of the phantom to place its tip on the target location. Once placed on the target, the pointer tip location was recorded and averaged to represent a single point. The TRE was obtained by calculating the Euclidean Distance between the ground truth location from the CT scan, and the location obtained by the pointer when placed on the screw after being converted to CT coordinates with the point-based registration.

To replicate the tracked endoscope AR system, the surface of the tumour was segmented. This was done with the NifTK manual segmentation toolkit method slice by slice and a polygon model was then generated also using NifTK (22). The AR display was obtained using the Smartliver software (18–20) and recorded. The endoscope camera was calibrated using Zhang's camera calibration method (14). The location of the board was tracked with an NDI reference marker part number 8700449.

To illustrate the errors caused by the tracking system, the tracked reprojection error was calculated. Since the coordinates of the chessboard corners are known, their 3D locations can be converted to the endoscope camera coordinates and projected to 2D to obtain the difference between the detected corners and the projected corners. A video of the overlay was also obtained for qualitative illustration purposes.

3. Results

Figure 4 shows a summary of the TREs involved in each of the simulations.

3.1. Simulations

3.1.1. Tracked pointer

The tracking and surface-based registration errors of the pointer were simulated. As mentioned previously, the length of

[5]storz website: https://www.karlstorz.com/gb/en/index.htm
[6]NDI website: https://www.ndigital.com

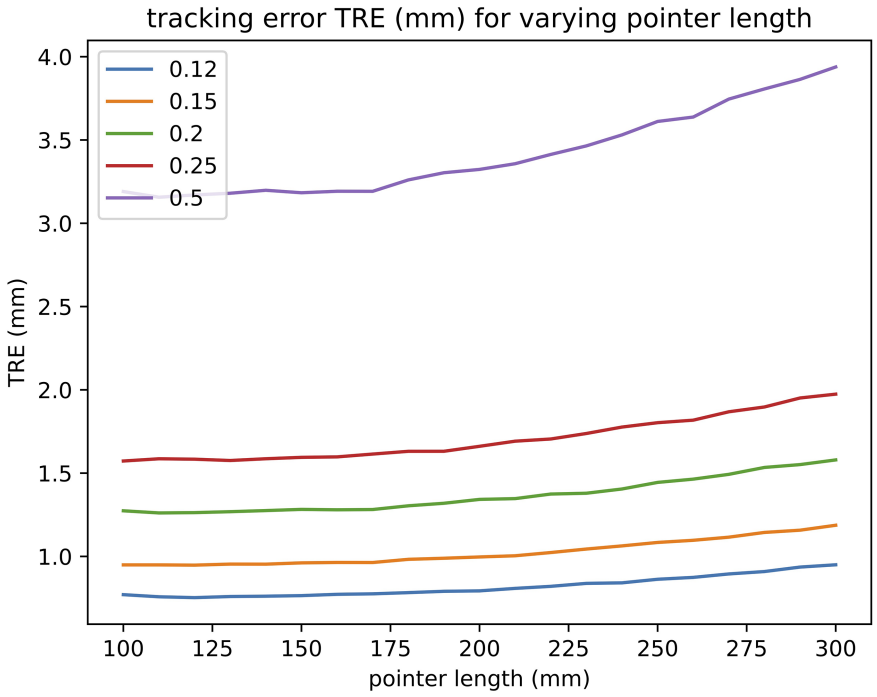


FIGURE 5
Effect of tool lengths on TRE. The different lines represent different values of σ- 0.12, 0.15, 0.20, 0.25 and 0.5 mm.

the pointer used in a single neurosurgical centre served as the model for this investigation. However, as pointer length can vary, the effect of tool length was also investigated on the tracking noise as seen in Figure 5. The total TRE of the pointer with σ 0.2 mm as seen in Figure 4 was composed of 2 main errors- tracking and surface-based registration errors, with values of 1.3 and 1.1 mm respectively. The TRE with all sources of error was 1.7 mm.

3.1.2. Tracked endoscope

In this simulation, two different accuracies can be quoted, one for 3D and one for the 2D (AR) errors. The 3D TRE of the 180 mm endoscope as seen in Figure 4 was composed of 3 main errors- a tracking error of 1.3 mm, a surface-based registration error of 1.9 mm and a hand-eye error of 0.9 mm. The TRE in 3D with all sources of error added was 2.5 mm. The total TRE in 2D for the AR display was 29 pixels.

3.2. Phantom

In the phantom study, the registration used for both simulations had an error of 1.03 mm. The calibration error with Zhang's calibration algorithm used for the tracked endoscope setup was 0.66 mm.

The TRE obtained by the tracked pointer phantom study was 2.14 mm. The tracked monocular reprojection error of the AR system was 3.21 mm, and the AR display obtained by the tracked endoscope can be seen in Figure 6

4. Discussion

4.1. Principal findings

In this simulation-based study we have, for the first time, demonstrated that tracker-based AR devices are likely to be insufficiently accurate to allow for guidance during pituitary adenoma resection. Therefore at present, the benefits of improved orientation using AR would be outweighed by the compounded errors associated with its use.

The typical size of a pituitary gland is approximately 10 mm in width and 5 mm in height (23). When the tracking camera has a σ of 0.2 mm, the error of the tracked pointer setup is 1.7 mm and that of the tracked endoscope setup is 2.5 mm. This would result in an AR display where the offset is almost half the pituitary itself, which is likely more confusing to a surgeon than helpful if used to guide resection. From the phantom study, we were able to obtain a visual representation of what the AR display may look like. As seen in Figure 6, the model of the tumour is not overlayed on the phantom's tumour and is therefore more distracting than helpful in aiding navigation. In order to achieve a reasonable visualisation with a tracked approach, the total error of tracking, registration and any other source such as hand-eye would need to be below 1 mm. In order to achieve this accuracy, we would need much more accurate tracking, better calibration and improved registration techniques.

4.2. Findings in context of existing literature

Navigation systems have a long history within neurosurgery, and particularly within neuro-oncology, where they can facilitate

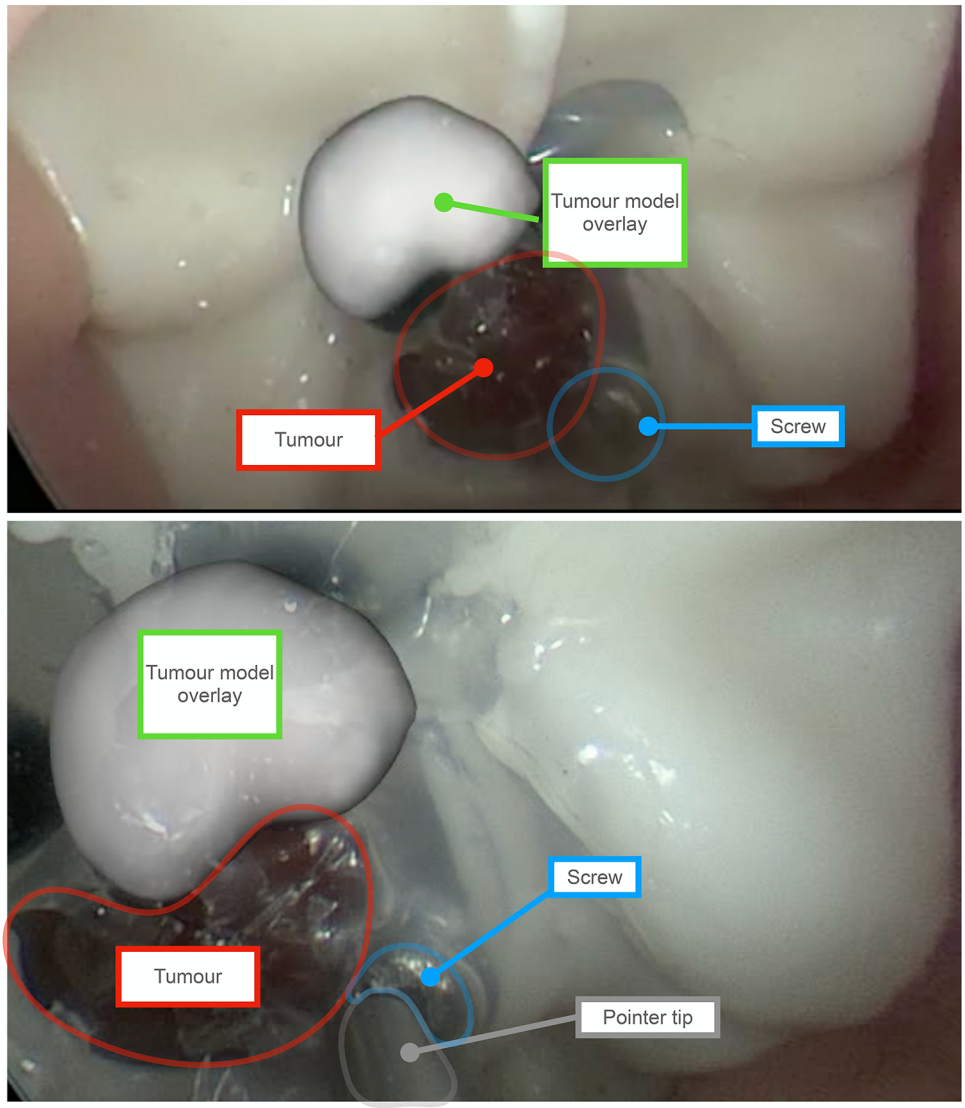


FIGURE 6
Augmented Reality (AR) overlay obtained using the tracked endoscope method. Outlines and labels are added to enhance readability. The red outline represents the phantom's tumor, and the green label indicates the AR overlay of the tumor, showcasing an offset caused by system errors. The overlay is a 3D model of the tumour generated by manually segmenting the tumor on each slice of a CT scan of the phantom. For scale perspective, the tumor on the phantom measures approximately 8 mm in diameter. The blue outline marks the screw, which serves as the target point for the tracked pointer phantom experiment. As shown in the bottom view, the pointer tip aligns with the screw's location. Since the precise position of the screw is known on the CT scan of the phantom, the error can be determined by calculating the difference from this location.

surgical trajectory planning, resulting in shorter incisions, smaller craniotomies, and more limited brain exposure and retraction (24). Within pituitary surgery, a recent registry study reported that the use of neuronavigation was associated with improved surgical safety (25), including a reduced rate of complications such as cerebrospinal fluid (CSF) leak (26). However, the same study found that navigation was only used in approximately one in ten pituitary operations.

Limitations of current navigation systems include the need for the surgeon to mentally map the location of a point identified on the pre-operative MRI onto the live endoscopic video, and the need to repeatedly interrupt the surgery to place the pointer. This was captured by a survey that was taken on the Society of British Neurosurgeons (SBNS). Key findings were the need for "better integration with image-guidance systems (20%)," and a call for "intra-operative visualisation and improvements in neuroendoscopy (49%)" (27). One potential solution to these limitations is the use of AR, which allows for the fusion of the pre-operative MRI and live endoscope video into a single display and can do so on-demand rather than requiring a probe to be placed.

To date, there are several reports of AR systems used within pituitary surgery, dating as early as 20 years ago (8, 12, 28–43). However, no AR systems have been widely adopted despite the aforementioned stated advantages. The findings from this paper suggest that one barrier to uptake is insufficient accuracy. Although only slightly more inaccurate than standard pointer-based navigation (1.7 versus 2.5 mm), the fact

that AR displays are overlaid onto the live endoscopic video make this obvious and distracting. By comparison, when using standard pointer-based techniques to identify a location on the pre-operative MRI, an experienced surgeon is more likely able to accommodate inaccuracies and interpolate the true location. The distractions posed by inaccurate AR may contribute to inattention blindness, which has also a recognised concern with such systems, and is thought to reflect cognitive overload (44, 45).

Automatic intra-operative CT scanning (iCT) is a promising technique that can be used to improve registration and can therefore boost the accuracy of the AR system (31, 46). However, its adoption faces many challenges. Firstly, it disrupts the surgical workflow as the operation has to be stopped while the image is acquired. It also increases radiation dose to both the patient and the hospital staff in the room. The head positioning needs to be altered and cannot be placed at the angle that is comfortable for operation. Finally, this method is expensive, rendering it unfeasible for low and middle-income countries.

Recently, alternatives to tracker-based navigation have emerged. The work of Mirota et al. (11, 12) introduces a vision-based system that directly matches the endoscopic video and the MRI scan without the need for a tracking system. The system functions by performing a 3D reconstruction of the endoscopic video by extracting and matching features between subsequent frames and estimating the motion between the frames. The 3D reconstructed image can then be registered with the pre-operative CT/MRI scan. Even though this paper was proposed in 2009–2011, this research has not yet reached any clinically viable solutions and there is more research to be done, especially given recent advances in deep learning techniques.

4.3. Strengths and limitations

The applicability of this study can be extended to various scenarios using the provided code in the **Supplementary material**. However, there are several points that were not accounted for in the simulation.

The primary limitation lies in how the setup estimates were derived. In order to perform the simulation, the relative positions of coordinate systems had to be determined. However, different centres may have different tools or a different setup than the one simulated. Although simulations of different tool lengths and IR camera volumetric accuracies were performed in the **Supplementary material**, other setup variations such as the patient-to-camera position or the relative locations of the reference coordinate systems, may influence the results as they can vary across cases. This is one of the reasons why the results of the tracked endoscope from the phantom study appeared larger than the simulations. The length of the endoscope in the simulation was 180 mm whereas the one used in the phantom experiment was 300 mm and therefore this adds to the total error of the system.

Another such point is the working range of the camera. If the tracking balls of the pointer are located outside the working range of the IR camera, the volumetric accuracy drops. Therefore, it is important to note the possibility that during a surgery, the localisation accuracy of the pointer setup changes if the surgical bed moves either too close or too far from the camera and lies outside the working range of the camera. Tracking can also stop if some of the markers are occluded.

Any tracked methods require an initial registration performed before each surgery to align the patient coordinates to the pre-operative MRI scan. Even if a highly-accurate IR camera were available in practice, the calibration errors such as the hand-eye calibration and the registration used when obtaining the transformation ${}^{PatRef}T_{MRI}$ would also need to be below 0.5 mm. This is currently unlikely in the case of surface-based registration. Other investigations comparing registration with point-based adhesive markers and surface-based registration quoted errors with surface-based registration averaging over 5 mm (47). Even though our simulations are highly analytical, the accuracy of a registration algorithm is ultimately determined by how well the surgeons perform the registration. That means that the errors will be dependent on the training provided and any time limitations the surgeon may have when performing the registration. This is also reflected in the phantom study, where the errors of the pointer and endoscope were larger than in the corresponding simulation studies by 0.41 mm and 0.71 mm respectively. The simulations we developed are simply a mathematical representation of the two setups. However, the errors are also ultimately dependent on factors that cannot be simulated such as how well the user performs registration or calibration, the model of the tracking camera, and its location in the room.

4.4. Conclusions and future work

The findings of this study demonstrate that a tracker-based system alone is insufficiently accurate to allow for AR in pituitary adenoma resection. To this end, future work is merited to develop either purely vision-based or hybrid vision- and tracker-based alternatives to support AR in this context.

Data availability statement

All data generated or analysed during this study is publicly available in the **Supplementary Material**.

Author contributions

AE: Conceptualization, Methodology, Software, Writing - Original Draft. MJC: Conceptualization, Methodology, Software, Writing - review and editing. HJM: Conceptualization, Writing - review and editing. JR, TD: Methodology, Writing - review and

editing. MI, JB, DZK: Conceptualization, Writing - review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Intelligent, Integrated Imaging in Healthcare (i4Health) [EP/S021930/1], the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145Z/16/Z), and EPSRC grant [EP/W00805X/1]. Hani J Marcus is supported by WEISS and by the UCL/UCLH BRC Neuroscience. Danyal Z Khan is supported by an NIHR Academic Clinical Fellowship and a Cancer Research UK Predoctoral Fellowship. João Ramalhinho is funded by the EPSRC grant EP/T029404/1. For the purpose of open access, the author has applied a CC BY public copyright licence to any author-accepted manuscript version arising from this submission.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsurg.2023.1222859/full#supplementary-material

References

1. Solari D, Villa A, De Angelis M, Esposito F, Cavallo LM, Cappabianca P. Anatomy, surgery of the endoscopic endonasal approach to the skull base. *Transl Med UniSa*. (2012) 2:36.

2. Hamid O, El Fiky L, Hassan O, Kotb A, El Fiky S. Anatomic variations of the sphenoid sinus, their impact on trans-sphenoid pituitary surgery. *Skull Base*. (2008) 18:9–15. doi: 10.1055/s-2007-992764

3. Daly AF, Rixhon M, Adam C, Dempegioti A, Tichomirowa MA, Beckers A. High prevalence of pituitary adenomas: a cross-sectional study in the province of liege, Belgium. *J Clin Endocrinol Metab*. (2006) 91:4769–75. doi: 10.1210/jc.2006-1668

4. Marcus HJ, Hughes-Hallett A, Cundy TP, Di Marco A, Pratt P, Nandi D, et al. Comparative effectiveness of 3-dimensional vs 2-dimensional, high-definition vs standard-definition neuroendoscopy: a preclinical randomized crossover study. *Neurosurgery*. (2014) 74:375. doi: 10.1227/NEU.0000000000000249

5. Collins T, Pizarro D, Gasparini S, Bourdel N, Chauvet P, Canis M, et al. Augmented reality guided laparoscopic surgery of the uterus. *IEEE Trans Med Imaging*. (2020) 40:371–80. doi: 10.1109/TMI.2020.3027442

6. Shuhaiber JH. Augmented reality in surgery. *Arch Surg*. (2004) 139:170–4. doi: 10.1001/archsurg.139.2.170

7. Onishi K, Fumiyama S, Nonaka M, Koeda M, Noborio H. Study on the image overlay approach to ar navigation system for transsphenoidal surgery. In *International Conference on Human-Computer Interaction*. Springer (2021). p. 625–43.

8. Kawamata T, Iseki H, Shibasaki T, Hori T. Endoscopic augmented reality navigation system for endonasal transsphenoidal surgery to treat pituitary tumors. *Neurosurgery*. (2002) 50:1393–7. doi: 10.1227/00006123-200206000-00038

9. Chung TK, Riley KO, Woodworth BA. The use of image-guidance during transsphenoidal pituitary surgery in the united states. *Am J Rhinol Allergy*. (2015) 29:215–20. doi: 10.2500/ajra.2015.29.4166

10. Louis RG, Steinberg GK, Duma C, Britz G, Mehta V, Pace J, et al. Early experience with virtual and synchronized augmented reality platform for preoperative planning and intraoperative navigation: a case series. *Oper Neurosurg*. (2021) 21:189. doi: 10.1093/ons/opab188

11. Mirota D, Wang H, Taylor RH, Ishii M, Hager GD. Toward video-based navigation for endoscopic endonasal skull base surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (2009). p. 91–9.

12. Mirota DJ, Wang H, Taylor RH, Ishii M, Gallia GL, Hager GD. A system for video-based navigation for endoscopic endonasal skull base surgery. *IEEE Trans Med Imaging*. (2011) 31:963–76. doi: 10.1109/TMI.2011.2176500

13. Prince S, *Computer vision: models learning and inference*. New York: Cambridge University Press (2012).

14. Zhang Z. A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell*. (2000) 22:1330–4. doi: 10.1109/34.888718

15. Koivukangas T, Katisko J, Koivukangas JP. Technical accuracy of optical, the electromagnetic tracking systems. *SpringerPlus*. (2013) 2:1–7. doi: 10.1186/2193-1801-2-90

16. Schlegel M. Predicting the accuracy of optical tracking systems. München, Germany: Technical University of Munich.

17. Newall N, Khan DZ, Hanrahan JG, Booker J, Borg A, Davids J, et al. High fidelity simulation of the endoscopic transsphenoidal approach: validation of the UpSurgeOn TNS Box. *Front Surg*. (2022) 9:1049685. doi: 10.3389/fsurg.2022.1049685

18. Thompson S, Totz J, Song Y, Johnsen S, Stoyanov D, Ourselin S, et al. Smartliver image guidance system for laparoscopic liver resection. Imperial College London.

19. Schneider C, Thompson S, Totz J, Song Y, Allam M, Sodergren M, et al. Comparison of manual and semi-automatic registration in augmented reality image-guided liver surgery: a clinical feasibility study. *Surg Endosc*. (2020) 34:4702–11. doi: 10.1007/s00464-020-07807-x

20. Thompson S , Dowrick T , Ahmad M , Xiao G , Koo B , Bonmati E , et al. SciKit-Surgery: compact libraries for surgical navigation. *Int J Comput Assist Radiol Surg*. (2020) 15(7):1075–84. doi: 10.1007/s11548-020-02180-5

21. Arun KS, Huang TS, Blostein SD. Least-squares fitting of two 3-D point sets. *IEEE Trans Pattern Anal Mach Intell*. (1987) 9(5):698–700. doi: 10.1109/TPAMI.1987.4767965

22. Clarkson MJ, Zombori G, Thompson S, Totz J, Song Y, Espak M, et al. The NifTK software platform for image-guided interventions: platform overview and NiftyLink messaging. *Int J Comput Assist Radiol Surg*. (2015) 10:301–16. doi: 10.1007/s11548-014-1124-7

23. Doraiswamy PM, Potts JM, Axelson DA, Husain MM, Lurie SN, Na C, et al. Mr assessment of pituitary gland morphology in healthy volunteers: age-and gender-related differences. *Am J Neuroradiol*. (1992) 13:1295–9.

