frontiers | Frontiers in Neurology

TYPE Study Protocol
PUBLISHED 26 August 2022
DOI 10.3389/fneur.2022.933940

Profiling hearing aid users through big data explainable artificial intelligence techniques

Eleftheria Iliadou[1]*†, Qiqi Su[2]*†, Dimitrios Kikidis[1], Thanos Bibas[1]‡ and Christos Kloukinas[2]‡

[1]1st Department of Otorhinolaryngology-Head and Neck Surgery, National and Kapodistrian University of Athens Medical School, Athens, Greece, [2]Department of Computer Science, University of London, London, United Kingdom

Debilitating hearing loss (HL) affects ∼6% of the human population. Only 20% of the people in need of a hearing assistive device will eventually seek and acquire one. The number of people that are satisfied with their Hearing Aids (HAids) and continue using them in the long term is even lower. Understanding the personal, behavioral, environmental, or other factors that correlate with the optimal HAid fitting and with users' experience of HAids is a significant step in improving patient satisfaction and quality of life, while reducing societal and financial burden. In SMART BEAR we are addressing this need by making use of the capacity of modern HAids to provide dynamic logging of their operation and by combining this information with a big amount of information about the medical, environmental, and social context of each HAid user. We are studying hearing rehabilitation through a 12-month continuous monitoring of HL patients, collecting data, such as participants' demographics, audiometric and medical data, their cognitive and mental status, their habits, and preferences, through a set of medical devices and wearables, as well as through face-to-face and remote clinical assessments and fitting/fine-tuning sessions. Descriptive, AI-based analysis and assessment of the relationships between heterogeneous data and HL-related parameters will help clinical researchers to better understand the overall health profiles of HL patients, and to identify patterns or relations that may be proven essential for future clinical trials. In addition, the future state and behavioral (e.g., HAids Satisfiability and HAids usage) of the patients will be predicted with time-dependent machine learning models to assist the clinical researchers to decide on the nature of the interventions. Explainable Artificial Intelligence (XAI) techniques will be leveraged to better understand the factors that play a significant role in the success of a hearing rehabilitation program, constructing patient profiles. This paper is a conceptual one aiming to describe the upcoming data collection process and proposed framework for providing a comprehensive profile for patients with HL in the context of EU-funded SMART BEAR project. Such patient profiles can be invaluable in HL treatment as they can help to identify the characteristics making patients more prone to drop out and stop using their HAids, using their HAids sufficiently long during the day, and being more satisfied by their HAids experience. They can also help decrease the number of needed remote sessions with their Audiologist for counseling, and/or HAids

Check for updates

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Pedro A. Moreno-Sanchez,
Tampere University, Finland
Mareike Buhl,
Carl von Ossietzky Universität Oldenburg, Germany

*CORRESPONDENCE
Eleftheria Iliadou
iliadoue@med.uoa.gr
Qiqi Su
qiqi.su@city.ac.uk

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

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 01 May 2022
ACCEPTED 08 August 2022
PUBLISHED 26 August 2022

CITATION
Iliadou E, Su Q, Kikidis D, Bibas T and Kloukinas C (2022) Profiling hearing aid users through big data explainable artificial intelligence techniques. *Front. Neurol.* 13:933940.
doi: 10.3389/fneur.2022.933940

fine tuning, or the number of manual changes of HAids program (as indication of poor sound quality and bad adaptation of HAids configuration to patients' real needs and daily challenges), leading to reduced healthcare cost.

KEYWORDS

explainable AI (XAI), Deep Learning, big data, hearing loss, Hearing Aids, prognosis prediction, Long Short-Term Memory (LSTM), attention mechanism

Introduction

Hearing Loss (HL) is a public health problem that affects one out of three people over the age of 65, while debilitating HL is estimated to affect 6% of the population (466 million people) according to World Health Organization (WHO) statistics[1]. As per the same statistics, its annual management cost is estimated at more than 555 billion Euros (1) for the European countries and at 750 billion Dollars globally. HL should not be considered as an isolated health problem. Apart from the associated financial cost, HL severely affects communication and is associated with various comorbidities. Multiple studies have suggested that hearing impairment is associated with psychological and physical illness, such as cognitive disorders and dementia. An increase in the hearing threshold of 25 decibels (dB) corresponds to a loss of 7 cognitive years (2), and is associated with increased anxiety and depression (3), and even higher mortality rate (4). On the other hand, adults with hearing impairment tend to isolate themselves by limiting their participation in social events (5), thereby reducing their quality of life significantly (6).