24. Paleologos TS, Wadley JP, Kitchen ND, Thomas DG. Clinical utility and cost-effectiveness of interactive image-guided craniotomy: clinical comparison between conventional and image-guided meningioma surgery. *Neurosurgery*. (2000) 47:40–8. doi: 10.1227/00006123-200007000-00010

25. Achey RL, Karsy M, Azab MA, Scoville J, Kundu B, Bowers CA, et al. Improved surgical safety via intraoperative navigation for transnasal transsphenoidal resection of pituitary adenomas. *J Neurol Surg B Skull Base*. (2019) 80:626–31. doi: 10.1055/s-0039-1677677

26. Chung TK, Riley KO, Woodworth BA. The use of image-guidance during transsphenoidal pituitary surgery in the united states. *Am J Rhinol Allergy*. (2015) 29:215–20. doi: 10.2500/ajra.2015.29.4166

27. Marcus HJ, Cundy TP, Hughes-Hallett A, Yang G-Z, Darzi A, Nandi D. Endoscopic and keyhole endoscope-assisted neurosurgical approaches: a qualitative survey on technical challenges and technological solutions. *Br J Neurosurg*. (2014) 28:606–10. doi: 10.3109/02688697.2014.887654

28. Thavarajasingam SG, Vardanyan R, Arjomandi Rad A, Thavarajasingam A, Khachikyan A, Mendoza N, et al. The use of augmented reality in transsphenoidal surgery: a systematic review. *Br J Neurosurg*. (2022) 36:457–71. doi: 10.1080/02688697.2022.2057435

29. Pennacchietti V, Stoelzel K, Tietze A, Lankes E, Schaumann A, Uecker FC, et al. First experience with augmented reality neuronavigation in endoscopic assisted midline skull base pathologies in children. *Childs Nerv Syst*. (2021) 37:1525–34. doi: 10.1007/s00381-021-05049-3

30. Dixon BJ, Daly MJ, Chan H, Vescan A, Witterick IJ, Irish JC. Augmented image guidance improves skull base navigation and reduces task workload in trainees: a preclinical trial. *Laryngoscope*. (2011) 121:2060–4. doi: 10.1002/lary.22153

31. Lai M, Skyrman S, Shan C, Babic D, Homan R, Edström E, et al. Fusion of augmented reality imaging with the endoscopic view for endonasal skull base surgery; a novel application for surgical navigation based on intraoperative cone beam computed tomography and optical tracking. *PLoS ONE*. (2020) 15: e0227312. doi: 10.1371/journal.pone.0227312

32. Bong JH, Song H-J, Oh Y, Park N, Kim H, Park S. Endoscopic navigation system with extended field of view using augmented reality technology. *Int J Med Robot Comput Assist Surg*. (2018) 14:e1886. doi: 10.1002/rcs.1886

33. Citardi MJ, Agbetoba A, Bigcas J-L, Luong A. Augmented reality for endoscopic sinus surgery with surgical navigation: a cadaver study. *Int Forum Allergy Rhinol*. (2016) 6:523–8. doi: 10.1002/alr.21702

34. Lapeer R, Chen M, Gonzalez G, Linney A, Alusi G. Image-enhanced surgical navigation for endoscopic sinus surgery: evaluating calibration, registration and tracking. *Int J Med Robot Comput Assist Surg*. (2008) 4:32–45. doi: 10.1002/rcs.175

35. Linxweiler M, Pillong L, Kopanja D, Kühn JP, Wagenpfeil S, Radosa JC, et al. Augmented reality-enhanced navigation in endoscopic sinus surgery: a prospective, randomized, controlled clinical trial. *Laryngoscope Investig Otolaryngol*. (2020) 5:621–9. doi: 10.1002/lio2.436

36. Dixon BJ, Chan H, Daly MJ, Vescan AD, Witterick IJ, Irish JC. The effect of augmented real-time image guidance on task workload during endoscopic sinus surgery. *Int Forum Allergy Rhinol*. (2012) 2:405–10. doi: 10.1002/alr.21049

37. Zeiger J, Costa A, Bederson J, Shrivastava RK, Iloreta AM. Use of mixed reality visualization in endoscopic endonasal skull base surgery. *Oper Neurosurg*. (2020) 19:43–52. doi: 10.1093/ons/opz355

38. Caversaccio M, Langlotz F, Nolte L-P, Häusler R. Impact of a self-developed planning and self-constructed navigation system on skull base surgery: 10 years experience. *Acta Otolaryngol*. (2007) 127:403–7. doi: 10.1080/00016480601002104

39. Li L, Yang J, Chu Y, Wu W, Xue J, Liang P, et al. A novel augmented reality navigation system for endoscopic sinus and skull base surgery: a feasibility study. *PLoS ONE*. (2016) 11:e0146996. doi: 10.1371/journal.pone.0146996

40. Carl B, Bopp M, Voellger B, Saß B, Nimsky C. Augmented reality in transsphenoidal surgery. *World Neurosurg*. (2019) 125:e873–83. doi: 10.1016/j.wneu.2019.01.202

41. Dixon BJ, Daly MJ, Chan H, Vescan A, Witterick IJ, Irish JC. Augmented real-time navigation with critical structure proximity alerts for endoscopic skull base surgery. *Laryngoscope*. (2014) 124:853–9. doi: 10.1002/lary.24385

42. Prisman E, Daly MJ, Chan H, Siewerdsen JH, Vescan A, Irish JC. Real-time tracking and virtual endoscopy in cone-beam CT-guided surgery of the sinuses and skull base in a cadaver model. *Int Forum Allergy Rhinol*. (2011) 1:70–7. doi: 10.1002/alr.20007

43. Onishi K, Fumiyama S, Miki Y, Nonaka M, Koeda M, Noborio H. Study on the development of augmented-reality navigation system for transsphenoidal surgery. In *Human-Computer Interaction. Human Values and Quality of Life: Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Proceedings, Part III 22*; 2020 Jul 19–24; Copenhagen, Denmark. Springer (2020). p. 623–38.

44. Marcus HJ, Pratt P, Hughes-Hallett A, Cundy TP, Marcus AP, Yang G-Z, et al. Comparative effectiveness and safety of image guidance systems in neurosurgery: a preclinical randomized study. *J Neurosurg*. (2015) 123:307–13. doi: 10.3171/2014.10.JNS141662

45. Hughes-Hallett A, Mayer EK, Marcus HJ, Pratt P, Mason S, Darzi AW, et al. Inattention blindness in surgery. *Surg Endosc*. (2015) 29:3184–9. doi: 10.1007/s00464-014-4051-3

46. Bopp MH, Saß B, Pojskić M, Corr F, Grimm D, Kemmling A, et al. Use of neuronavigation and augmented reality in transsphenoidal pituitary adenoma surgery. *J Clin Med*. (2022) 11:5590. doi: 10.3390/jcm11195590

47. Mongen MA, Willems PW. Current accuracy of surface matching compared to adhesive markers in patient-to-image registration. *Acta Neurochir (Wien)*. (2019) 161:865–70. doi: 10.1007/s00701-019-03867-8

frontiers | Frontiers in Surgery

TYPE Systematic Review
PUBLISHED 27 September 2023
DOI 10.3389/fsurg.2023.1227510

OPEN ACCESS

EDITED BY
Mohammed Ali Alvi,
University Health Network (UHN), Canada

REVIEWED BY
Cristian Luciano,
University of Illinois Chicago, United States
Bernhard Preim,
Otto von Guericke University Magdeburg, Germany

*CORRESPONDENCE
Elisa Colombo
✉ Elisa.colombo@usz.ch

RECEIVED 31 May 2023
ACCEPTED 18 September 2023
PUBLISHED 27 September 2023

CITATION
Colombo E, Lutters B, Kos T and van Doormaal T (2023) Application of virtual and mixed reality for 3D visualization in intracranial aneurysm surgery planning: a systematic review.
Front. Surg. 10:1227510.
doi: 10.3389/fsurg.2023.1227510

Application of virtual and mixed reality for 3D visualization in intracranial aneurysm surgery planning: a systematic review

Elisa Colombo[1]*, Bart Lutters[2], Tessa Kos[3] and Tristan van Doormaal[1]

[1]Department of Neurosurgery and Klinisches Neurozentrum Zurich ZH, Universität Zürich; Universitätsspital Zürich, Zurich, Switzerland, [2]Julius Center for Health Sciences and Primary Care, Medical Humanities, University Medical Center Utrecht, Utrecht, Netherlands, [3]Image Science Institute, University Medical Center Utrecht, Utrecht, Netherlands

Background: Precise preoperative anatomical visualization and understanding of an intracranial aneurysm (IA) are fundamental for surgical planning and increased intraoperative confidence. Application of virtual reality (VR) and mixed reality (MR), thus three-dimensional (3D) visualization of IAs could be significant in surgical planning. Authors provide an up-to-date overview of VR and MR applied to IA surgery, with specific focus on tailoring of the surgical treatment.
Methods: A systematic analysis of the literature was performed in accordance with the PRISMA guidelines. Pubmed, and Embase were searched to identify studies reporting use of MR and VR 3D visualization in IA surgery during the last 25 years. Type and number of IAs, category of input scan, visualization techniques (screen, glasses or head set), inclusion of haptic feedback, tested population (residents, fellows, attending neurosurgeons), and aim of the study (surgical planning/rehearsal, neurosurgical training, methodological validation) were noted.
Results: Twenty-eight studies were included. Eighteen studies (64.3%) applied VR, and 10 (35.7%) used MR. A positive impact on surgical planning was documented by 19 studies (67.9%): 17 studies (60.7%) chose the tailoring of the surgical approach as primary outcome of the analysis. A more precise anatomical visualization and understanding with VR and MR was endorsed by all included studies (100%).
Conclusion: Application of VR and MR to perioperative 3D visualization of IAs allowed an improved understanding of the patient-specific anatomy and surgical preparation. This review describes a tendency to utilize mostly VR-platforms, with the primary goals of a more accurate anatomical understanding, surgical planning and rehearsal.

KEYWORDS
cerebrovascular surgery, intracranial aneurysms, virtual reality, mixed reality, 3D visualization

Introduction

Intracranial aneurysms (IAs) are pathological dilatations of cerebral arteries. IAs are relatively commonly acquired lesions occurring with a frequency ranging between 0.5% and 3% in the general population, and accounting for about 80%–85% of non-traumatic subarachnoid hemorrhages (1). Upon detection of an IA, tailoring of the optimal treatment strategy is based on careful consideration of the patient history and specific

aneurysm characteristics. Treatment approaches are surgical and/or endovascular. With advances in endovascular approaches, the indications for surgical clipping of IAs have been decreased. Currently, open IA clipping is generally reserved for complex aneurysms. Successful and safe surgery of these cases depends on accurate surgical planning, which implies precise pre-operative characterization of lesion-specific anatomical features. The current gold standard imaging modality for the preoperative study of IAs is digital subtraction angiography (DSA). DSA allows a comprehensive anatomical examination of the most relevant IAs' features (i.e.,: relation to the parent vessels, neck's width, dome's regularity and orientation) at the cost of invasiveness. The role of magnetic resonance flow (MR-flow) has indeed been increasing for the diagnosis and the preoperative analysis of IAs. Nonetheless, MR studies are mostly black-and-white and visualized on two-dimensional (2D) screens. When compared to two-dimensional images, three-dimensional (3D) anatomical visualization with virtual reality (VR) and mixed reality (MR) offers a more comprehensive anatomical visualization and understanding in the perioperative phase. In a VR environment, the user is fully immersed in a simulated world. To create an immersive environment, each eye is provided with a separate image by the displays in the VR device. The user's physical movement is registered by cameras in the VR device and matched to the digital world. An MR device enhances the user's physical environment with a digital overlay, a so-called hologram. MR provides the opportunity to interact with the digital objects in the physical world through (depth) cameras and a motion sensor in the device that map out the user's surroundings and track their movements (2).

Both VR and MR techniques are increasingly adopted in neurosurgical preparation to provide a safe environment to plan surgical procedures, rehearse and foresee possible technical difficulties, and make the intraoperative phase more efficient (3).

Despite their substantial promise, a systematic analysis of the literature examining the role of MR and VR applications and their benefits as perioperative adjuncts in open IA surgery has been lacking. Authors present a comprehensive review on the topic, with the primary goal to study the true measurable benefits of using 3D visualization with MR and VR in preparation of IA surgery. This analysis thereby provides an overview of the technology used, its drawbacks and the potential future improvements.

Materials and methods

A systematic review was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (4). Two reviewers (EC and TK) screened records independently, and disagreements at any stage were resolved by discussion and consensus. Two additional records were identified through reference search. The critical appraisal of the included studies was performed by means of a risk of bias score using a modified version of the Cochrane Risk of Bias tool as shown in **Table 1** (5).

TABLE 1 Risk of bias score.

Author	A	B	C	D
Fellner et al. (6)	+	-	+	NA
Koyama et al. (7)	+	+	+	NA
Wong et al. (8)	+	+	+	NA
Bu et al. (9)	+	+	NA	NA
Mo et al. (8)	+	+	-	NA
Mori et al. (10)	+	+	-	-
Agarwal et al. (11)	+	+	+	NA
Nakabayashi et al. (12)	+	+	NA	NA
Bambakidis et al. (13)	+	+	+	NA
Di Somma et al. (14)	+	+	NA	NA
Cabrilo et al. (15)	+	+	NA	NA
Alaraj et al. (16)	+	-	+	NA
Kockro et al. (17)	+	+	+	-
Chugh et al. (18)	+	-	+	NA
Shono et al. (19)	+	+	NA	NA
Tucker et al. (20)	+	+	NA	NA
Eftekhar et al. (21)	+	-	-	NA
Gmeiner et al. (22)	+	+	+	NA
Toyooka et al. (23)	+	+	-	-
Neyazi et al. (24)	+	+	+	NA
Zawy Alsofy et al. (25)	+	+	+	NA
Haridas et al. (26)	+	+	NA	NA
Deib et al. (27)	+	-	NA	NA
Allgaier et al. (28)	+	+	+	NA
Li et al. (29)	+	+	NA	NA
Perin et al. (30)	+	+	+	-
Steineke et al. (31)	+	+	+	NA
Stifano et al. (32)	+	+	NA	NA

A: Appropriate eligibility criteria; B: Exposure/outcome measurement; C: Failure to adequately control confounding; D: Incomplete follow-up; NA, not applicable.

Search strategy

The PubMed and EMBASE databases were searched to identify eligible papers. The query was performed using the Boolean operators "AND" or "OR", and database-related filters to maximize the chance to identify articles focusing on 3D visualization through MR and VR system applied to IA surgery. The following string was entered:

((“neurosurg*"[Title/Abstract] OR “Neurosurgery"[MeSH Terms] OR “Neurosurgical Procedures"[MeSH Terms] OR “ventriculostom*"[Title/Abstract] OR “lobectom*"[Title/Abstract] OR “craniotom*"[Title/Abstract] OR “neuro surg*"[Title/Abstract] OR “neurologic surg*"[Title/Abstract]) AND (“augmented realit*"[Title/Abstract] OR “Augmented Reality"[MeSH Terms] OR “mixed realit*"[Title/Abstract] OR “virtual realit*"[Title/Abstract] OR “extended realit*"[Title/Abstract] OR “hologra*"[Title/Abstract] OR “Holography"[MeSH Terms] OR “head mounted display*"[Title/Abstract] OR “head up display*"[Title/Abstract] OR “head worn display*"[Title/Abstract] OR “Smart Glasses"[MeSH Terms])).

The most recent search was performed on November 28th 2022.

Selection criteria

Articles were included if the following criteria were met: (1) Studies published after 1997; (2) Studies analyzing specifically the

role of MR and VR in IA pre-surgical and intraoperative phases; (3) A specified 2D or 3D visualization technique as a mean to study angioarchitecture; (4) English, Italian, French or German language.

Data extraction

The following information was extracted from all included publications: (1) study group and year of publication; (2) type and number of IA included in the analysis; (3) imaging data source (computed tomography angiography (CTA), magnetic resonance angiography (MRA), digital subtraction angiography (DSA)); (4) category of visualization techniques (screen, glasses, head-mounted device (HMD); (5) inclusion of haptic feedback; (6) aim of the study (surgical planning/rehearsal, neurosurgical training, methodological validation); (7) study population (residents, fellows, attending neurosurgeons).

Statistical analysis

The descriptive statistical analyses were performed using R Studio. Data were presented as numbers and percentages.

Results

A PRISMA flowchart is displayed in Figure 1. A total of 1,763 publications were screened, 40 full-text articles were assessed for eligibility and 28 studies were included in this review. Studies were excluded when considered beyond the scope for the aims of the present analysis, and/or when their outcomes were not of interest. An overview of the included studies highlighting their major goals and advantages/disadvantages of augmented reality application as perceived by the authors of the publications is illustrated respectively in Table 2 and Table 3. In Table 3, where no data was specified, it means that the authors of the publication did not express it. Table 1 provides a visual summary of the quality review of the included studies.

Virtual reality

Virtual reality implies the use of a system which generates a complete immersion in a digital environment, that could provide a realistic simulation of the surgical approach (16). This type of technology was applied by 18 of the studies (64.3%)

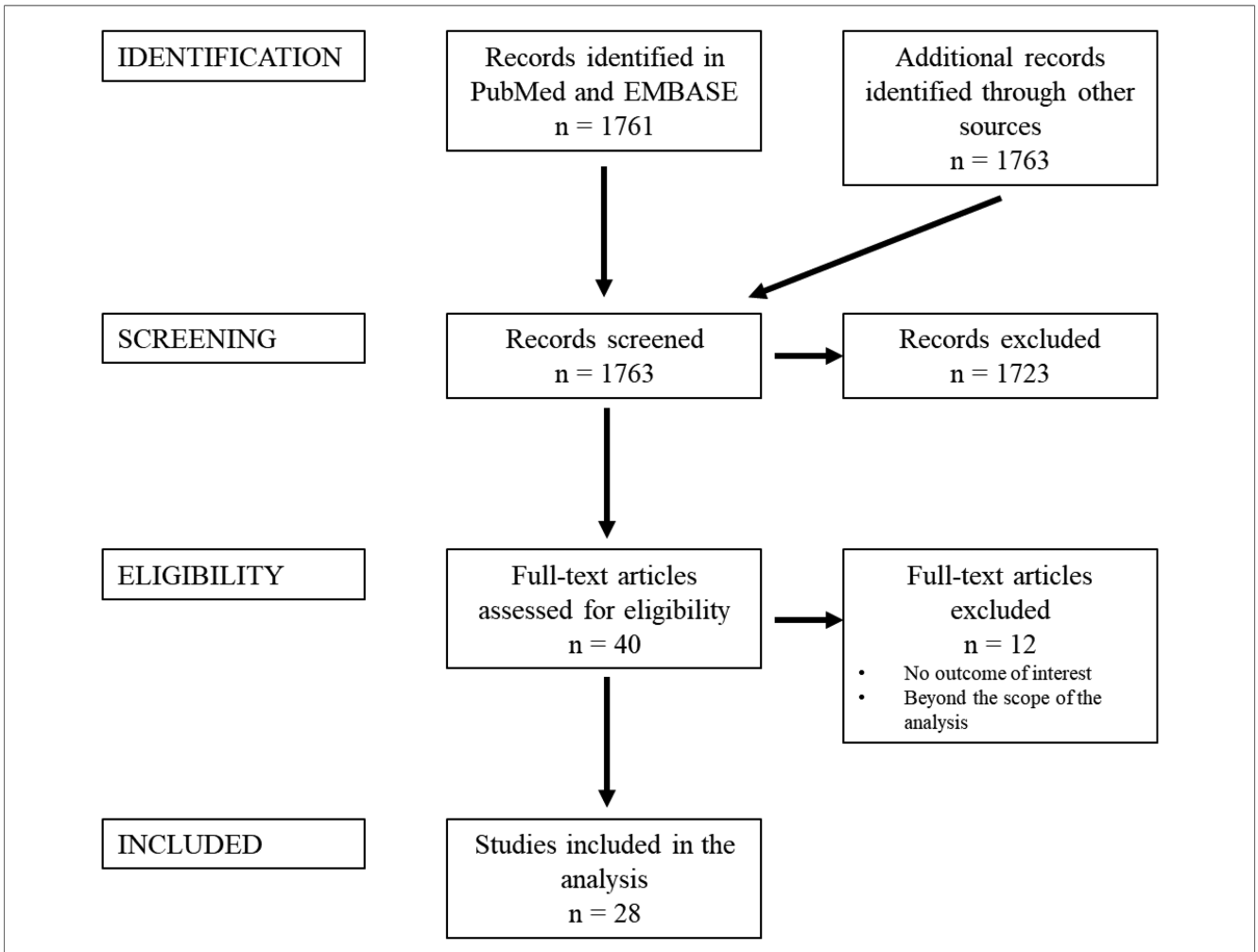


FIGURE 1
Summary of search strategy (PRISMA flow chart) for relevant studies.

TABLE 2 Studies' major goals, imaging elaboration techniques and display types, and source imaging.

Author	Year	AR	Major goals	Technology (+/-) processing tools	Display type	Source imaging
Fellner F et al.	1998	MR	Intraoperative visualization	Virtual cisternoscopy and Voxel View software	Screen	CT + MRI
Koyama T et al.	2000	MR	Surgical planning	New application program and Visual C++	Screen	NA
Wong et al.	2007	VR	• Surgical planning • Training	Dextroscope system (Bracco Diagnostics Inc.)	• Screen • Glasses/Head set	CT
Bu B et al.	2009	VR	Surgical planning	Dextroscope system (Bracco Diagnostics Inc.)	• Screen • Glasses/Head set	CT + MRI
Mo D et al.	2010	VR	Surgical planning	Dextroscope system (Bracco Diagnostics Inc.)	• Screen • Glasses/Head set	CT
Mori et al	2011	VR	Surgical planning	Mimics software	Screen	CT
Agarwal N et al	2012	VR	Training	Dextroscope system (Bracco Diagnostics Inc.)	• Screen • Glasses/Head set	CT + MRI
Nakabayashi H et al.	2012	MR	Surgical planning	AW VolumeShare, Stereo Movie Maker and QuickTime Virtual Reality	• Screen • Glasses/Head set	CT
Bambakidis NC et al.	2013	VR	Training	Selman Surgical Rehearsal Platform (Surgical Theater)	• Screen • Glasses/Head set	CT
Di Somma A et al.	2014	VR	Method validation	Dextroscope system (Bracco Diagnostics Inc.)	• Screen • Glasses/Head set	CT
Cabrilo I et al.	2014	MR	Intraoperative visualization	iPlan Workstation, Brainlab	Screen	CT + MRI + DSA
Alaraj A et al.	2015	VR	Training	Immersive Touch platform	• Screen • Glasses/Head set	CT
Kockro R et al.	2016	VR	Surgical planning	Dextroscope system (Bracco Diagnostics Inc.)	• Screen • Glasses/Head set	CT + MRI
Chung AJ et al.	2017	VR	Clipping	Selman Surgical Rehearsal Platform (surgical Theater)	• Screen • Glasses/Head set	CT
Shono N et al.	2017	MR	Surgical planning	Unity game engine, Avizo, Maya, and Leap Motion	• Screen • Glasses/Head set	CT + MRI + DSA
Tucker et al.	2017	VR	Training	Surgical Theatre	• Screen • Glasses/Head set	CT + MRI
Eftekhar B et al.	2017	MR	Intraoperative visualization	Sketchfab.com and Virtual Reality Modeling Language	Screen	DSA
Gmeiner M et al.	2018	VR	Training	New simulator and RISC Software—MEDVIS 3D	Screen	CT + DSA
Toyooka et al.	2018	MR	• Intraoperative visualization • Surgical planning	Head-up Display system and iPlan Workstation, Brainlab	Glasses/Head set	CT + MRI
Neyazi et al.	2019	VR	• Surgical planning • Training	Unity game engine and Virtual Reality Toolkit and MeVisLab	Glasses/Head set	MRI
Alsofy SZ et al.	2020	MR	Surgical planning	VR workstation connected to HTC Vive goggles and the SteamVR system and 3D Slicer	• Screen • Glasses/Head set	CT
Haridas A et al.	2020	VR	Surgical planning	Surgical Theatre	• Screen • Glasses/Head set	CT
Deib G et al.	2020	MR	Intraoperative visualization	Magic Leap One device	• Screen • Glasses/Head set	CT
Allgaier M et al.	2021	VR	Surgical planning	Unity game engine and XR Interaction Toolkit	Glasses/Head set	MRI
Li Z et al.	2021	VR	Surgical planning	Visualization Tool Kit and Unity3D platform	Glasses/Head set	CT
Perin A et al.	2021	VR	• Surgical planning • Training	Surgical Theatre	• Screen • Glasses/Head set	CT
Steineke TC et al.	2021	VR	Surgical planning	Surgical Theatre	• Screen • Glasses/Head set	CT + MRI
Stifano V et al.	2021	MR	Surgical planning	New MR application, Unity 3D, 3D Slicer and Blender	Screen Glasses/ Head set	CT

AR, augmented reality; MR, mixed reality; VR, virtual reality; CT, computer tomography; MRI, magnetic resonance imaging; DSA, digital subtraction angiography; NA, not applicable.

(8–11, 13, 14, 16–18, 20, 22, 24, 26, 28–31, 33), with a total of 321 aneurysms included in the studies. The VR systems that were mostly used were the Dextroscope system (Bracco Diagnostics Inc., Milan, Italy), documented by 6 studies (33%) (8, 9, 11, 14, 17, 33), and Surgical Theater (Surgical Theater Inc., Los Angeles, CA), utilized by 4 studies (22%) (20, 26, 30, 31). Thirteen studies (72%) chose the combination of screen and glasses/HMD as preferred visualization method (8, 9, 11, 13, 14, 16–18, 20, 26, 30, 31, 33). Exclusive use of a 2D visualization of the CT images represented the most relevant imaging source in 10 studies (56%) (8, 10, 13, 14, 16, 18, 26, 29, 30, 33).

TABLE 3 Advantages and disadvantages.

Author	Year	AR	Advantages	Disadvantages
Fellner F et al.	1998	MR	• Depth, perspective, lighting, color • Correct therapeutic decision	• Difficult visualization of perforators Intraoperative application • Operator-dependence • Time consuming
Koyama T et al.	2000	MR	Virtual manipulation	Imperfect reproduction of reality
Wong et al.	2007	VR	• Overview of the vasculature from any perspective • Haptic feedback • Training and education	• Small vessels running horizontally tend to be underestimated • Intraaneurysmal features not displayed • Time consuming
Bu B et al.	2009	VR	• Surgical rehearsal • Preoperative risk assessment	NA
Mo D et al.	2010	VR	Quick simulation	• Intraaneurysmal features not displayed
Mori et al.	2011	VR	More precise minimally invasive craniotomy planning	NA
Agarwal N et al.	2012	VR	Nonthreatening learning environment with immediate feedback	• Intraaneurysmal hemodynamics not displayed • No intrinsic information of the vessel's wall • No information about surrounding structures
Nakabayashi H et al.	2012	MR	• More effective realistic surgical simulation • Design of minimally invasive procedures	NA
Bambakidis NC et al.	2013	VR	Improved training experience, rehearsal and safety	No data on patients' outcomes
Di Somma A et al.	2014	VR	Improved anatomical understanding	NA
Cabrilo I et al.	2014	MR	• Optimization of patient positioning and operative trajectory • Better anatomical understanding • Supportive to intraoperative orientation	• Too small cohort to objectively evaluate the real impact of Mixed Reality
Alaraj A et al.	2015	VR	Improved anatomical understanding, intuitive training experience, haptic feedback	• Intraaneurysmal hemodynamics not displayed • No intrinsic information of the vessel's wall
Kockro R et al.	2016	VR	• Intraoperative deja-vu': enhancement of surgical confidence • Stereoscopic display and manipulation • Steep learning curve • Depth perception	Retrospective analysis lacking control groups
Chung AJ et al.	2017	VR	Statistically significant improvement in time per clip used	No patient outcome nor safety of surgical clipping
Shono N et al.	2017	MR	• Optimization of intraoperative trajectory and clip placement • Archive of cases that could be used for training • Incorporation of sense, touch and hearing	• Force feedback not incorporated • Feasibility not validated • Quality of the model dependent on quality of source imaging • Cumbersome workflow
Tucker et al.	2017	VR	• Better appreciation of the surgical anatomy • Nonthreatening environment for surgical simulation and training	The application of VR for neurosurgical training should be further and better implemented
Eftekhar B et al.	2017	MR	Improved anatomical orientation	Privacy concerns
Gmeiner M et al.	2018	VR	Improved anatomical understanding, realistic experience, improved training, satisfactory haptic feedback	Not realistic for calcified aneurysms, small perforators and wall irregularities like mini-blebs
Toyooka et al.	2018	MR	Understanding of anatomy, geometry and approach	Poorer image quality of the HUD
Neyazi et al.	2019	VR	Benefit on surgical trajectory and education	No patient-specific data
Alsofy SZ et al.	2020	MR	Benefit on aneurysm detection, anatomical understanding, surgical approach, and clipping planning	• No display of small branches and perforators, nor adhesions • Great dependence on the quality of input data
Haridas A et al.	2020	VR	• Detailed evaluation of the patient-specific anatomy prior to surgery • Better understanding of the complex anatomy in high resolution	NA
Deib G et al.	2020	MR	• Better preoperative anatomical understanding	NA
Allgaier M et al.	2021	VR	• Steep learning curve after adequate training • Improved anatomical understanding and planning of the surgical approach	Difficult first-time use. No real patients data
Li Z et al.	2021	VR	• Better anatomical understanding • Training environment • Enhanced surgical confidence	• Small number of experiments • Experimental equipment is relatively backward, resulting in inaccurate results
Perin A et al.	2021	VR	Improved anatomical understanding, realistic experience, improved training	• No haptic feedback: no simulation of dissection and tissue handling • Small study sample • Costs: limited diffusion
Steineke TC et al.	2021	VR	• Benefits on preoperative planning and rehearsal with decreased intraoperative times • Improved training experience and increased intraoperative efficiency	• Lack of an agreed on and validated complexity scoring system • Small sample size and not a varied population of surgeons tested
Stifano V et al.	2021	MR	• Better anatomical understanding • Optimization and customization of surgical planning • Valuable training tool	• Low comfort and maneuverability • Dependence of the model on the quality of the source imaging • Limited number of patients and users

AR, augmented reality; MR, mixed reality; VR, virtual reality; NA, not applicable.

- Preoperative planning:

None of the studies of this subgroup applied VR intraoperatively. 16 of the studies (89%) focused on the pre-operative planning of the surgical approach. The benefit of VR application for preoperative planning was qualitatively assessed using Likert scales and the Think Aloud Method, specifically for evaluation of anatomical understanding, depth perception and visualization of the surgical trajectory perceived by nine study groups (8, 14, 16, 17, 22, 24, 28, 30, 31).

- Benefit on training:

Eight studies (44%) assessed the impact of this technology on training neurosurgical residents, focusing on the benefits of VR with regard to realistic anatomical understanding, haptic feedback satisfaction and enhancement of surgical confidence (8, 11, 13, 16, 20, 24, 30). User satisfaction was assessed by means of Likert Scales and the Think-Aloud Method.

- Impact on patients:

Three studies (17%) evaluated the potential effect of 3D visualization in VR on clinical outcomes (10, 17, 30). None of the 18 studies aimed to evaluate the impact of surgical planning with VR on patient safety, and only one study (5.6%) aimed to assess the benefits of VR on patient education and understanding of the surgical procedure (8).

- Perceived disadvantages:

The lack of information on intra-aneurysmal hemodynamics and vessel wall characteristics was also reported as a disadvantage (8, 11, 16, 22). Furthermore, small vessels and perforating arteries tended to be underestimated or not displayed (8, 22).

Mixed reality

Conceptually, MR differs from VR in that it integrates a virtual environment with the real world, whereas the latter is a full immersion in a virtual environment. MR provides an interaction with digital objects in the real world. Ten of the collected studies applied this technology (6, 7, 12, 15, 19, 21, 23, 25, 27, 32), with a total of 183 analyzed IAs. In this subgroup, there was indeed no homogeneity among the MR systems used: each group utilized a center-specific system and different softwares for the segmentations and the post-processing of the images. 3D visualization occurred by means of a combination of screen and glasses/HMD in 6 out of 10 studies (60%) (7, 12, 19, 25, 27, 32). The remaining 4 studies (40%) utilized solely screen for image visualization (6, 7, 15, 21). CT as exclusive imaging input source was used by 3 studies (30%) (12, 25, 27), and 4 studies (40%) chose a multimodal imaging source (6, 15, 19, 23). Only one group (10%) used DSA only as the input source (21). None of the studies in this subgroup utilized MRI as the exclusive imaging source.

- Preoperative planning:

Four studies (40%) used this technology for both the surgical planning and intra-operative guidance (6, 15, 21, 27). The major goal documented in this subgroup was again planning of the best surgical approach, as documented by 5 studies (50%). Similarly to the VR-subgroup, the advantages perceived for the preoperative planning were based on an improved anatomical orientation, better depth perception and more adequate understanding of the surgical approach (6, 21, 25). The major outcomes of these studies were mostly evaluated through Likert Scales for a qualitative assessment. Only 2 studies (20%) performed a structured statistical analysis to examine the outcomes (23, 25).

- Benefit on training:

One of the studies in this subgroup aimed to assess the impact of MR visualization on neurosurgical training, testing the technology on residents neurosurgeons (32).

- Impact on patients:

None of the studies in this subgroup aimed to validate the impact of MR visualization on patient education/safety or clinical outcomes.

- Perceived disadvantages:

The most relevant drawback reported in the MR-subgroup was the difficult, if not impossible, visualization of small vessels and perforators, and the dependence of the segmentation on the quality of the input data (5/10 studies, 50%) (6, 7, 21, 23, 25).

Discussion

The present analysis represents an up-to-date systematic review of all published studies, which applied perioperative 3D visualization through MR and VR to IA microsurgery from 1997 to November 2022.

A relevant aspect emerging from the present analysis is the lack of measurable hardcore values to quantitatively examine the real added value of VR and MR applied to open IA surgery. While a qualitative assessment of the benefits of these 3D technologies is possible using Likert Scales and the Think Aloud Method, the absence of objective qualitative parameters makes the analysis partial and may hinder objective comparisons among the different 3D modalities, especially when a structured statistical analysis is not performed. Under this premise, this systematic review suggests an improved anatomical understanding, a better depth perception and a nonthreatening learning environment to be the most relevant perceived advantages of VR, MR applied for IA surgery planning, compared to conventional visualization strategies. The 3D and realistic replication of the cerebrovascular anatomy could help the acquisition of procedural motor skills, and enhance surgical orientation and confidence (34).

Most of the included studies used multimodal imaging input to create a more informative 3D vascular model, overcoming the disadvantages of exclusive use of one imaging modality. While CTA allows for a precise understanding of aneurysmal size and shape, provides detailed information on the parent vessel, and anatomical relationships with the skull base, combination with digital subtraction angiography (DSA) or MR-flow adds information on the flow patterns (22, 35). Nonetheless, none of the studies in the present cohort integrated hemodynamic information to the 3D visualization. Furthermore, the combination of CT and MR imaging provides important

information on vessel/aneurysm spatial relationships with the parenchyma and the cisternal system, which allow a better surgical orientation (11).

As far as visualization techniques are concerned, merging glasses or HMD's with 2D visualization of 3D vascular models enhances the perception of spatial position and surgical orientation (8). Glasses and HDMs may also allow a more intuitive and immersive interaction with the 3D models (36).

The use of VR. MR and RV does not come without limitations. The studies published so far are mostly retrospective, with small sample sizes and no control groups. The analyzed studies examine almost exclusively aneurysms treated in elective settings, with specific focus on anterior circulation IAs, and rarely provide information on patient functional outcomes. The lack of an objective strategy to qualitatively assess the benefits of these technologies represents a major bias as well. While Likert-scales or Think-Aloud Method are mostly applied to evaluate the intuitiveness and the satisfaction of the users, no standardized, agreed on quantitative scores have been provided yet. To obviate this absence, objective parameters such as size of the aneurysmal dome, width of the neck, orientation of the dome, distance of the aneurysm from relevant anatomical structures should be noted, when using VR and/or MR, validated and combined into quantitative scores. The difficulty of these analyses may lie in the paucity of data and in the novelty of these technologies, which are still not available in every center. Their diffusion may also be limited by their often not affordable costs. Another relevant aspect resulting from the paucity and diversity of the available data is the lack of unified criteria to provide an objective appraisal of the current literature. The advantages and the disadvantages reported for each paper come mostly from the appraisal and experience of the original authors. With further implementation of these technologies and gathering of more extensive and unified data, this limitation could be obviated. Furthermore, segmentation of intracranial vessels and fine anatomical structures is still highly dependent on the quality of input data, which makes the integration of hemodynamic information, small vessels or intramural particularities difficult. The integration of hemodynamic information into a 3D preoperative study of IAs may help characterize their angioarchitecture more accurately. This information may provide major advantages on the tailoring of their treatment in a pathology-specific way. Such a tailoring could potentially increase intraoperative safety and therapeutic efficiency.

Conclusion

This analysis endorses the promising role of MR and VR to provide a more accurate aneurysm-specific anatomical visualization and understanding. The absence of a standardized set of quantitative parameters to provide an objective assessment of the real benefit of these technologies on training in IA surgery should be a major drive for future studies on the topic.

Furthermore, integration of hemodynamic analysis to the 3D visualization may also be a promising avenue for future research.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

EC conceptualized the study with TD; EC wrote the manuscript text, prepared all tables and the figure. BL and TK helped with the systematic review of the literature and the amendment of the manuscript. TD supported the gathering and the processing of the data. All authors contributed to the article and approved the submitted version.

Funding

This work is part of "SURGENT" under the auspices of University Medicine Zurich/Hochschulmedizin Zürich.

Conflict of interest

TD is Co-founder and CMO of Augmedit bv, an augmented reality company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Brown RD, Broderick JP. Unruptured intracranial aneurysms: epidemiology, natural history, management options, and familial screening. *Lancet Neurol.* (2014) 13(4):393–404. doi: 10.1016/S1474-4422(14)70015-8

2. Andrews C, Southworth MK, Silva JNA, Silva JR. Extended reality in medical practice. *Curr Treat Options Cardiovasc Med.* (2019) 21(4):18. doi: 10.1007/s11936-019-0722-7

3. Zhang C, Gao H, Liu Z, Huang H. The potential value of mixed reality in neurosurgery. *J Craniofac Surg*. (2021) 32(3):940–3. doi: 10.1097/SCS.0000000000007317

4. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Br Med J*. (2009) 339:b2535. doi: 10.1136/bmj.b2535

5. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence–study limitations (risk of bias). *J Clin Epidemiol*. (2011) 64(4):407–15. doi: 10.1016/j.jclinepi.2010.07.017

6. Fellner F, Blank M, Fellner C, Böhm-Jurkovic H, Bautz W, Kalender WA. Virtual cisternoscopy of intracranial vessels: a novel visualization technique using virtual reality. *Magn Reson Imaging*. (1998) 16(9):1013–22. doi: 10.1016/s0730-725x(98)00113-1

7. Koyama T, Hongo K, Tanaka Y, Kobayashi S. Simulation of the surgical manipulation involved in clipping a basilar artery aneurysm: concepts of virtual clipping. Technical note. *J Neurosurg*. (2000) 93(2):355–60. doi: 10.3171/jns.2000.93.2.0355

8. Wong GKC, Zhu CXL, Ahuja AT, Poon WS. Craniotomy and clipping of intracranial aneurysm in a stereoscopic virtual reality environment. *Neurosurgery*. (2007) 61(3):564–8; discussion 568–569. doi: 10.1227/01.NEU.0000290904.46061.0D

9. Bu B, Zhou D-B, Xu B-N, Yu X-G, Zhang Y-Z, Wei S-B. Application of dextroscope operation planning workstation in intracranial lesion operations: preoperative images reconstruction, simulation and dissection. *J Clin Rehab Tissue Eng Res*. (2009) 13(4):789–92. doi: 10.1093/eurheartj/ehab777

10. Mori K, Esaki T, Yamamoto T, Nakao Y. Individualized pterional keyhole clipping surgery based on a preoperative three-dimensional virtual osteotomy technique for unruptured middle cerebral artery aneurysm. *Minim Invasive Neurosurg*. (2011) 54(5–6):207–13. doi: 10.1055/s-0031-1286335

11. Agarwal N, Schmitt PJ, Sukul V, Prestigiacomo CJ. Surgical approaches to complex vascular lesions: the use of virtual reality and stereoscopic analysis as a tool for resident and student education. *BMJ Case Rep*. (2012) 2012:bcr0220125859. doi: 10.1136/bcr.02.2012.5859

12. Nakabayashi H, Shimizu K. Stereoscopic virtual realistic surgical simulation in intracranial aneurysms. *Neurol India*. (2012) 60(2):191–7. doi: 10.4103/0028-3886.96399

13. Bambakidis NC, Selman WR, Sloan AE. Surgical rehearsal platform: potential uses in microsurgery. *Neurosurgery*. (2013) 73(Suppl 1):122–6. doi: 10.1227/NEU.0000000000000099

14. Somma AD, Bronzoni C, Guadagno E, Solari D, Dell'aversana GO, De Caro BS, et al. Extended endoscopic endonasal approaches for cerebral aneurysms: anatomical, virtual reality and morphometric study. *Biomed Res Int*. (2014) 2014:703792. doi: 10.1155/2014/703792

15. Cabrilo I, Bijlenga P, Schaller K. Augmented reality in the surgery of cerebral aneurysms: a technical report. *Neurosurgery*. (2014) 10(Suppl 2):252–60; discussion 260–261. doi: 10.1227/NEU.0000000000000328

16. Alaraj A, Luciano CJ, Bailey DP, Elsenousi A, Roitberg BZ, Bernardo A, et al. Virtual reality cerebral aneurysm clipping simulation with real-time haptic feedback. *Neurosurgery*. (2015) 11(Suppl 2):52–8. doi: 10.1227/NEU.0000000000000583

17. Kockro RA, Killeen T, Ayyad A, Glaser M, Stadie A, Reisch R, et al. Aneurysm surgery with preoperative three-dimensional planning in a virtual reality environment: technique and outcome analysis. *World Neurosurg*. (2016) 96:489–99. doi: 10.1016/j.wneu.2016.08.124

18. Chugh AJ, Pace JR, Singer J, Tatsuoka C, Hoffer A, Selman WR, et al. Use of a surgical rehearsal platform and improvement in aneurysm clipping measures: results of a prospective, randomized trial. *J Neurosurg*. (2017) 126(3):838–44. doi: 10.3171/2016.1.JNS152576

19. Shono N, Kin T, Nomura S, Miyawaki S, Saito T, Imai H, et al. Microsurgery simulator of cerebral aneurysm clipping with interactive cerebral deformation featuring a virtual arachnoid. *Oper Neurosurg (Hagerstown)*. (2018) 14(5):579–89. doi: 10.1093/ons/opx155

20. Tucker AM, Beckett JS, Martin NA. Next generation case report: supraorbital craniotomy for anterior communicating artery aneurysm clipping in annotated virtual reality environment. *Oper Neurosurg (Hagerstown)*. (2018) 15(5):E73–6. doi: 10.1093/ons/opy039

21. Eftekhar B. Smartphone as a remote touchpad to facilitate visualization of 3D cerebral angiograms during aneurysm surgery. *J Neurol Surg A Cent Eur Neurosurg*. (2017) 78(5):502–6. doi: 10.1055/s-0037-1598049

22. Gmeiner M, Dirnberger J, Fenz W, Gollwitzer M, Wurm G, Trenkler J, et al. Virtual cerebral aneurysm clipping with real-time haptic force feedback in neurosurgical education. *World Neurosurg*. (2018) 112:e313–23. doi: 10.1016/j.wneu.2018.01.042

23. Toyooka T, Otani N, Wada K, Tomiyama A, Takeuchi S, Fujii K, et al. Head-up display may facilitate safe keyhole surgery for cerebral aneurysm clipping. *J Neurosurg*. (2017) 129(4):883–9. doi: 10.3171/2017.5.JNS162692

24. Neyazi B, Saalfeld P, Berg P, Skalej M, Preim B, Erol İ, et al. *VR Craniotomy for optimal intracranial aneurysm surgery planning*. Reutlingen: Conference: Computer and Robotic Assisted Surgery (CURAC) (2019). p. 234–9.