Although the only available and validated management solution that currently exists for HL is the fitting and use of hearing assistive devices, only one in five people in need of a Hearing Aid (HAid) will eventually seek, acquire, and continue to use one efficiently (7, 8). A “HAid experience” refers to the process of living with a HAid and involves all the real-life challenges, coping strategies, and facilitations that the uses of HAid may evoke. Improvements in the HAid experience can lead to minimization of drop-out risk and enhancement of the overall quality of life (9).

The key factors in improving the HAid experience include, but are not limited to, proper fitting, affordability and accessibility of the follow-up services, and their combination with thorough and evidence-based personalized counseling and training on how to use the selected HAid (10). Since everyday patient needs and HL degree are not static and might change over time, there are still many factors that audiologists find challenging to address, including selecting optimal HAid configurations or best counseling approach according to individual patient profile and lifestyle (7, 11–13). Dynamic monitoring and collecting information about a patient's hearing and cognitive capacity, as well as their ability to control settings in real time in order to cope in different sound environments, could be very helpful toward this direction (14, 15). The development and validation of prediction models using the collected information and making accurate prognoses of how each patient's HAid experience will unfold are of major priority.

The use of Artificial Intelligence (AI) models in prognosis studies has gained traction increasingly in recent years due to its ability to handle large amounts of messy data (16), to learn from different types of data (17), and to facilitate clinical management of patients (18). Researchers have incorporated AI models in prognosis in clinical cancer research, such as breast cancer with Support Vector Machine (SVM) (19), colorectal cancer with Long Short-Term Memory (LSTM) (20), and glioblastoma with Prognosis Enhanced Neural Network (PENN) (21). As well as the prognosis for adult congenital heart disease with Convolutional Neural Network (CNN)-LSTM (22), rate of kidney disease with an ensemble of Logistic Regression, Decision Tree, Random Forest (RF), and K-Nearest Neighbor (KNN) (23), and COVID-19 with a segmentation network (24).

The effectiveness of AI models in HL prognosis has also been investigated by many researchers. Sensorineural Hearing Loss (SNHL) is the most common form of permanent HL resulting from the damage to the auditory nerve and/or the hair cells in the inner ear. Abdollahi et al. (25) constructed eight Machine Learning (ML) models to predict SNHL after chemoradiotherapy, including Decision Stump, Hoeffding, C4.5, Bayesian Network, Naïve, Adaptive Boosting (AdaBoost), Bootstrap Aggregating, Classification *via* Regression, and Logistic Regression (LR). The average predictive power of all models was found to be more than 70% in terms of accuracy, precision, and Area Under Curve (AUC). Idiopathic Sensorineural Hearing Loss (ISSHL) is characterized by an acute dysfunction of the inner ear. Zhao et al. (26) developed several ML models for ISSHL prediction, including SVM, Multilayer Perceptron (MLP), RF, and AdaBoost. A similarly high level of accuracy is also reported and varies between 78.6 and 80.1%. Bing et al. (27) evaluated several Deep Learning (DL)

[1] https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

and ML models to predict the dichotomised hearing outcome of ISSHL in order to identify the best predictive model for clinical application. Six input feature collections derived from 149 potential predictors have been used with Deep Belief Network, LR, SVM, and MLP. Best predictive performance was achieved by Deep Belief Network when evaluated with accuracy, precision, recall, F-score, Receiver Operating Characteristic Curve (ROC), and AUC, achieving 77.58% of accuracy and 0.84 of AUC. Ototoxic-induced HL, more specifically, the ototoxic effects in participants who were exposed to cigarette smoke and/or pesticides were evaluated by Artificial Neural Network, KNN, and SVM (28). While all models showed a good performance during training, KNN achieved the highest training accuracy with about 90% in two of the five datasets.