25. Zawy Alsofy S, Sakellaropoulou I, Nakamura M, Ewelt C, Salma A, Lewitz M, et al. Impact of virtual reality in arterial anatomy detection and surgical planning in patients with unruptured anterior communicating artery aneurysms. *Brain Sci*. (2020) 10(12):E963. doi: 10.3390/brainsci10120963

26. Haridas A, Miller M. Middle cerebral artery aneurysm clipping with immersive 360° virtual reality model: 2-dimensional operative video. *Oper Neurosurg (Hagerstown)*. (2021) 20(4):E314. doi: 10.1093/ons/opaa416

27. Deib G, Smith D, Chaudhary B, Boo S, Tarabishy AJ, Carpenter J, et al. E-195 A mixed reality spatial computing framework for preprocedural evaluation of cerebral aneurysms: approach and preliminary results. *J Neurointervent Surg*. (2020) 12(1):A134–5. doi: 10.1136/neurintsurg-2020-SNIS.226

28. Allgaier M, Amini A, Neyazi B, Sandalcioglu IE, Preim B, Saalfeld S. VR-based training of craniotomy for intracranial aneurysm surgery. *Int J Comput Assist Radiol Surg*. (2022) 17(3):449–56. doi: 10.1007/s11548-021-02538-3

29. Li Z, Huo G, Feng Y, Ma Z. Application of virtual reality based on 3D-CTA in intracranial aneurysm surgery. *J Healthc Eng*. (2021) 2021:9913949. doi: 10.1155/2021/9913949

30. Perin A, Gambatesa E, Galbiati TF, Fanizzi C, Carone G, Rui CB, et al. The "STARS-CASCADE" study: virtual reality simulation as a new training approach in vascular neurosurgery. *World Neurosurg*. (2021) 154:e130–46. doi: 10.1016/j.wneu.2021.06.145

31. Steineke TC, Barbery D. Microsurgical clipping of middle cerebral artery aneurysms: preoperative planning using virtual reality to reduce procedure time. *Neurosurg Focus*. (2021) 51(2):E12. doi: 10.3171/2021.5.FOCUS21238

32. Stifano V, Palumbo MC, Chidambaram S, Sturiale CL, Albanese A, Marchese E, et al. The use of mixed reality for the treatment planning of unruptured intracranial aneurysms. *J Neurosurg Sci*. (2023) 67(4):491–7. doi: 10.23736/S0390-5616.21.05356-X

33. Mo D, Bao S, Li L, Yi Z, Zhang J, Zhang Y. Virtual reality system for diagnosis and therapeutic planning of cerebral aneurysms. *Chin Med J (Engl)*. (2010) 123(16):2206–10. doi: 10.1007/s00701-016-2892-3

34. Wurm G, Lehner M, Tomancok B, Kleiser R, Nussbaumer K. Cerebrovascular biomodeling for aneurysm surgery: simulation-based training by means of rapid prototyping technologies. *Surg Innov*. (2011) 18(3):294–306. doi: 10.1177/1553350610395031

35. Andereggen L, Gralla J, Andres RH, Weber S, Schroth G, Beck J, et al. Stereolithographic models in the interdisciplinary planning of treatment for complex intracranial aneurysms. *Acta Neurochir (Wien)*. (2016) 158(9):1711–20. doi: 10.1007/s00701-016-2892-3

36. Stadie AT, Kockro RA, Reisch R, Tropine A, Boor S, Stoeter P, et al. Virtual reality system for planning minimally invasive neurosurgery. Technical note. *J Neurosurg*. (2008) 108(2):382–94. doi: 10.3171/JNS/2008/108/2/0382

frontiers | Frontiers in Surgery

TYPE Original Research
PUBLISHED 23 October 2023
DOI 10.3389/fsurg.2023.1259756

OPEN ACCESS

EDITED BY
Mohammed Ali Alvi,
University Health Network (UHN), Canada

REVIEWED BY
Satoshi Obata,
Fukuoka University School of Medicine and Hospital, Japan
Leonardo Tariciotti,
University of Milan, Italy

*CORRESPONDENCE
Vivek P. Buch
✉ vpbuch@stanford.edu

†These authors have contributed equally to this work

RECEIVED 17 July 2023
ACCEPTED 20 September 2023
PUBLISHED 23 October 2023

CITATION
Park JJ, Doiphode N, Zhang X, Pan L, Blue R, Shi J and Buch VP (2023) Developing the surgeon-machine interface: using a novel instance-segmentation framework for intraoperative landmark labelling.
Front. Surg. 10:1259756.
doi: 10.3389/fsurg.2023.1259756

Developing the surgeon-machine interface: using a novel instance-segmentation framework for intraoperative landmark labelling

Jay J. Park[1,2†], Nehal Doiphode[1,3†], Xiao Zhang[4], Lishuo Pan[5], Rachel Blue[6], Jianbo Shi[3] and Vivek P. Buch[1*]

[1]Department of Neurosurgery, The Surgical Innovation and Machine Interfacing (SIMI) Lab, Stanford University School of Medicine, Stanford, CA, United States, [2]Centre for Global Health, Usher Institute, Edinburgh Medical School, The University of Edinburgh, Edinburgh, United Kingdom, [3]Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, United States, [4]Department of Computer Science, University of Chicago, Chicago, IL, United States, [5]Department of Computer Science, Brown University, Providence, RI, United States, [6]Department of Neurosurgery, Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA, United States

Introduction: The utilisation of artificial intelligence (AI) augments intraoperative safety, surgical training, and patient outcomes. We introduce the term Surgeon-Machine Interface (SMI) to describe this innovative intersection between surgeons and machine inference. A custom deep computer vision (CV) architecture within a sparse labelling paradigm was developed, specifically tailored to conceptualise the SMI. This platform demonstrates the ability to perform instance segmentation on anatomical landmarks and tools from a single open spinal dural arteriovenous fistula (dAVF) surgery video dataset.

Methods: Our custom deep convolutional neural network was based on SOLOv2 architecture for precise, instance-level segmentation of surgical video data. Test video consisted of 8520 frames, with sparse labelling of only 133 frames annotated for training. Accuracy and inference time, assessed using F1-score and mean Average Precision (mAP), were compared against current state-of-the-art architectures on a separate test set of 85 additionally annotated frames.

Results: Our SMI demonstrated superior accuracy and computing speed compared to these frameworks. The F1-score and mAP achieved by our platform were 17% and 15.2% respectively, surpassing MaskRCNN (15.2%, 13.9%), YOLOv3 (5.4%, 11.9%), and SOLOv2 (3.1%, 10.4%). Considering detections that exceeded the Intersection over Union threshold of 50%, our platform achieved an impressive F1-score of 44.2% and mAP of 46.3%, outperforming MaskRCNN (41.3%, 43.5%), YOLOv3 (15%, 34.1%), and SOLOv2 (9%, 32.3%). Our platform demonstrated the fastest inference time (88ms), compared to MaskRCNN (90ms), SOLOV2 (100ms), and YOLOv3 (106ms). Finally, the minimal amount of training set demonstrated a good generalisation performance –our architecture successfully identified objects in a frame that were not included in the training or validation frames, indicating its ability to handle out-of-domain scenarios.

Discussion: We present our development of an innovative intraoperative SMI to demonstrate the future promise of advanced CV in the surgical domain. Through successful implementation in a microscopic dAVF surgery, our framework demonstrates superior performance over current state-of-the-art segmentation architectures in intraoperative landmark guidance with high sample efficiency, representing the most advanced AI-enabled surgical inference platform to date. Our future goals include transfer learning paradigms for

scaling to additional surgery types, addressing clinical and technical limitations for performing real-time decoding, and ultimate enablement of a real-time neurosurgical guidance platform.

KEYWORDS

artificial intelligence, intraoperative guidance, machine learning, surgical guidance, spine, arteriovenous fistula, surgeon-machine interface, global neurosurgery

1. Background

Intraoperative application of Artificial Intelligence (AI) is a rapidly advancing area in surgical innovation. AI technology offers various capabilities within the operating room, such as automating workflows and aiding in intraoperative decision-making (1, 2). The ultimate objective is to leverage AI's potential to learn, interpret, predict, and solve problems by training Machine Learning (ML) algorithms. These algorithms can process vast amounts of real-world data and guide decisions comparable to those of expert surgeons (3). We have introduced the term Surgeon-Machine Interface (SMI) to describe the advanced and innovative fusion of surgeons and machine interfaces, creating a new realm of collaboration. Computer Vision (CV) plays a pivotal role in facilitating interaction with intraoperative data, enabling machines to comprehend surgical images and videos (4). It also serves as the foundation of current endeavours in intraoperative landmark guidance. However, the availability of literature and regulatory-approved devices for real-time AI-based anatomical landmark labelling is limited, indicating that this technology is still in its early stages (5, 6). Nevertheless, recent advancements in artificial neural networks (ANNs), a subfield of ML and the backbone of deep learning (DL), show promise in enabling AI to achieve even higher levels of performance in this field (5, 6).

In a surgical setting, there are two crucial CV tasks: recognition and tracking. Object recognition employs machine learning (ML) to identify objects within an image, similar to human perception. When combined with an object localization algorithm, object detection can be achieved. This algorithm generates a bounding box that encompasses the object and provides a label for it. However, in surgical applications, where anatomical structures have intricate contours and unclear boundaries, a single bounding box may not accurately capture the desired area (7). Object segmentation addresses this limitation by producing pixel-wise masks that offer a detailed labelling of individual objects within an image (8). There are two types of segmentation: semantic segmentation groups similar pixels into a single classification, while instance segmentation distinguishes and segments each individual entity. In essence, instance segmentation allows for the pixelwise classification of individual objects within a surgical field, whether they are anatomical structures or surgical instruments. Although, it is the most preferred recognition technique for intraoperative guidance (9), previous attempts have been limited to segmenting rigid surgical instruments (10), as opposed to anatomical structures, which are often characterised by semi-rigid boundaries and thus pose a more difficult segmentation task.

Although few, there have been some platforms and clinical evidence in general surgery that attests to the accuracy and the significance of AI software in an intraoperative setting (7, 9, 11–14). Neurosurgery, at the forefront of cutting-edge technology, has witnessed numerous advancements in AI applications; however, these applications are limited to surgical phase recognition (15), detection and surveillance (16), diagnosis (17, 18), endovascular navigation (16), training and preoperative planning (2, 16, 19–21), intraoperative imaging, and workflow automation (22). To our knowledge, there is no other literature or technological reports that demonstrate a scalable surgical video analysis system in neurosurgery. In this study, we aim to demonstrate the most advanced surgical CV architecture to-date, and for the first time applied to a neurosurgical context. Although a fully functioning SMI will incorporate real-time implementation with a user interface, in this manuscript we seek to introduce the core technology for the conceptualisation of our future real-time enabled SMI. We demonstrate our custom instance segmentation core architecture and prediction model in a proof-of-concept for open spinal dural arteriovenous fistula (dAVF) surgery.

2. Methodology

2.1. Developing the AI framework

Our first prototype framework consisted of three parallel components (**Figure 1**). (1) A single frame would be processed through Mask Region-based Convolutional Neural Network (MaskRCNN) training, producing class-agnostic boxes. We train MaskRCNN in a supervised manner on a few annotated frames. (2) We augment the prediction capability of MaskRCNN by including temporal information from Colourisation, an unsupervised technique to learn the frame-wise feature correlation. Specifically, we extract the unsupervised feature flow from Colourisation. Then we use it to propagate the instance mask from the previous frame to the consecutive frames serving as complementary predictions. (3) Propagation *via* feature flow also yields pseudo ground truth instance segmentation to further improve our Mask R-CNN model. To filter out the noisy label, we build a rolling updated memory bank to collect high-quality predictions and score the incoming pseudo prediction. Only examples above the threshold will be used to fine-tune the Mask R-CNN model. If few annotated labels are present, only process (1) and (2) will take place. This architecture rendered inadequate results with anatomical landmarks being poorly demarcated as shown in **Figure 2**. The major issue was that the embedding flow

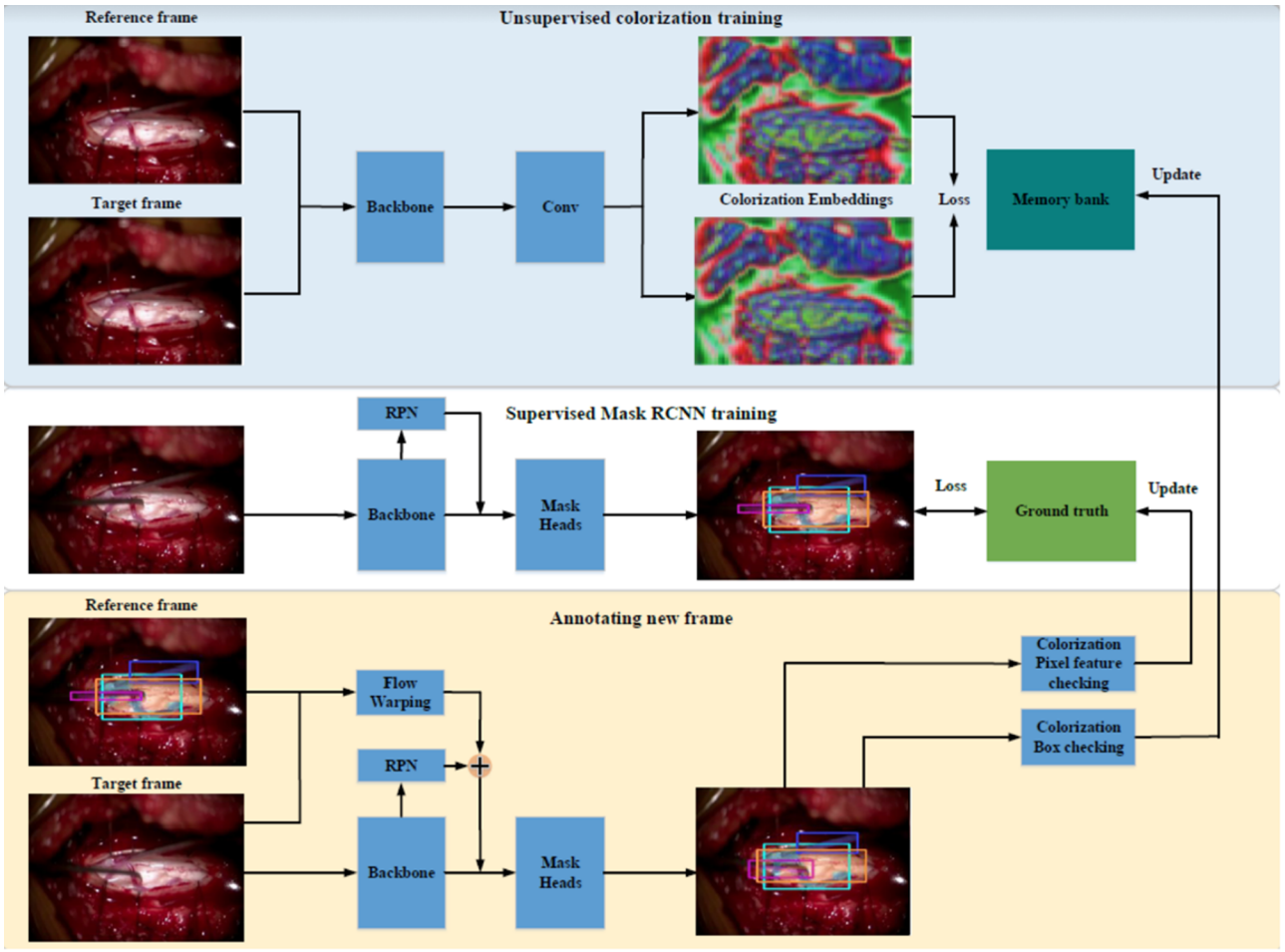


FIGURE 1
The initial mask-RCNN architecture.

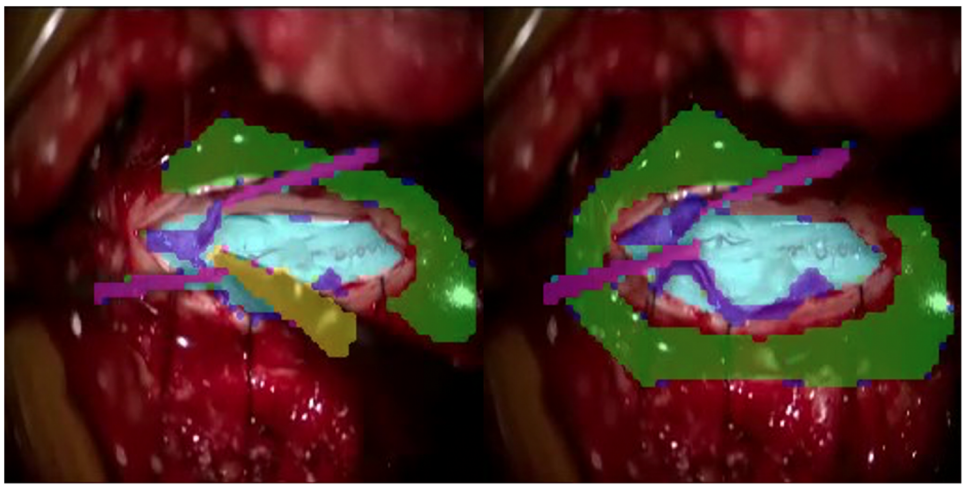


FIGURE 2
Poorly demarcated annotation using colourisation technique.

from colourisation is biassed to the low-level texture. Tracking without mid and high-level features is fragile in the challenging case.

Our next attempt was to utilise SOLOv1 architecture with Kanade-Lucas-Tomasi (KLT) tracking system, as Figure 3. In SOLOv1, the image is divided into a grid and objects are

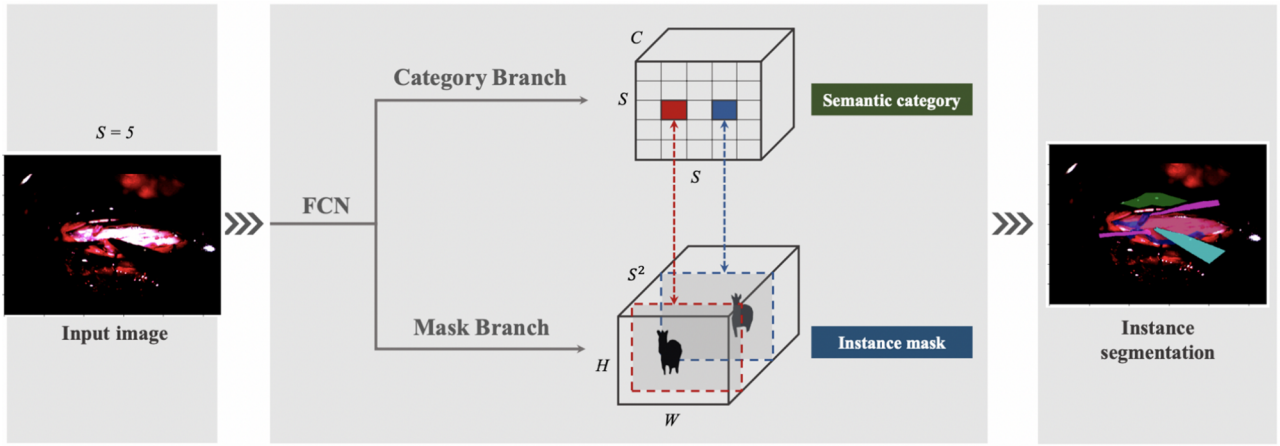


FIGURE 3
Initial SOLOv1 architecture.

located for every cell in the grid. Instance category and instance mask are computed in parallel across the grid. Since frame-by-frame instance segmentation has a disadvantage of lacking temporal information we embedded a KLT tracking system into the segmentation system to recognize the objects moving across video. The tracking system generates initial feature points over the masks in the first frame. KLT feature tracker tracks these feature points crossing the video. The identity of a segmentation is determined by a majority vote over all the feature points.

To account for the challenges faced with our previous instance segmentation frameworks, we developed a novel segmentation algorithm (Figure 4) based on the state-of-the-art SOLOv2 architecture, a dynamic and fast framework for real-time object detection. In SOLOv2, the mask head is further decomposed into 2 branches, namely, feature branch and dynamic convolution kernel branch. Instead of forwarding the feature map directly through another layer of convolution, the feature map is used to learn a dynamic convolution head, which is a kernel map that convolves with the feature map to output the final mask

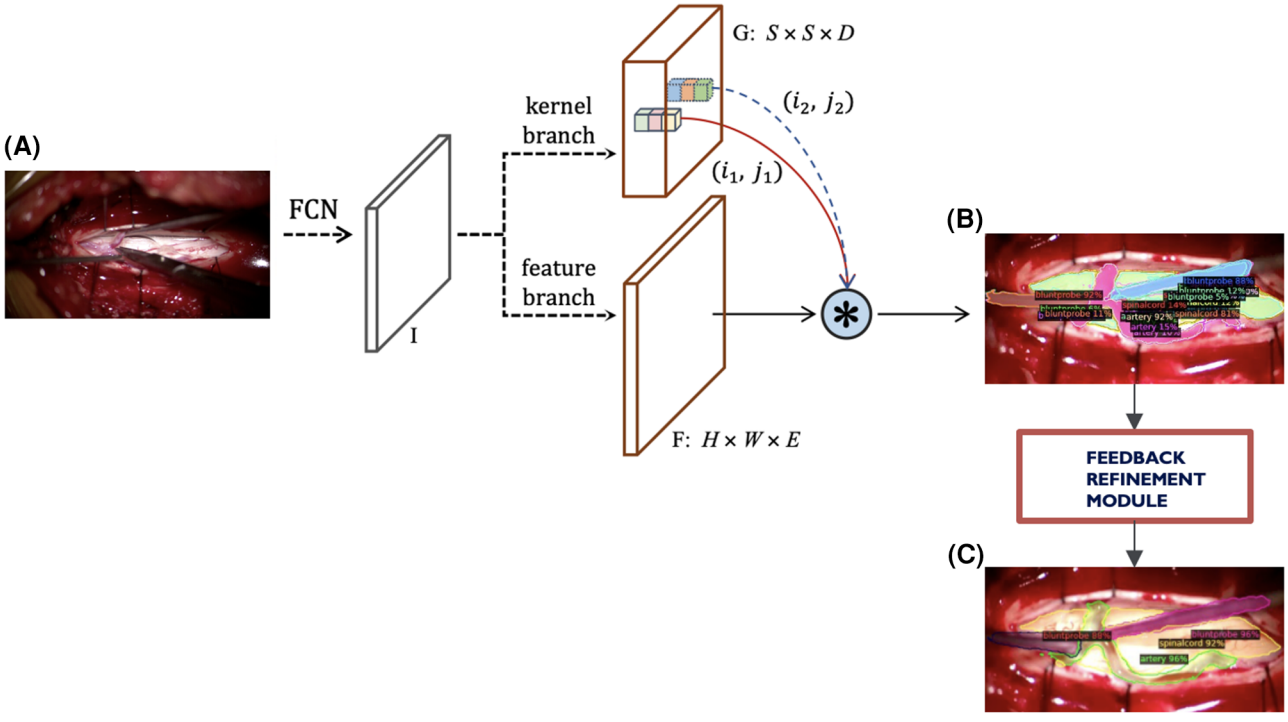


FIGURE 4
Current modified SOLOv2 architecture. (A) Input video frame, (B) Model output predictions, (C) Refined predictions.

prediction. However, in our experiments we face a practical limitation of the SOLOv2 framework and obtain many false positive detections. To alleviate this issue, we customise the framework to refine this low confidence but plausible detections into high confidence and obtain higher performance in instance segmentation (Figure 4).

All algorithms were constructed on Ubuntu 20.04.4 LTS (x86_64; Canonical Ltd., London, United Kingdom) and developed using the Detectron2, a Pytorch object detection library in Python (Language).

2.2. Dataset preparation

A single test video of spinal dAVF surgery was recorded with Zeiss Opmi Pentero 800 (Carl Zeiss AG, Jena, Germany) at The Hospital of the University of Pennsylvania. This video constitutes 8,520 frames. A trainee neurosurgeon (RB) annotated 133 frames using Computer Vision Annotation Tool (CVAT 7.4.0; Irvine, California, United States). 133 frames were sampled every 30 frames within the first 1.5 min of the surgical video (in-domain). These frames consist of the operator dissecting the arachnoid and separating the two abnormal dorsal spinal arteries (part of the dural AVF) with a blunt probe, dissector, and micro-scissors followed by the temporary clipping a dorsal spinal artery with an aneurysm clip after. Any ambiguities with anatomical structure were clarified by an attending neurosurgeon (VPB).

2.3. Statistical analysis

To assess segmentation accuracy and computing speed, we applied different frameworks to the single surgical video and validated it across 85 frames. These frames were uniformly sampled every 3.33 s from entire 4.73 min of the video. The validation set comprised of 27 frames which were in-domain and 58 of these frames that were out-of-domain frames, which was a hold-out test set as a temporal partition from the same video. The in-domain frames constitute unlabelled frames within the period of the training set as opposed to the out-of-domain frames that were unlabelled frames after the last time period of the training set. In these out-of-domain frames, the surgeon irrigated the surgical field, suctioned pools of blood outside the dura, manipulated the arteries with a suction and blunt probe, and finally removed the aneurysm clip with the applier. Instance segmentation was performed by MaskRCNN + Feature Pyramid Networks (FPN), SOLOv2 + R101DCN, SparseRCNN, YOLOv3, and our SOLOv2-based modified architecture. We then calculated certain metrics to measure precision for each of these architectures: mean Average Precision (mAP), and F1 score. Accuracy was tested on in-domain frames, which were part of the validation set, as well as out-of-domain frames which were not part of the test or validation set. Furthermore, we evaluated the inference time for the computing speed of each of the networks.

All these frameworks were processed in a single computer on Ubuntu 20.04.4 LTS (x86_64; Canonical Ltd., London, United Kingdom) with GeForce RTX 2080 Ti (NVIDIA, Santa Clara, United States) mounted. Statistical analysis was all performed using the Detectron2 (software).

3. Results

Our SMI framework outperformed any other known frameworks reported in literature for intraoperative landmark guidance in terms of accuracy and computing speed as shown in Table 1. F1-score and mAP of our model was 17% and 15.2% respectively, in comparison to the original SOLOv2 architecture which was 3.1% and 10.4%, YOLOv3 with 5.4% and 11.9%, and MaskRCNN with 15.2% and 13.9%, respectively. Taking into consideration detections that surpassed the Intersection over Union (IoU) threshold of 50%, our SMI had F1-score of 44.2% and mAP of 46.3%. This was followed by the MaskRCNN architecture with F1-score of 41.3% and mAP of 43.5%, YOLOv3 with 15% and 34.1%, and finally SOLOv2 with 9% and 32.3% respectively.

Qualitatively, our SMI architecture was successful in identifying objects in both in-domain frames (Figures 5A,B) and out-of-domain frames (Figures 6A,B), which indicates a good generalisation based on training on a quarter of the full surgical video. Our in-domain frame predictions demonstrated a high mAP score of >0.50 for anatomical structures and surgical tools and achieved >0.30 for distractions such as pool of blood. Most importantly, it was able to identify the blunt probe which was not annotated in the ground truth frame and was also able to predict a segment of the artery that was not annotated (Figure 5B). Identification of objects in out-of-domain frames also demonstrated visually promising results and high accuracy (Figure 6B). To highlight, the model was able to identify the suction as a different instrument in the out-of-domain frames, even though it has never seen the object before within the training set (Figure 6B).

The computation speed of our SMI was the fastest amongst all other frameworks (Table 1) with inference time of 88 ms, followed by MaskRCNN (90 ms), SOLOv2 (100 ms), and YOLOv3 (106 ms).

4. Discussion

This study introduces an advanced surgical platform that utilises instance segmentation for intraoperative guidance. The obtained results demonstrate the feasibility of our SMI

TABLE 1 Accuracy and computing speed of our SMI architecture and other commonly reported frameworks in literature.

	F1	F1 (iou 50)	mAP	mAP50	Inference time (ms)
SOLOv2	0.031	0.09	10.4	32.3	100
YOLOv3	0.054	0.15	11.9	34.1	106
MaskRCNN	0.152	0.413	13.9	43.5	90
Our SMI	**0.17**	**0.442**	**15.2**	**46.3**	**88**

Bold values are results from novel architecture.

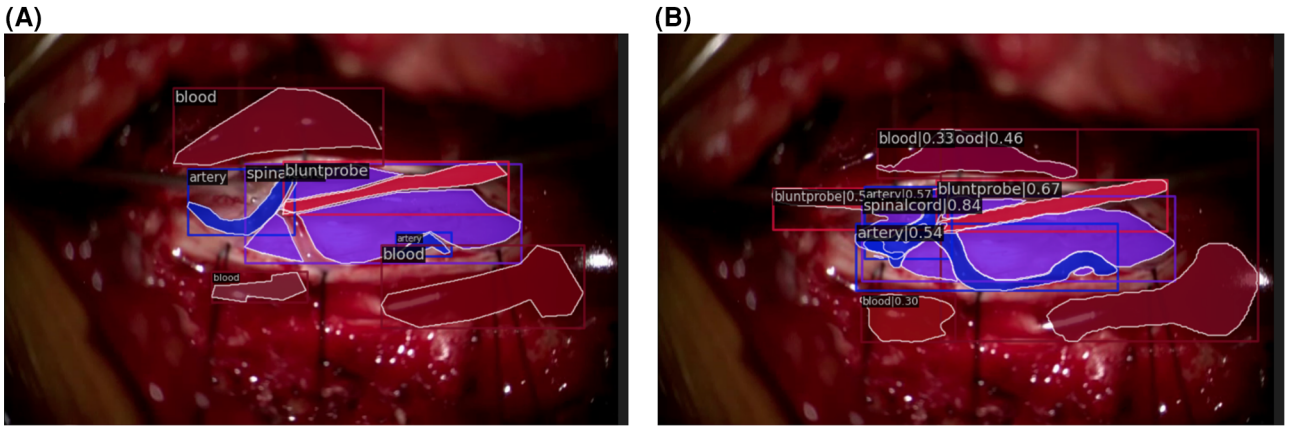


FIGURE 5
In-domain segmentation frames. (A) Ground truth frame of in-domain segmentation. (B) Prediction frame of in-domain segmentation.

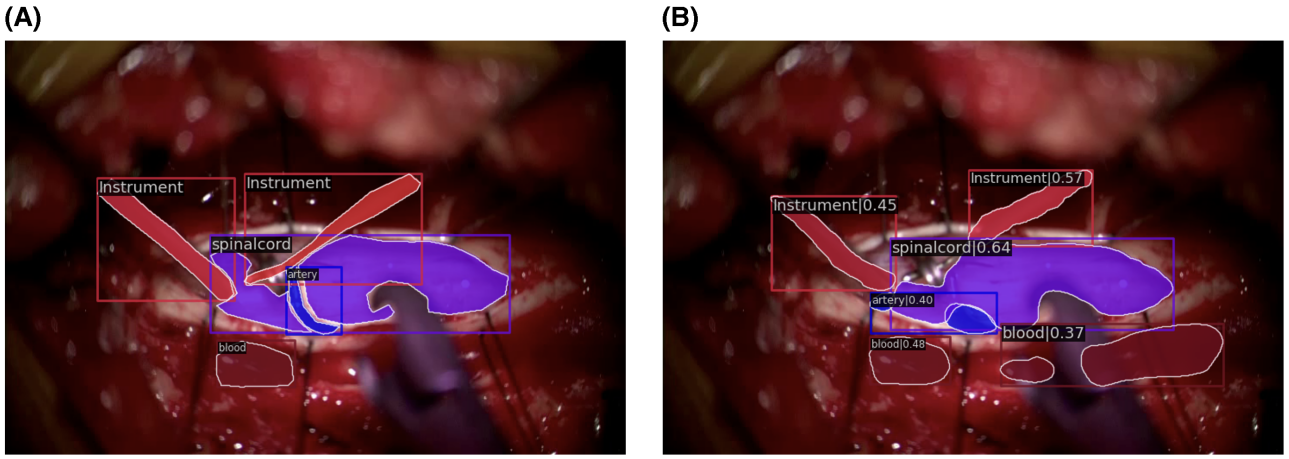


FIGURE 6
Out-of-domain segmentation frames. (A) Ground truth frame of out-of-domain segmentation. (B) Prediction frame of out-of-domain segmentation.

framework for real-time application in spinal dAVF surgeries and its potential for adaptation to other neurosurgical cases. Our developed framework for the guidance system surpasses previously described frameworks in the literature in terms of precision and computational speed. Prior efforts in the field of general surgery demonstrated various elements of utility; however, none reports segmentation of both anatomical landmarks and surgical tools intraoperatively. Moreover, all segmentation efforts in general surgery are limited to YOLOv3-based bounding boxes and semantic segmentation techniques. Nakanuma et al. recently published a feasibility trial (J-SUMMIT-C-01) for a YOLOv3-based object detection framework to be used for intraoperative guidance in laparoscopic cholecystectomy (LC). Although they used the YOLOv3 framework, they were able to demonstrate an objective usefulness of an AI-powered surgical guidance platform (11, 12). In addition, Liu et al. provided supporting evidence that their YOLOv3 based framework identified anatomical structures within LC more accurately than their trainees and senior surgeons (13). Laplante et al., Madani et al., and Mascagni et al., on the other hand, utilised a semantic segmentation method to successfully determine safe and danger zones for surgical dissection and labelling anatomical structures relevant in LC (7, 14). Moving forward, Kitaguchi and his team developed an instance segmentation model but they have only classified surgical instruments in laparoscopic colorectal surgeries (9). In neurosurgery, Bouget et al. reports an attempt to identify intraoperative tools by an outdated method of semantic labelling and shape-based detection using supervised-vector machine (SVM) training (22), which has been further improved by Kalavakonda et al., with binary and instance segmentation approach (23). Therefore, our study reports the first and the most advanced use of instance segmentation in the surgical field to date.

Our extensive experience and previous failed attempts have significantly contributed to the development of a scalable segmentation framework in surgery. We have encountered challenges related to colourisation, the MaskRCNN architecture,

SOLOv1, and KLT feature trackers, which have informed our understanding of these issues. After careful evaluation, we opted to modify the SOLOv2 architecture due to its superior performance compared to counterparts like YOLOv3 (24). Moreover, another compelling reason to use SOLOv2 was due to its ease of debugging. With SOLOv2, it becomes possible to visualize the features for each grid point, considering various kernel choices. This ability enhances the network's expressive power, allowing for a deeper understanding of the underlying processes and facilitating effective troubleshooting. Ultimately, through extensive experimentation and framework modifications, we have successfully established a platform that facilitates the swift adoption of new segmentation frameworks for improved outcomes.

Additionally, we have gained valuable insights into the use of more objective metrics for instance segmentation. While mean Average Precision (mAP) has been commonly endorsed in clinical literature, its susceptibility to false positives makes it less suitable as a metric for instance segmentation in the surgical field (25). Instead, metrics such as the F1-score, precision, and recall provide a more accurate evaluation (25). Nakanuma et al. used the DICE coefficient as a metric for accuracy; however, we have incorporated this into our loss function (12).

In this study, we present a novel comparison of sample efficiency among different intraoperative CV architectures, which, to our knowledge, is the first attempt of its kind in the literature. To measure sample efficiency, we calculated a ratio for each version of our SMI framework. This ratio represents the percentage of total available data used for training compared to the percentage used for testing, while achieving the same level of segmentation accuracy. Our findings reveal a positive trend in reducing the reliance on training annotation (**Figure 7**), distinguishing our study from previous efforts in utilising CV for intraoperative guidance, which have shown limited exploration in this particular area.

4.1. Limitations

There are some technological limitations in our platform. An unknown object inserted in the surgical field, such as a metal clip, can interfere with landmark recognition. This issue has not been previously reported or addressed in the existing literature. The current algorithms used in surgery are supervised and primarily designed for two-dimensional analysis of a three-dimensional surgical field. However, in the presence of unknown objects that occlude the known objects, these algorithms may encounter difficulties in recognition. In this study, we have not evaluated the performance of the algorithms in cases involving occlusion by unknown objects. In addition to occlusion, there is a significant translation of the view when the microscope is moved, which impacts the tracking of the algorithm. Specifically, we plan to incorporate considerations for translation in out-of-frame surgeries and for instances where the anatomy is obscured by surgical tools. We intend to address both these issues in the next update of our algorithm. Lastly, we observed instances where frames or anatomical structures appeared blurry interfering with recognition and tracking. This blurriness can be attributed to the inherent challenges of object detection in microscopic surgery, where the lens may be out-of-focus. Additionally, anatomical structures can appear hazy due to fluid or bleeding. To mitigate these issues, we intend to enhance our model's ability to recognise such barriers and refrain from providing anatomical predictions when these barriers are identified.

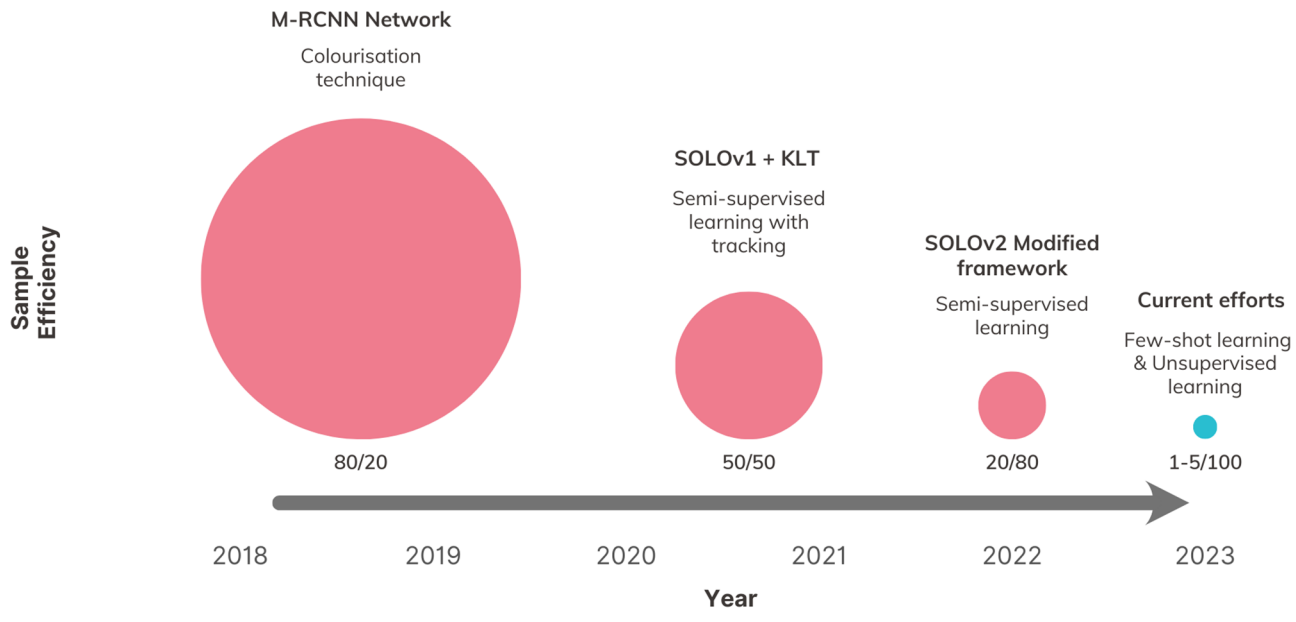


FIGURE 7
Comparison of sample efficiency across the versions of our SMI.

We evaluated the architecture on a single surgical video from one patient. However, to ensure the clinical applicability of the platform, it is essential to conduct further multi-centred trials involving multiple patients and various types of surgeries. However, previous research by Tokuyaso et al. has reported poor concordance among surgeons when labelling anatomical data. Therefore, to minimise bias and subjectivity (7, 26), we plan to involve multiple experts in annotating the surgical images and assess the inter-rater reliability. Additionally, it is important to note that spinal dAVF surgery is a microscopic procedure that does not currently allow our platform to provide superimposed guidance during surgery, as seen in many laparoscopic studies (7, 11, 12). To address this limitation, our research group is currently exploring an upgraded architecture that can process sparsely labelled data from multiple patients who have undergone endoscopic microvascular decompression (MVD) surgery (27).

4.2. Future directions

Our plans involve developing our network to overcome the technological limitations associated with identifying unknown objects, visual obstruction by tools, and out-of-focus frames (7, 11). To address the issue of unknown objects, we propose the network to adapt to and recognise previously unseen objects. For accurate tracking, our objective is to implement video-based segmentation that enables continuous tracking even when objects appear and disappear within the frame whilst retaining information about object trajectories and make more informed predictions. Furthermore, we aim to develop task-driven segmentation, as segmentation itself can be an ill-posed task, even with the availability of ground truth annotations. By implementing these advancements, we aim to address the technological limitations and enhance the performance and versatility of our network for intraoperative guidance.

In future updates, our goal is to incorporate both spatial and temporal annotation by utilizing a combination of semantic, instance, and phase recognition techniques, which were not explored in this study due to its scope limitations. By integrating these techniques effectively, we anticipate significant improvements in both the accuracy of spatial annotation and the overall understanding of surgical procedures. This expanded annotation approach holds great potential in providing valuable insights for surgery, including early error detection, surgical decision support, and performance feedback in complex neurosurgical cases. These advancements have the potential to enhance surgical care by enabling a more comprehensive analysis of surgical procedures and facilitating continuous improvement in surgical outcomes (28–30).

As demonstrated with a previous version of our architecture (27), we are currently extending our framework to be used in endoscopic surgeries that involve decompression of cranial nerves at the skull base. In the preliminary work, we conduct experiments using multiple patient videos and leverage Few-Shot Learning techniques within a sparsely labelled paradigm (27). The need for numerous expert annotations and validation poses a challenge for the generalisation of our framework across various surgical cases and specialties. Therefore, we are currently exploring different models of unsupervised learning techniques to improve the utilisation of sparsely labelled datasets and sample efficiency. With transfer learning capabilities of our CV methodology, we are actively scaling the core technology for future iterations and further development of the SMI.