Attention-based DL models have also gained popularity in the medical domain recently. Bahdanau et al. (29) proposed the first attention mechanism, also known as the Soft Attention, for a Neural Machine Translation task using LSTM. The advantage of using attention mechanisms with LSTM is that it prevents the LSTM from forgetting certain input features when analyzing long-term dependencies and from putting too much weight on certain input features. Despite the lack of research using attention-based LSTM for HL patients specifically, a similar approach has been adapted for other comorbidities. Park et al. (30) used a Frequency-aware Attention-based LSTM (FA-Attn-LSTM) to investigate medical features that can be considered as critical for predicting the risk of cardiovascular disease. Wall et al. (31) proposed a framework for audio classification, specifically for chronic and non-chronic lung disease and COVID-19 diagnosis, with attention-based bidirectional LSTM (A-BiLSTM).

AI, particularly DL models, in general are appreciated for their ability to achieve high prediction accuracy. However, for sensitive domains, such as health care, accuracy is not the only determining factor (32). The inherent limitation of many AI systems is their black box nature, which means that humans are unable to easily understand the inner workings of these systems or how they arrive at their conclusions. Thus, automated decision-making systems that employ AI models are not widely accepted (32) due to a lack of trust from the end users. The integration of AI models into medical domains also faces criticisms where the models may fail to adhere to high standards of accountability, reliability, and transparency for medical decisions (33). It also complicates the issue of accountability in the event of a wrong decision (34).

Explainable AI (XAI) aims to overcome these limitations by explaining the learned decisions of AI models, thus giving end-users the ability to trust the models (35) and understanding why the models made certain decisions (32). Different XAI methods have been proposed over the years, particularly in the fields of computer vision and natural language processing. Yet very few studies have explored the potential applications of XAI methods to the medical field (34), especially in prognosis studies. A number of researchers have adapted Local Interpretable Model-agnostic Explanation (LIME) (36) to explain a CNN-based diagnostic model, including chronic wound classification (37), gastral image classification (38), and Alzheimer's diagnosis (39). Gu et al. (40) proposed an auxiliary decision system for breast cancer diagnosis and prediction with Extreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP) (41). Chakraborty et al. (17) developed a similar framework that was inspired by Gu et al. (40) using XGBoost and SHAP for prognosis in breast cancer patients. In the HL domain, Lenatti et al. (42) applied SHAP to explain the classification results of RF in predicting whether or not a patient has HL. In particular, SHAP is used to investigate the local predictions for each of the two output classes in four scenarios: true positive, true negative, false positive, and false negative. They have found that Age is the most important feature that impacts the classifier. In particular, values of age equal to 74 contribute positively to the model correctly predicting participants with HL (true positive), whereas values of age equal to 25 contribute negatively to the model correctly predicting participants without HL (true negative).

To the best of our knowledge, this is the first conceptual paper on a framework that leverages AI and XAI for prognosis for HL benefit and usage. ML techniques have been implemented previously in studies focusing on the prognosis of SNHL, ISSHL, and HL induced by ototoxic drugs and other substances (25–28), and modeling has also been attempted with synthetic data in more progressive types of HL, such as age-related or noise-induced HL (43). Nevertheless, we are unaware of any such attempts with real multi-source big data to date.

In the EU-funded SMART BEAR project[2], we are developing and validating a prognosis framework to address this scientific gap for HL patients. AI and XAI techniques will help identify and explain particular trends and factors in the large amount of heterogeneous data collected that correlate with the success or failure of hearing rehabilitation. In particular, the proposed framework composes the predictive power of LSTM with Attention Mechanism with the explanatory abilities of SHAP, and it will be used to answer several questions to provide a comprehensive profiling of HL patients.

The purpose of this article is to describe the planned data collection process, as well as the upcoming analyses to identify and explain particular trends and factors that correlate with the success or failure of hearing rehabilitation: drop-out of HAids usage, more hours of HAids usage and higher benefit from it, and less frequent need for manual adjustments or fine tuning of the HAids. As this is a conceptual paper, data collection is expected to begin in autumn 2022, followed by the experiments of the proposed methods.

[2] https://www.smart-bear.eu/