Addressing these limitations provides an outlook that can have a revolutionising impact on global neurosurgery, by improving the standard of neurosurgical care and training. Our ongoing study exemplifies the foundational technology behind a SMI, a concept that aims to enhance patient outcomes and provide better training opportunities.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

Ethics statement

The studies involving humans were approved by University of Pennsylvania Institutional Review Board (IRB: 831602). The studies were conducted in accordance with the local legislation and institutional requirements. The Ethics Committee/ institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because Informed consent was attained for surgery however this is a retrospective review without patient identifiable information. Therefore, informed consent has been waived.

Author contributions

JP: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. ND: Conceptualization, Data curation, Formal Analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. XZ: Methodology, Software, Validation, Visualization, Writing – review & editing. LP: Methodology, Software, Validation, Visualization, Writing – review & editing. RB: Data curation, Resources, Writing – review & editing. JS: Supervision, Writing – review & editing. VB: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Software, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

Funding provided by the Penn Health Tech Grant.

Conflict of interest

VB and JS filed intellectual property that initiated this work. VB has equity ownership in TAIRIS, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Birkhoff DC, van Dalen ASHM, Schijven MP. A review on the current applications of artificial intelligence in the operating room. *Surg Innov*. (2021) 28 (5):611–9. doi: 10.1177/1553350621996961

2. Tariciotti L, Palmisciano P, Giordano M, Remoli G, Lacorte E, Bertani G, et al. Artificial intelligence-enhanced intraoperative neurosurgical workflow: current knowledge and future perspectives. *J Neurosurg Sci*. (2022) 66(2):139–50. doi: 10.23736/S0390-5616.21.05483-7

3. Gordon L, Grantcharov T, Rudzicz F. Explainable artificial intelligence for safe intraoperative decision support. *JAMA Surg*. (2019) 154(11):1064. doi: 10.1001/jamasurg.2019.2821

4. Filicori F, Meireles OR. Artificial intelligence in surgery. In: *Artificial intelligence in medicine*. Cham: Springer International Publishing (2022). 855–62.

5. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. (2018) 268(1):70–6. doi: 10.1097/SLA.0000000000002693

6. Kavlakoglu E. IBM. 2020. AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?.

7. Madani A, Namazi B, Altieri MS, Hashimoto DA, Rivera AM, Pucher PH, et al. Artificial intelligence for intraoperative guidance. *Ann Surg*. (2022) 276(2):363–9. doi: 10.1097/SLA.0000000000004594

8. pawangfg. GeeksForGeeks. 2022. p. 06 Object Detection vs Object Recognition vs Image Segmentation. Available at: https://www.geeksforgeeks.org/object-detection-vs-object-recognition-vs-image-segmentation/ (Cited June 21, 2023).

9. Kitaguchi D, Lee Y, Hayashi K, Nakajima K, Kojima S, Hasegawa H, et al. Development and validation of a model for laparoscopic colorectal surgical instrument recognition using convolutional neural network–based instance segmentation and videos of laparoscopic procedures. *JAMA Netw Open*. (2022) 5 (8):e2226265. doi: 10.1001/jamanetworkopen.2022.26265

10. Cerón JCÁ, Ruiz GO, Chang L, Ali S. Real-time instance segmentation of surgical instruments using attention and multi-scale feature fusion. *Med Image Anal*. (2022) 81:102569. doi: 10.1016/j.media.2022.102569

11. Tokuyasu T, Iwashita Y, Matsunobu Y, Kamiyama T, Ishikake M, Sakaguchi S, et al. Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. *Surg Endosc*. (2021) 35 (4):1651–8. doi: 10.1007/s00464-020-07548-x

12. Nakanuma H, Endo Y, Fujinaga A, Kawamura M, Kawasaki T, Masuda T, et al. An intraoperative artificial intelligence system identifying anatomical landmarks for laparoscopic cholecystectomy: a prospective clinical feasibility trial (J-SUMMIT-C-01). *Surg Endosc*. (2023) 37(3):1933–42. doi: 10.1007/s00464-022-09678-w

13. Liu R, An J, Wang Z, Guan J, Liu J, Jiang J, et al. Artificial intelligence in laparoscopic cholecystectomy: does computer vision outperform human vision? *Artificial Intelligence Surgery*. (2022) 2(2):80–92. doi: 10.20517/ais.2022.04

14. Laplante S, Namazi B, Kiani P, Hashimoto DA, Alseidi A, Pasten M, et al. Validation of an artificial intelligence platform for the guidance of safe laparoscopic cholecystectomy. *Surg Endosc*. (2023) 37(3):2260–8. doi: 10.1007/s00464-022-09439-9

15. Jumah F, Raju B, Nagaraj A, Shinde R, Lescott C, Sun H, et al. Uncharted waters of machine and deep learning for surgical phase recognition in neurosurgery. *World Neurosurg*. (2022) 160:4–12. doi: 10.1016/j.wneu.2022.01.020

16. Bravo J, Wali AR, Hirshman BR, Gopesh T, Steinberg JA, Yan B, et al. Robotics and artificial intelligence in endovascular neurosurgery. *Cureus*. (2022) 14:1–12. doi: 10.7759/cureus.23662

17. Hollon T, Orringer DA. Label-free brain tumor imaging using Raman-based methods. *J Neurooncol*. (2021) 151(3):393–402. doi: 10.1007/s11060-019-03380-z

18. Zhu APJ, Xu B. Artificial intelligence (AI) in neurosurgery. In: *Learning and career development in neurosurgery*. Cham: Springer International Publishing (2022). 263–81.

19. Dundar TT, Yurtsever I, Pehlivanoglu MK, Yildiz U, Eker A, Demir MA, et al. Machine learning-based surgical planning for neurosurgery: artificial intelligent approaches to the cranium. *Front Surg*. (2022) 9:1–10. doi: 10.3389/fsurg.2022.863633

20. Hendricks BK, Rumalla K, Benner D, Lawton MT. Cavernous malformations and artificial intelligence. *Neurosurg Clin N Am*. (2022) 33(4):461–7. doi: 10.1016/j.nec.2022.05.007

21. Park JJ, Tiefenbach J, Demetriades AK. The role of artificial intelligence in surgical simulation. *Front Med Technol*. (2022) 4:1–8. doi: 10.3389/fmedt.2022.1076755

22. Bouget D, Benenson R, Omran M, Riffaud L, Schiele B, Jannin P. Detecting surgical tools by modelling local appearance and global shape. *IEEE Trans Med Imaging*. (2015) 34(12):2603–17. doi: 10.1109/TMI.2015.2450831

23. Kalavakonda N, Hannaford B, Qazi Z, Sekhar L. *2019 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). Autonomous neurosurgical instrument segmentation using End-to-End learning* IEEE, (2019). p. 514–6.

24. Gong D, Wu L, Zhang J, Mu G, Shen L, Liu J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol*. (2020) 5(4):352–61. doi: 10.1016/S2468-1253(19)30413-3

25. Jena R, Zhornyak L, Doiphode N, Chaudhari P, Buch V, Gee J, et al. Beyond mAP: Towards better evaluation of instance segmentation. ArXiv. 2022.

26. Mascagni P, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, et al. Artificial intelligence for surgical safety. *Ann Surg*. (2022) 275(5):955–61. doi: 10.1097/SLA.0000000000004351

27. Blue R, Doiphode N, Jena R, Madsen P, Lee JYK, Shi J, et al. Designing and Developing a Novel Deep Computer Vision Platform for Intraoperative Prediction and Analytics in Skull Base Surgery. In 2023.

28. Ward TM, Fer DM, Ban Y, Rosman G, Meireles OR, Hashimoto DA. Challenges in surgical video annotation. *Computer Assisted Surgery*. (2021) 26(1):58–68. doi: 10.1080/24699322.2021.1937320

29. Garrow CR, Kowalewski KF, Li L, Wagner M, Schmidt MW, Engelhardt S, et al. Machine learning for surgical phase recognition. *Ann Surg*. (2021) 273(4):684–93. doi: 10.1097/SLA.0000000000004425

30. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, et al. Computer vision analysis of intraoperative video. *Ann Surg*. (2019) 270(3):414–21. doi: 10.1097/SLA.0000000000003460

frontiers | Frontiers in Surgery

TYPE Original Research
PUBLISHED 18 December 2023
DOI 10.3389/fsurg.2023.1271775

OPEN ACCESS

EDITED BY
Mohammed Ali Alvi,
University Health Network (UHN), Canada

REVIEWED BY
Alessandro Boaro,
University of Verona, Italy
Redi Rahmani,
Barrow Neurological Institute (BNI), United States

*CORRESPONDENCE
Sayan Biswas
✉ sayan.biswas@nca.nhs.uk

†These authors have contributed equally to this work and share first authorship

RECEIVED 02 August 2023
ACCEPTED 27 November 2023
PUBLISHED 18 December 2023

CITATION
Biswas S, McMenemy L, Sarkar V, MacArthur J, Snowdon E, Tetlow C and George KJ (2023) Natural language processing for the automated detection of intra-operative elements in lumbar spine surgery.
Front. Surg. 10:1271775.
doi: 10.3389/fsurg.2023.1271775

Natural language processing for the automated detection of intra-operative elements in lumbar spine surgery

Sayan Biswas[1*†], Lareyna McMenemy[1†], Ved Sarkar[2], Joshua MacArthur[1], Ella Snowdon[1], Callum Tetlow[3] and K. Joshi George[4]

[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom, [2]College of Letters and Sciences, University of California, Berkeley, CA, United States, [3]Division of Data Science, The Northern Care Alliance NHS Group, Manchester, United Kingdom, [4]Department of Neurosurgery, Manchester Centre for Clinical Neurosciences, Salford Royal Hospital, Manchester, United Kingdom

Background: The aim of this study was to develop natural language processing (NLP) algorithms to conduct automated identification of incidental durotomy, wound drains, and the use of sutures or skin clips for wound closure, in free text operative notes of patients following lumbar surgery.

Methods: A single-centre retrospective case series analysis was conducted between January 2015 and June 2022, analysing operative notes of patients aged >18 years who underwent a primary lumbar discectomy and/or decompression at any lumbar level. Extreme gradient-boosting NLP algorithms were developed and assessed on five performance metrics: accuracy, area under receiver-operating curve (AUC), positive predictive value (PPV), specificity, and Brier score.

Results: A total of 942 patients were used in the training set and 235 patients, in the testing set. The average age of the cohort was 53.900 ± 16.153 years, with a female predominance of 616 patients (52.3%). The models achieved an aggregate accuracy of >91%, a specificity of >91%, a PPV of >84%, an AUC of >0.933, and a Brier score loss of ≤0.082. The decision curve analysis also revealed that these NLP algorithms possessed great clinical net benefit at all possible threshold probabilities. Global and local model interpretation analyses further highlighted relevant clinically useful features (words) important in classifying the presence of each entity appropriately.

Conclusions: These NLP algorithms can help monitor surgical performance and complications in an automated fashion by identifying and classifying the presence of various intra-operative elements in lumbar spine surgery.

KEYWORDS

clips, dural tears, natural language processing, sutures, spine surgery, wound drains

Abbreviations

AUC, area under the receiver-operating curve; CI, confidence intervals; CPT, current procedural terminology; DCA, decision curve analysis; ICD, international statistical classification of diseases; NLP, natural language processing; PPV, positive predictive value; TRIPOD, transparent reporting of multivariable prediction models for individual prognosis or diagnosis; XGBoost, extreme gradient boosting.

1. Introduction

Administrative, billing, and coding tasks are a major source of financial and economic burden on healthcare systems worldwide (1). With the increase in healthcare and labour costs in recent years, major health systems are shifting towards minimising financial expenditure while maximising patient care. A key component in this process is optimising the clinical coding pipeline by reducing the burden on labour with limited manual review and intervention. The clinical coding process involves transforming medical records, usually presented as free text written by clinicians, into structured codes using the standardised Current Procedural Terminology (CPT) and the International Statistical Classification of Diseases (ICD) codes. The purpose of such clinical coding is to characterise the use of hospital services, document patient outcomes, and quantify clinical and surgical practices to allow for optimal financial reimbursement and to inform healthcare service planning and policy (2, 3).

Natural language processing (NLP) is a domain of machine learning that focuses on the analysis of structured and unstructured free text. NLP techniques are well suited for clinical coding due to their ability to analyse free text in real time with great precision. In the United Kingdom, the General Medical Council states that maintaining accurate and detailed clinical documentation is essential across all specialties for good medical practice (4), in addition to providing information for research, audits, and medicolegal records (5, 6). The current epidemic of defensive practice due to fear of medicolegal repercussions has had an extensive impact on neurosurgical documentation practices, resulting in more detailed documentation of procedures (7). Despite guidelines being available for the documentation of operative notes (8), many studies have demonstrated the inadequate quality of operative notes with much salient information missing, including the nature of the surgery, indication of surgery, estimated blood loss, incidence of complications, and postoperative instructions (6, 9–11). Such non-standardised documentation can lead to greater manual review times, making the extraction of relevant information more labour-intensive. The creation of accurate NLP algorithms trained on a large number of heterogeneous documents can be used to supplement the current clinical coding process, reducing the need for extensive and tedious manual reviews.

Spine surgery comprises the majority of operative cases in neurological surgery. Incidental durotomy, lumbar drains, and type of skin closure (sutures or clips) are important elements included in operative notes and are associated with patient outcomes, and therefore accurate documentation is vital to inform best clinical practice (12–16). At present, CPT and ICD-10 codes are used to identify incidental durotomies and "dural tears" within operative notes. However, these modalities have been shown to lack sensitivity, resulting in the underreporting of these complications (17–19). To the best of our knowledge, no such codes exist for the identification of the use of drains or wound closure technique used. Hence, the aim of this study is to develop NLP algorithms to conduct automated surveillance for identification of incidental durotomy, wound drains, and the use of sutures or skin clips for wound closure, in free text operative notes of patients following lumbar surgery. Towards this, in this study we attempted to evaluate if NLP techniques could be harnessed to analyse operative notes to detect the three important elements of spine surgery: incidental durotomy, the use of wound drains, and type of skin closure (suture or clips).

2. Materials and methods

2.1. Guidelines

The following guidelines were followed in this study: the *Journal of Medical Internet Research* (JMIR) Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research, and the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis (TRIPOD) checklist (20, 21).

2.2. Data source and outcome measure

A single tertiary neurosurgical centre retrospective case series analysis was conducted for all patients who underwent lumbar spine surgery between January 2015 and June 2022. The inclusion criteria for this study were as follows: (1) patient age more than or equal to 18 years, (2) patient underwent a primary lumbar discectomy and/or decompression at any lumbar level, and (3) availability of index surgery operation notes in our electronic health records. The exclusion criteria included any patients with incomplete data and patients who underwent primary lumbar discectomy and/or fusion. The hospital's electronic patient records were examined and a total of 1,177 patients were identified. Each patient's operation note was then blinded and extracted in an anonymised manner. Our study was approved by the local hospital's institutional review board because of the retrospective and anonymised operative note data collection method. The study was registered as a health improvement project with the requirement for patient consent being waived. All methods were conducted in accordance with local and national guidelines and regulations.

Along with the operation notes, the age (continuous) and gender of the patient (male or female) were also collected as independent variables. There were three primary outcomes for each operation note: (1) the presence of intra-operative durotomies (binary outcome), (2) the placement of wound drains (binary outcome), and (3) the use of clips or sutures for skin closure (binary outcome). The terms durotomies and dural tears are used interchangeably in this paper. Each patient's operation note was reviewed and annotated by blinded researchers (LM and SB) who were not involved in the care of these patients. The results of each outcome category were then verified by the senior author.

2.3. Data pre-processing

The data acquisition, pre-processing, model development, and evaluation pipeline have been highlighted in Figure 1. The dataset was initially cleaned with a custom data-cleaning function that consisted of the removal of special characters retrieved from the Natural Language Toolkit (NLTK) such as "@/{$#%&" and stopwords including "*and*", "*or*", and "the". These words do not carry significant meaning or information in text analysis tasks, hence their removal helps to de-noise the text data resulting in the better efficiency and performance of NLP models. Stemming and lemmatisation are two common techniques used in the data-cleaning function, both of which aim to normalise words by reducing them to their base or root forms. Stemming achieves this by removing any suffixes at the end of a word, while lemmatisation is the process of reducing a word to its base or dictionary form (known as the lemma) while taking into account the context and part of speech of the word.

Lastly, the CountVectorizer library function was used to pre-process the cleaned data. By default, CountVectorizer uses the "term frequency" weighting for single tokenisation, which means it represents each word by the number of times it appears in a document. This results in a document-term matrix where each element represents the frequency of a particular word in a specific document. The resulting matrix is then used as the input to various machine learning algorithms such as clustering, classification, and topic modelling. By representing text data in a numerical format, the CountVectorizer enables machine learning (ML) algorithms to process and analyse textual data, which would otherwise be difficult due to the unstructured nature of natural language.

2.4. Model development

An 80:20 training–testing split was carried out on the total cohort of 1,177 patients, with 942 patients in the training set and 235 patients in the testing set. The datasets were stratified for the outcome variables to account for class imbalances. An extreme gradient-boosting (XGBoost) NLP classifier was developed to predict each outcome category. XGBoost was selected as the classifier of choice owing to a number of factors: (1) its ability to handle high-dimensional feature spaces such as word to vector embeddings used in NLP, (2) the ability to handle and adjust for sparse and imbalanced datasets using weighted loss functions and subsampling, (3) its faster computational run time and scalability, and (4) explicit feature importance calculation for each input attribute (11, 22). Three individual models were created for identifying each outcome category, and the outputs from the models were concatenated to produce a multilabel ensemble output with the predicted probabilities for each outcome. The three ML models will be referred to as the dural tear, drains, and clips vs. sutures models in this paper. An iterative process termed Grid Search was used to optimise the model hyperparameters. In grid search, a predefined set of hyperparameter values is defined, and the model is trained and evaluated on all possible combinations of these values to achieve the highest level of accuracy.

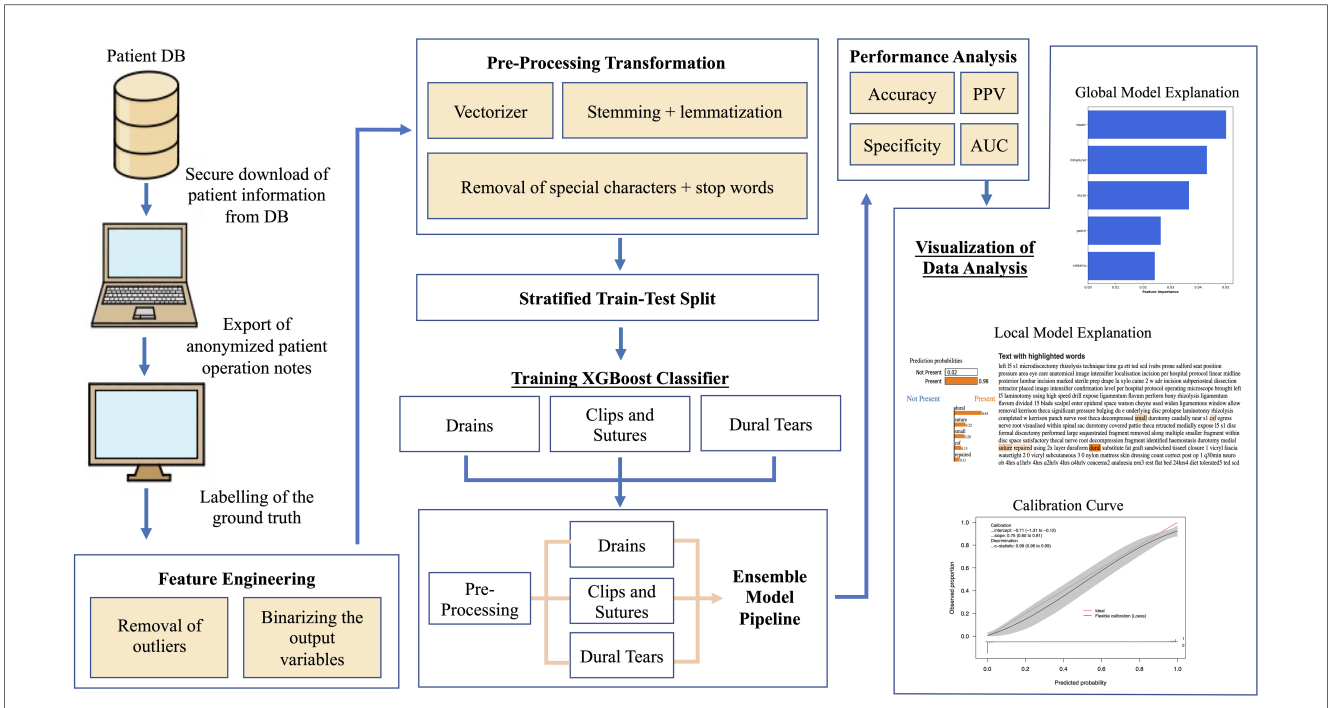


FIGURE 1
Data acquisition, processing, analysis, and visualisation pipeline. DB, database.

The models were trained on fivefold stratified K-fold cross-validation with five repeats on the training dataset. The training and testing datasets were stratified by each of the outcome categories to standardise the class imbalances within our outcome variables and provide us with the best overall performance results for the models. The performance of the models were evaluated via five performance metrics on the training and testing sets: accuracy, precision/positive predictive value (PPV), specificity, area under the receiver-operating curve (AUC)/discrimination, and the Brier score loss. All metrics were bootstrapped with 1,000 resamples to derive the associated 95% confidence intervals (CIs). Each model was then calibrated on the testing set. Calibration refers to how well a model's predicted probabilities align with the true observed probabilities in the study population. This is evaluated using a calibration curve, which is ideally a 45° straight line starting from the origin, with a slope of 1 (indicating the spread of the model's estimated probabilities over the observed probabilities), and an intercept of 0 (indicating how much the model tends to over- or underestimate the true probability). In this study, the preferred method of calibration was Platt scaling or sigmoid binned calibration, which involved dividing the probability range into 10 bins and evaluating the shape of the calibration curve, as well as its slope, intercept, and the Brier score loss metric. In addition, the decision curve analysis (DCA) was used to evaluate and plot the clinical benefit of using the NLP algorithms to predict the presence of each outcome variable over a wide range of predicted threshold probabilities. The DCA illustrates the net benefit defined as the number of true positives detected for each outcome class when using the NLP algorithms on individual patient operation notes.

A model-specific global feature importance analysis was conducted on the trained models via retrieval of each model's relative feature weights that were averaged across all training folds. Furthermore, the Local Interpretable Model-agnostic Explanation analysis was performed to predict and highlight the important features on an individual patient operation note level.

2.5. Statistical analysis

All statistical analyses were conducted using IBM SPSS software (Statistical Package for the Social Science; SPSS Inc., Chicago, IL, USA) Version 25 for Mac, Microsoft Excel (Office 365, Microsoft, Seattle, WA, USA), and the R coding language (R Foundation for Statistical Computing, Vienna, Austria). Histogram plots and the Kolmogorov–Smirnov test were utilised for tests of normality for the continuous variables. The chi-squared tests were used to compare all categorical variables, and the independent samples *t*-test was used to compare the means of the continuous variables. Temporal trend analysis with a linear line of best fit was conducted for all variables, within our retrospective observation time period. A *p*-value <0.05 was considered statistically significant.

TABLE 1 Cohort demographics of the total patient cohort.

	Total cohort ($n = 1,177$)
Age	53.900 ± 16.153
Sex	
Female	616 (52.3%)
Male	561 (47.7%)
Drain(s)	
Yes	373 (31.6%)
No	801 (67.8%)
Closure	
Clips	458 (38.9%)
Sutures	710 (60.3%)
Dural tear(s)	
Yes	117 (9.9%)
No	1,060 (90.1%)

3. Results

3.1. Cohort demographics

A total of 1,177 patients were included in the study, with 942 patients in the training set and 235 patients in the testing set. Table 1 demonstrates the total cohort demographics. The average age of the cohort was 53.900 ± 16.153 years, with a female predominance of 616 patients (52.3%). The rates of intra-operative durotomy and the use of wounds drains were 9.9% (117/1,177) and 31.6% (373/1,177), respectively. Overall, the use of sutures [710 (60.3%)] was more common for skin closure compared with the use of metal surgical clips [458 (38.9%)]. The inter-variable comparative analysis (Table 2) demonstrated a significant relationship between increasing patient age and the use of sutures (p-value = 0.001). We also noted that with an ageing population, the operative age of our patients significantly increased over our observation period (p-value = 0.013). There was also a statistically significant relationship between the use of sutures for skin closure in cases with intra-operative dural tears and wound drains (p-value < 0.001). However, there was no

TABLE 2 Inter-variable statistical correlation analysis using *t*-tests for continuous variables and Chi-square tests for categorical variables.

	p-value					
	Age	Sex	Drains	Closure	Dural tear (s)	Year of surgery
Age		0.137	0.283	0.001[a]	0.501	0.013[a]
Sex						
Female	0.137		0.278	0.294	0.217	0.322
Male						
Drain(s)						
Yes	0.283	0.278		<0.001[a]	0.554	0.906
No						
Closure						
Clips	0.001[a]	0.294	<0.001[a]		<0.001[a]	0.017[a]
Sutures						
Dural tear(s)						
Yes	0.501	0.217	0.554	<0.001[a]		0.853
No						

[a]Statistically significant *p*-value.

statistically significant relationship between the use of wound drains and the presence of intra-operative dural tears (p-value = 0.554).

3.2. Temporal trend analysis

The Mann–Kendall test was used to analyse the temporal trends of the variables across our observation time period as shown in Supplementary Figure S1. During the study period, the total number of lumbar discectomies and/or decompressions decreased significantly from 220 surgeries in 2015 to 56 in the first half of 2022 (−112 estimated in a year) (tau = −0.929, p-value = 0.002). This decline was observed in all the years with the exception of 2019, which saw an increase of one operation from the previous year. It was noted that there was also a decrease in all spinal procedures post COVID-19, which may account for the decrease. The frequency of intra-operative durotomies/dural tears did decrease over the study period; however, no statistical significance was observed (tau = −0.286, p-value = 0.386), with rates ranging from 5.8% to 14.2%. The frequency of intra-operative placement of wound drains also statistically significantly increased over the study period, rising from 18.6% in 2015 to 41.1% in 2022 (tau = 0.643, p-value = 0.035). The preferred method of skin closure also changed over the study period, demonstrating a preference for closure with sutures in later years (tau = 0.5, p-value = 0.108) with a rise from 54% in 2015 to 75% in 2022. We observed an exact yet complementary decrease in the use of surgical clips for skin closure over the years (tau = −0.5, p-value = 0.108).

3.3. Model performance

Table 3 provides the performance metrics for the three ML models on the testing dataset. The dural tears model achieved an accuracy of 91.7615 (95% CI: 88.636–94.602), a PPV of 84.211% (95% CI: 80.667–90.000), a specificity of 99.032% (95% CI: 96.959–99.750), and an AUC of 0.946 (95% CI: 0.917–0.970). The drains model achieved an accuracy of 94.894% (95% CI: 92.330–97.160), a PPV of 88.696% (95% CI: 82.308–94.000), a specificity of 94.694% (95% CI: 90.886–97.025), and an AUC of 0.950 (95% CI: 0.923–0.973). The clips vs. sutures model achieved an accuracy of 93.750% (95% CI: 91.193–96.307), a PPV of 94.495% (95% CI: 91.379–97.260), a specificity of 91.177% (95% CI: 84.770–95.153), and an AUC of 0.933 (95% CI: 0.923–0.973). Figure 2 shows the calibration curves for each of the models. The dural tears model had a propensity to underpredict the presence of a dural tear, with a Brier score loss of 0.082 (95% CI: 0.054–0.114), an intercept of 0.91 (95% CI: 0.46–1.36), and a slope of 0.99 (95% CI: 0.76–1.23). The drains model demonstrated excellent calibration across all predicted probabilities with a Brier score loss of 0.051 (95% CI: 0.028–0.076), an intercept of −0.71 (95% CI: −1.31 to −0.12), and a slope of 0.75 (95% CI: 0.60–0.91). The clips vs. sutures model demonstrated a tendency to overpredict the use of sutures for skin closure, with a Brier score loss of 0.063 (95% CI: 0.037–0.088), an intercept of −0.01 (95% CI: −0.61–0.60), and a slope of 0.65 (95% CI: 0.53–0.77). Lastly, the decision curve analysis on the testing set revealed that all NLP algorithms ensured greater clinical net benefit at all possible threshold probabilities relative to the default decisions of changes made for all or none patients (Figure 3).

3.4. Model explainability

The global feature importance calculations for the NLP algorithms are presented in Figure 4. These explanations highlight that for identification of an intra-operative durotomy, the five most meaningful features (words) are: "repair," "intradural," "dural," "patch," and "Valsalva." The five most important features (words) for detecting the intra-operative placement of a lumbar drain are: "drain," "fascial," "scoliosis," "clotting," and "incision." Similarly, for detecting whether surgical clips or traditional sutures were utilised for skin closure, the following five words were the most important: "clip," "staple," "warmer," "lamina," and "clamp." In addition, the local feature importance analysis for an example patient level operation note demonstrates that the dural tear model is able to identify the five most important clinically meaningful features to detect the presence of an intra-operative dural tear (Figure 5). Interestingly, the local feature importance analysis for the drains and clips vs. sutures model demonstrated that the algorithm primarily searched only for the words "drain" and "clip," respectively, to make the prediction, with the other aforementioned features possessing very little impact on the outcome.

4. Discussion

This study analysed the trends in the use of various intra-operative elements in spine surgery and developed NLP algorithms capable of reliably identifying these elements in operative notes. The automated identification of these elements

TABLE 3 Performance metrics of the machine learning model on the testing set with 95% confidence intervals.

Model	Accuracy (%)	Precision/PPV (%)	Specificity (%)	AUC	Brier score loss
Testing set (n = 235)					
Dural tears	91.761 (88.636–94.602)	84.211 (80.667–90.000)	99.032 (96.959–99.750)	0.946 (0.917–0.970)	0.082 (0.054–0.114)
Drains	94.894 (92.330–97.160)	88.696 (82.308–94.000)	94.694 (90.886–97.025)	0.950 (0.923–0.973)	0.051 (0.028–0.076)
Clips vs. sutures	93.750 (91.193–96.307)	94.495 (91.379–97.260)	91.177 (84.770–95.153)	0.933 (0.923–0.973)	0.063 (0.037–0.088)

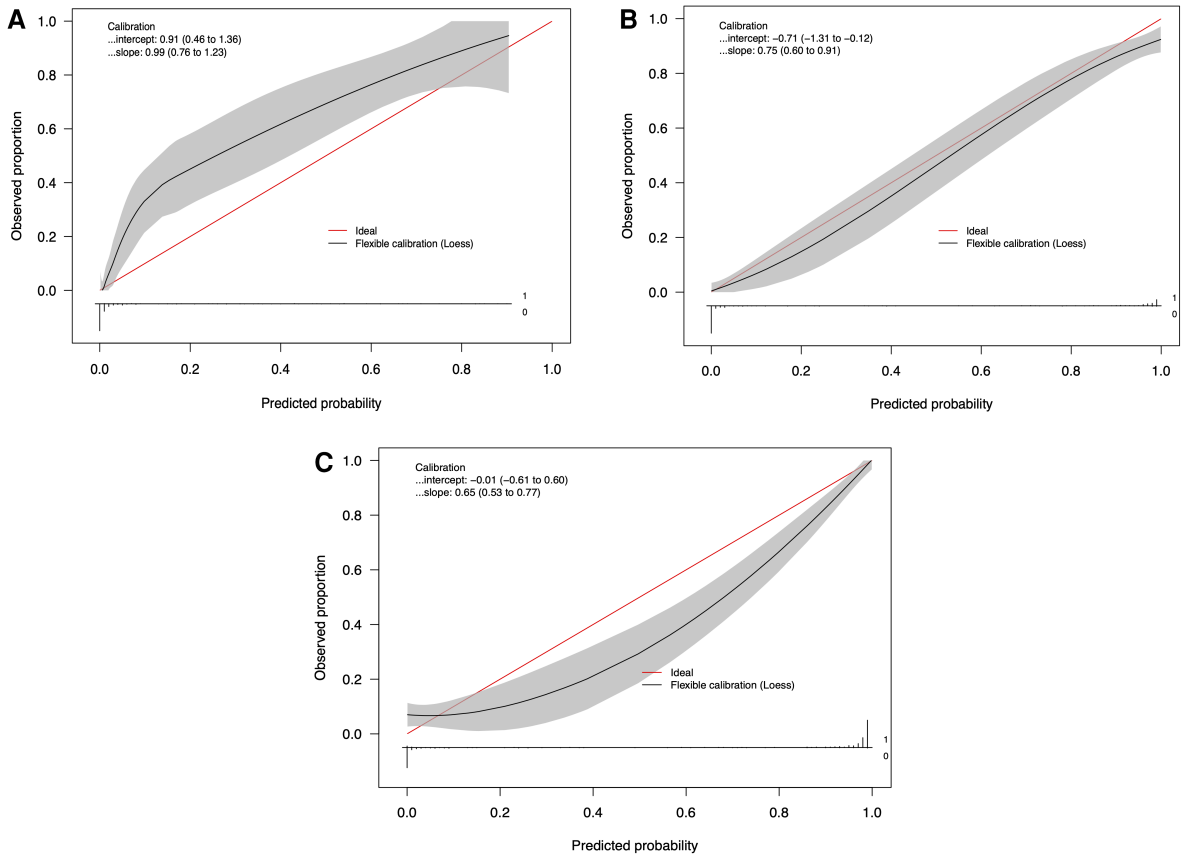


FIGURE 2
Calibration curves of the natural language processing models for identifying (**A**) dural tears, (**B**) wound drains, and (**C**) clips vs. sutures, in the testing set (*n* = 235).

can facilitate more efficient clinical coding and billing processes, help optimise hospital quality improvement and safety efforts, assist clinicians in auditing surgical practices, and guide overall resource allocation. This study demonstrates that our NLP algorithms are capable of reliably and accurately identifying the placement of intra-operative wound drains, the presence of incidental dural tears, and whether surgical clips or sutures were utilised for skin closure. This is the first ever study from a country with a public healthcare system that has demonstrated the feasibility of using automated NLP systems in operative notes to potentially guide both surgical practices and resource allocation.

The use of NLP techniques in spine surgery has seen a rise in the recent years and is projected to rapidly grow in the future (23, 24). The ability of NLP to perform precise automated surveillance of operative notes, to answer clinically relevant questions, serves to reduce the burden of time-intensive and error-prone reviews by clinical coders (25, 26). The delays in clinical coding within the National Health Service (NHS) impose a significant burden, with the potential for funding to be blocked if coding is not completed within a prerequisite timeframe (27). The average accuracy of this coding has been reported at approximately 83%, with large inter-study variability (28). Such problems exist in majority of healthcare systems worldwide and necessitate the development of automated techniques capable of facilitating these burdensome manual record review processes. Within this realm, Zaidat et al. have already developed an XLNet model capable of automatically generating CPT billing codes from operative notes for three specific surgical procedures: anterior cervical discectomy and fusion (ACDF), posterior cervical discectomy and fusion (PCDF), and cervical disc arthroplasty (CDA) (2). Such models have the potential to greatly reduce manual review/input, minimise errors in the coding process, and promote standardisation. Most recently, Shost et al. have also demonstrated a model capable of reliably identifying the type of spinal surgery performed via analysis of patient consent forms (29). The ability to rapidly classify surgical practices can be beneficial to both hospitals and the practicing surgeons. This will help track surgical volume, surgery-specific patient outcomes, and also provide trainees with a method of tracking individual surgical experience. In addition, NLP algorithms have also demonstrated predictive value in classifying lumbar spine imaging findings and in determining the need for surgical intervention in patients with low back pain via analysis of radiological and clinical reports (30). These examples highlight the importance of NLP techniques in improving the provision of patient care and demonstrate the clinical utility of such models in enhancing hospital and surgical practices.

Our NLP algorithms were developed to identify the presence of three important intra-operative factors that play a role in guiding

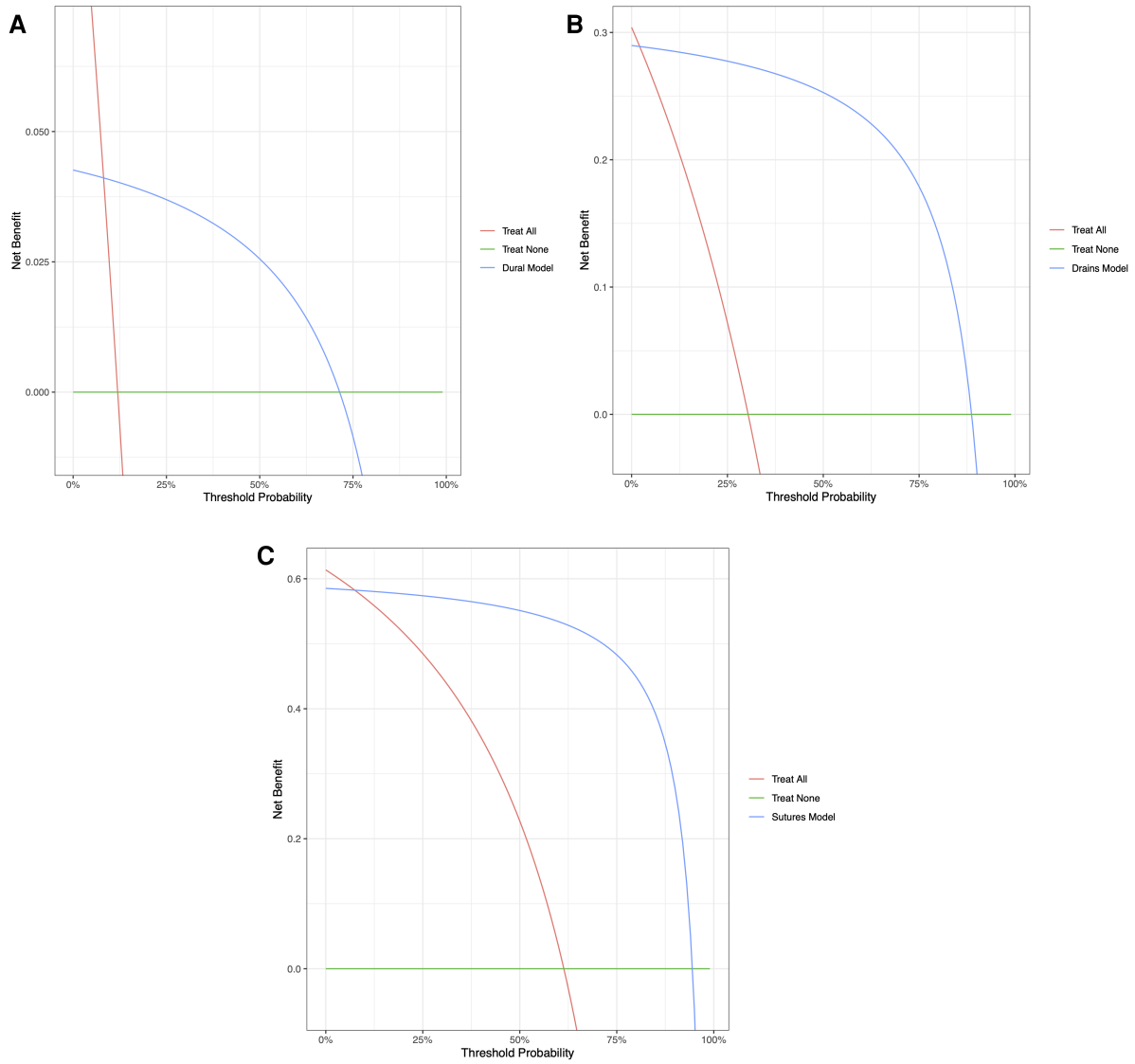


FIGURE 3
Decision curve analyses comparing expected clinical net benefit of the three models: (A) dural tears, (B) wound drains, and (C) clips vs. sutures, on the testing set.

the resource allocation and surgical practices of a neurosurgical department. In this study, the prevalence of incidental durotomy was 9.9%, in line with the recent literature on lumbar surgery (18, 31, 32). Our model demonstrated adequate discrimination and performance in identifying intra-operative dural tears and highlighted the use of clinically relevant features (words) to make its predictions. Previous studies by Karhade et al. have also been successful in the identification of incidental durotomy with an accuracy of 99%, surpassing the performance of CPT and ICD-10 codes, which demonstrated an accuracy of only 64% (18). Interestingly, however, the feature importance in their NLP algorithm showed different features compared with ours, further underscoring the potential variability in NLP algorithm performance across different cohorts that are geographically separated, and highlighting the need for broader validation studies (17). The importance of reliably identifying cases of intra-operative incidental durotomy is highlighted by evidence suggesting that patients with durotomies tend to have increased operative durations and inpatient length of stay (LOS) (33). Thus, accurate depiction of the rates of incidental durotomy can aid postoperative patient counselling, quantify surgical complication rates, and help track surgical performance.

For the wound drains model, our study demonstrated an accuracy of almost 95%. Previous studies have concluded that postoperative drains are currently being overused in spinal surgery, potentially imposing an increased risk of unnecessary complications, while not lending substantial benefit (34). Most notably, reports have suggested an elevated risk of surgical site infections (SSI) (35, 36), although this has been refuted by other papers (15, 37–39). Ho et al. interestingly report that both the absence of a wound drain and increased drainage when drains are used indicate an increased risk of delayed infection after posterior

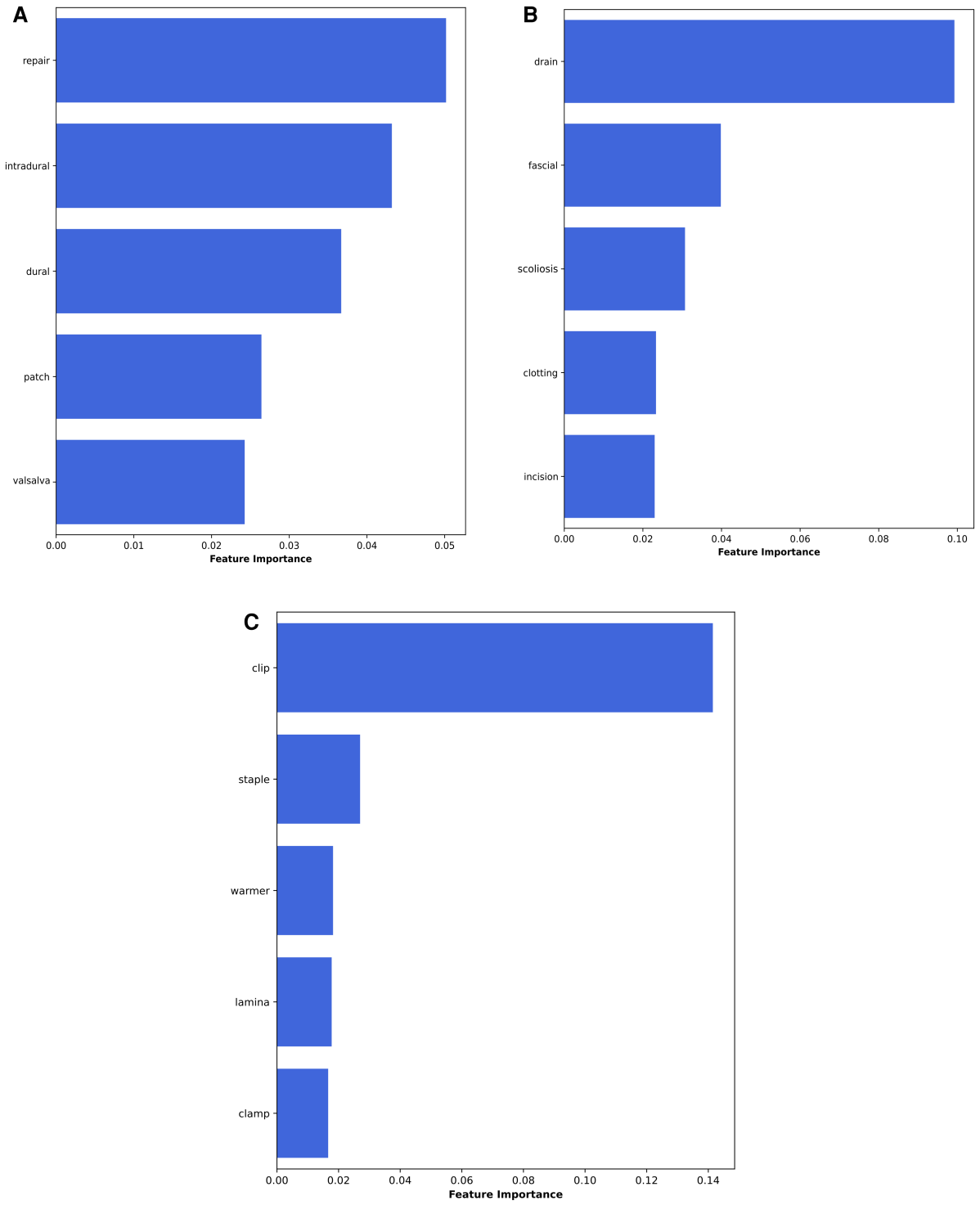
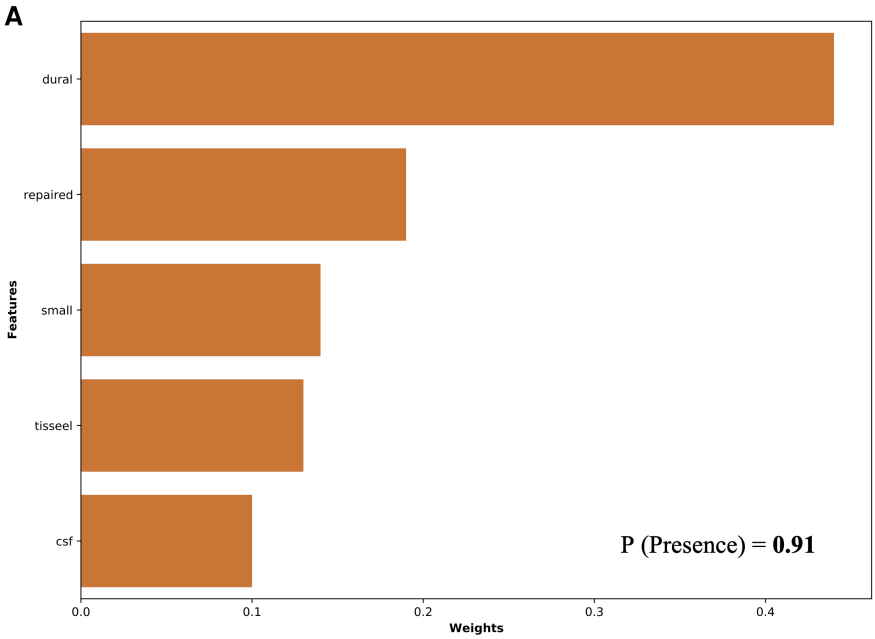


FIGURE 4
Global feature importance values for the NLP algorithms: (**A**) dural tears, (**B**) wound drains, and (**C**) clips vs. sutures.

spine surgery (36). Walid et al. additionally found that the use of postoperative drains was linked to increased post-haemorrhagic anaemia, and a subsequent requirement of allogenic blood transfusions (40), which may impose greater costs to the healthcare system. Adogwa et al. have also demonstrated that patients with postoperative drains have a significantly longer LOS compared with patients with no drains (37). The combination of such factors highlights the importance of tracking and quantifying the use of drains in spine surgery, and therefore the development of our NLP algorithm will allow for its automated and reliable detection. The future application of this algorithm in tracking wound drain use and the associated SSI rates remains to be investigated.



B

left l5 s1 microdiscectomy rhizolysis technique time ga ett ted scd ivabs prone salford seat position pressure area eye care anatomical image intensifier localisation incision per hospital protocol linear midline posterior lumbar incision marked sterile prep drape la xylo caine 2 w adr incision subperiosteal dissection retractor placed image intensifier confirmation level per hospital protocol operating microscope brought left l5 laminotomy using high speed drill expose ligamentum flavum perform bony rhizolysis ligamentum flavum divided 15 blade scalpel enter epidural space watson cheyne used widen ligamentous window allow removal kerrison theca significant pressure bulging du e underlying disc prolapse laminotomy rhizolysis completed w kerrison punch nerve root theca decompressed small durotomy caudally near s1 csf egress nerve root visualised within spinal sac durotomy covered pattie theca retracted medially expose l5 s1 disc formal discectomy performed large sequestrated fragment removed along multiple smaller fragment within disc space satisfactory thecal nerve root decompression fragment identified haemostasis durotomy medial suture repaired using 2x layer duraform dural substitute fat graft sandwiched tisseel closure 1 vicryl fascia watertight 2 0 vicryl subcutaneous 3 0 nylon mattress skin dressing count correct post op 1 q30min neuro ob 4hrs q1hrly 4hrs q2hrly 4hrs q4hrly concerns2 analgesia prn3 rest flat bed 24hrs4 diet tolerated5 ted scd tinzaparin 40mg c od tomorrow evening6 monitor wound csf leak7

FIGURE 5
Local feature importance analysis for detecting dural tears in an example individual patient operation note as generated by the NLP algorithm.

Further, our clips vs. suture model demonstrated an accuracy of >93% accuracy, and our temporal trend analysis showed a preference for using sutures for wound closure. Various studies have concluded that suturing is more efficient when compared with the use of clips for good wound closure, resulting in lower rates of separation, prevention of SSI, and ultimately shorter hospital LOS (41, 42). Contrastingly, postoperative analysis of visual analogue pain scores comparing the use of clips to non-absorbable sutures have also demonstrated a significantly quicker and pain-free experience for patients with stapled wounds (43). From an economic perspective as well, studies have demonstrated that staples/clips are less expensive than sutures and that the financial gain appears to increase as laceration length increases (44). However, conflicting literature exists on the impact of sutures and clips on patients postoperatively (22), with the need for future robust randomised control trials to further investigate their effects. Nevertheless, such single-use surgical items are the largest contributors to the surgical carbon footprint and hence precise quantification of such use can guide both financial and environmental practices (45). Therefore, such automated NLP techniques can facilitate accurate data collection and analysis of the use of clips and sutures in neurosurgery. Nevertheless, the

utility of this NLP algorithm in identifying and predicting postoperative LOS, risk of SSI, and the estimated carbon footprint after a surgery remains to be explored in a future study.

4.1. Limitations

Despite these results, the study has several limitations. First, this was a retrospective analysis at a single centre and therefore the development and testing of the NLP algorithms was geographically limited to a specific region. This raises questions about the algorithms' generalisability and their performance in diverse linguistic and clinical contexts. Furthermore, the surgeons affiliated with the healthcare entities in the study likely share practices that influence the specific terminology used to document the various intra-operative characteristics, which could bias the results. Hence, future prospective and external validation of the algorithms needs to be performed to validate the clinical utility of the algorithms. In addition, there are other approaches that can be utilised to adapt our general model to geographically distinct regions. Geographically customisable models can be implemented via techniques such as federated learning and transfer learning. Federated learning enables the collaborative training of models across multiple centres without data sharing, preserving both privacy and centre-specific relationships and trends in the data. Transfer learning further facilitates rapid fine-tuning, which can efficiently adapt a base model to new regions by learning from small local datasets, boosting model performance and reliability. Secondly, though these models are able to reliably identify the outcomes of interest, a further manual review by clinical coders will still be required to exclude any cases of false positives or false negatives. Thus, the need for manual review will still exist, though with a considerably lower level of burden. Hence, multicentre, linguistically different validation studies in hospitals with varying coding/billing practices are required to determine the reliability of these models. Lastly, with the advent of state-of-the-art large language models such as Bidirectional Encoder Representations from Transformers and Generative Pre-trained Transformer models, the need for manual annotation of unstructured, free text data may exponentially reduce. These models are capable of independently performing named entity recognition and can understand the contextual nuances of each outcome of interest. For example, these models would be able to interpret the reason/context for using a drain, or the reason for a durotomy. Thus, in the future the goal would be to develop such models capable of functioning independently without the need for any manual annotation or review.

5. Conclusion

In conclusion, this study evaluated the feasibility and reliability of NLP algorithms in determining the presence of three intra-operative elements in lumbar spine surgery. We demonstrate that these NLP models possess great discriminative ability and accuracy in predicting the presence of wound drains, incidental dural tears, and the use of clips or sutures for wound closure. These models can help automate the clinical coding process, help optimise hospital quality improvement, and monitor surgical performance and practices. This is the first ever study from a country with a primarily public healthcare system that has demonstrated the feasibility of using automated NLP systems in operative notes to potentially guide both surgical practices and resource allocation.

Data availability statement

The original contributions presented in this study are included in the article/**supplementary materials** and further inquires should be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Northern Care Alliance Research and Innovation Team. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

SB: Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing. LM: Data curation, Formal analysis, Writing – original draft. VS: Formal analysis, Software, Writing – review & editing. JM: Conceptualization, Writing – review & editing. ES: Writing – original draft, Writing – review & editing. CT: Data curation, Methodology, Supervision, Writing – review & editing. KG: Methodology, Supervision, Writing – review & editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this paper.

Acknowledgments

The authors wish to acknowledge the Department of Neurosurgery at Salford Royal Hospital for providing the data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this paper can be found online at: https://www.frontiersin.org/articles/10.3389/fsurg.2023.1271775/full#supplementary-material

References

1. Chernew M, Mintz H. Administrative expenses in the US health care system: why so high? *JAMA*. (2021) 326(17):1679–80. doi: 10.1001/jama.2021.17318

2. Zaidat B, Tang J, Arvind V, Geng EA, Cho B, Duey AH, et al. Can a novel natural language processing model and artificial intelligence automatically generate billing codes from spine surgical operative notes? *Global Spine J*. (2023):21925682231164936. doi: 10.1177/21925682231164935. [Epub ahead of print]

3. Dong H, Falis M, Whiteley W, Alex B, Matterson J, Ji S, et al. Automated clinical coding: what, why, and where we are? *NPJ Digit Med*. (2022) 5(1):159. doi: 10.1038/s41746-022-00705-7

4. GMC. *Good medical practice* (2013). Available at: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice (accessed July 19, 2023).

5. Hamza A, Abdalrahim H, Idris S, Ahmed O. Evaluating the operative notes of patients undergoing surgery at Omdurman Teaching Hospital, Sudan. *Sch J App Med Sci*. (2013) 1(6):668–72. doi: 10.36347/sjams.2013.v01i06.003

6. Rogers A, Bunting M, Atherstone A. The quality of operative notes at a general surgery unit. *S Afr Med J*. (2008) 98(9):726–8.

7. Gadjradj PS, Ghobrial JB, Harhangi BS. Experiences of neurological surgeons with malpractice lawsuits. *Neurosurg Focus*. (2020) 49(5):E3. doi: 10.3171/2020.8.FOCUS20250

8. RCS. *Good surgical practice* (2019). Available at: https://www.rcseng.ac.uk/standards-and-research/gsp/ (accessed July 19, 2023).

9. Lefter LP, Walker SR, Dewhurst F, Turner RWL. An audit of operative notes: facts and ways to improve. *ANZ J Surg*. (2008) 78(9):800–2. doi: 10.1111/j.1445-2197.2008.04654.x

10. Malik S, Nogaro M, Shenoy R, Mitchell P. Improving trauma operation notes at an emerging trauma unit in the regional trauma network. *Bull R Coll Surg Engl*. (2013) 95(6):1–3. doi: 10.1308/003588413X13643054410386

11. Nyamulani N, Mulwafu W. The quality of hand-written operative notes in a surgical unit at Queen Elizabeth central hospital (QECH), Malawi: a prospective completed audit loop study. *Malawi Med J*. (2018) 30(2):86–9. doi: 10.4314/mmj.v30i2.6

12. Nandyala SV, Elboghdady IM, Marquez-Lara A, Noureldin MNB, Sankaranarayanan S, Singh K. Cost analysis of incidental durotomy in spine surgery. *Spine (Phila Pa 1976)*. (2014) 39(17):E1042–51. doi: 10.1097/BRS.0000000000000425

13. Saxler G, Krämer J, Barden B, Kurt A, Pförtner J, Bernsmann K. The long-term clinical sequelae of incidental durotomy in lumbar disc surgery. *Spine (Phila Pa 1976)*. (2005) 30(20):2298–302. doi: 10.1097/01.brs.0000182131.44670.f7

14. Strömqvist F, Sigmundsson FG, Strömqvist B, Jönsson B, Karlsson MK. Incidental durotomy in degenerative lumbar spine surgery—a register study of 64,431 operations. *Spine J*. (2019) 19(4):624–30. doi: 10.1016/j.spinee.2018.08.012

15. Kanayama M, Oha F, Togawa D, Shigenobu K, Hashimoto T. Is closed-suction drainage necessary for single-level lumbar decompression?: review of 560 cases. *Clin Orthop Relat Res*. (2010) 468(10):2690–4. doi: 10.1007/s11999-010-1235-6

16. Yilmaz E, Blecher R, Moisi M, Ankush C, O'Lynnger TM, Abdul-Jabbar A, et al. Is there an optimal wound closure technique for major posterior spine surgery? A systematic review. *Global Spine J*. (2018) 8(5):535–44. doi: 10.1177/2192568218774323

17. Karhade AV, Oosterhoff JHF, Groot OQ, Agaronnik N, Ehresman J, Bongers MER, et al. Can we geographically validate a natural language processing algorithm for automated detection of incidental durotomy across three independent cohorts from two continents? *Clin Orthop Relat Res*. (2022) 480(9):1766–75. doi: 10.1097/CORR.0000000000002200

18. Karhade A, Bongers M, Groot O, Kazarian E, Cha T, Fogel H, et al. Natural language processing for automated detection of incidental durotomy. *Spine J*. (2020) 20(5):695–700. doi: 10.1016/j.spinee.2019.12.006

19. Ehresman J, Pennington Z, Karhade A, Huq S, Medikonda R, Schilling A, et al. Incidental durotomy: predictive risk model and external validation of natural language process identification algorithm. *J Neurosurg Spine*. (2020) 33(3):342–8. doi: 10.3171/2020.2.SPINE20127

20. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med*. (2015) 13(1):1. doi: 10.1186/s12916-014-0241-z

21. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*. (2016) 18(12):e323. doi: 10.2196/jmir.5870

22. Selvadurai D, Wildin C, Treharne G, Choksy SA, Heywood MM, Nicholson ML. Randomised trial of subcuticular suture versus metal clips for wound closure after thyroid and parathyroid surgery. *Ann R Coll Surg Engl*. (1997) 79(4):303–6.

23. Huang BB, Huang J, Swong KN. Natural language processing in spine surgery: a systematic review of applications, bias, and reporting transparency. *World Neurosurg*. (2022) 167:156–64.e6. doi: 10.1016/j.wneu.2022.08.109

24. Agaronnik ND, Kwok A, Schoenfeld AJ, Lindvall C. Natural language processing for automated surveillance of intraoperative neuromonitoring in spine surgery. *J Clin Neurosci*. (2022) 97:121–6. doi: 10.1016/j.jocn.2022.01.015

25. Bacco L, Russo F, Ambrosio L, D'Antoni F, Vollero L, Vadalà G, et al. Natural language processing in low back pain and spine diseases: a systematic review. *Front Surg*. (2022) 9:957085. doi: 10.3389/fsurg.2022.957085

26. Groot O, Ogink P, Oosterhoff J, Beam A. Natural language processing and its role in spine surgery: a narrative review of potentials and challenges. *Semin Spine Surg*. (2021) 33(1):100877. doi: 10.1016/j.semss.2021.100877

27. Alonso V, Santos JV, Pinto M, Ferreira J, Lema I, Lopes F, et al. Problems and barriers during the process of clinical coding: a focus group study of Coders' perceptions. *J Med Syst*. (2020) 44(3):62. doi: 10.1007/s10916-020-1532-x

28. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. *J Public Health (Oxf)*. (2012) 34(1):138–48. doi: 10.1093/pubmed/fdr054

29. Shost MD, Meade SM, Steinmetz MP, Mroz TE, Habboub G. Surgical classification using natural language processing of informed consent forms in spine surgery. *Neurosurg Focus*. (2023) 54(6):E10. doi: 10.3171/2023.3.FOCUS2371

30. Krebs B, Nataraj A, McCabe E, Clark S, Sufiyan Z, Yamamoto SS, et al. Developing a triage predictive model for access to a spinal surgeon using clinical variables and natural language processing of radiology reports. *Eur Spine J*. (2023) 32(1):181–9. doi: 10.1007/s00586-023-07552-4

31. Mueller KB, Garrett CT, Kane S, Sandhu FA, Voyadzis JM. Incidental durotomy following surgery for degenerative lumbar disease and the impact of minimally invasive surgical technique on the rate and need for surgical revision: a case series. *Oper Neurosurg (Hagerstown)*. (2021) 21(5):351–5. doi: 10.1093/ons/opab282

32. Albayar A, Spadola M, Blue R, Saylany A, Dagli MM, Santangelo G, et al. Incidental durotomy repair in lumbar spine surgery: institutional experience and review of literature. *Global Spine J*. (2022):21925682221141370. doi: 10.1177/21925682221141368. [Epub ahead of print]

33. Desai A, Ball PA, Bekelis K, Lurie J, Mirza SK, Tosteson TD, et al. SPORT: does incidental durotomy affect longterm outcomes in cases of spinal stenosis? *Neurosurgery*. (2015) 76 Suppl 1(01):S57–63; discussion S63. doi: 10.1227/01.neu.0000462078.58454.f4

34. Reier L, Fowler JB, Arshad M, Siddiqi J. Drains in spine surgery for degenerative disc diseases: a literature review to determine its usage. *Cureus*. (2022) 14(3):312–23. doi: 10.7759/cureus.23129

35. Mirzai H, Eminoglu M, Orguc S. Are drains useful for lumbar disc surgery? A prospective, randomized clinical study. *J Spinal Disord Tech*. (2006) 19(3):171–7. doi: 10.1097/01.bsd.0000190560.20872.a7

36. Ho C, Sucato DJ, Richards BS. Risk factors for the development of delayed infections following posterior spinal fusion and instrumentation in adolescent idiopathic scoliosis patients. *Spine (Phila Pa 1976)*. (2007) 32(20):2272–7. doi: 10.1097/BRS.0b013e31814b1c0b

37. Adogwa O, Elsamadicy AA, Sergesketter AR, Shammas RL, Vatsia S, Vuong VD, et al. Post-operative drain use in patients undergoing decompression and fusion: incidence of complications and symptomatic hematoma. *J Spine Surg*. (2018) 4 (2):220–6. doi: 10.21037/jss.2018.05.09

38. Brown MD, Brookfield KFW. A randomized study of closed wound suction drainage for extensive lumbar spine surgery. *Spine (Phila Pa 1976)*. (2004) 29 (10):1066–8. doi: 10.1097/00007632-200405150-00003

39. Choi HS, Lee SG, Kim WK, Son S, Jeong TS. Is surgical drain useful for lumbar disc surgery? *Korean J Spine*. (2016) 13(1):20–3. doi: 10.14245/kjs.2016.13.1.20

40. Walid MS, Abbara M, Tolaymat A, Davis JR, Waits KD, Robinson JS, et al. The role of drains in lumbar spine fusion. *World Neurosurg*. (2012) 77(3–4):564–8. doi: 10.1016/j.wneu.2011.05.058

41. Zaman S, Mohamedahmed AYY, Peterknecht E, Zakaria RM, Mohamedahmed SYY, Hajibandeh S, et al. Sutures versus clips for skin closure following caesarean section: a systematic review, meta-analysis and trial sequential analysis of randomised controlled trials. *Langenbecks Arch Surg*. (2022) 407(1):37–50. doi: 10.1007/s00423-021-02239-0

42. Mostofi K, Peyravi M, Shirbacheh A, Shirbache K. A comparison between different suture techniques in lumbar spine surgery. *Int Wound J*. (2023) 20 (2):296–301. doi: 10.1111/iwj.13875

43. Meiring L, Cilliers K, Barry R, Nel CJ. A comparison of a disposable skin stapler and nylon sutures for wound closure. *S Afr Med J*. (1982) 62(11):371–2.

44. Orlinsky M, Goldberg RM, Chan L, Puertos A, Slajer HL. Cost analysis of stapling versus suturing for skin closure. *Am J Emerg Med*. (1995) 13(1):77–81. doi: 10.1016/0735-6757(95)90248-1

45. Rizan C, Lillywhite R, Reed M, Bhutta MF. The carbon footprint of products used in five common surgical operations: identifying contributing products and processes. *J R Soc Med*. (2023) 116(6):1410768231166135. doi: 10.1177/01410768231166135

Frontiers in Surgery

Explores and improves surgical practice and clinical patient management

A multidisciplinary journal which explores surgical practices - from fundamental principles to advances in microsurgery and minimally invasive techniques. It fosters innovation and improves the clinical management of patients.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact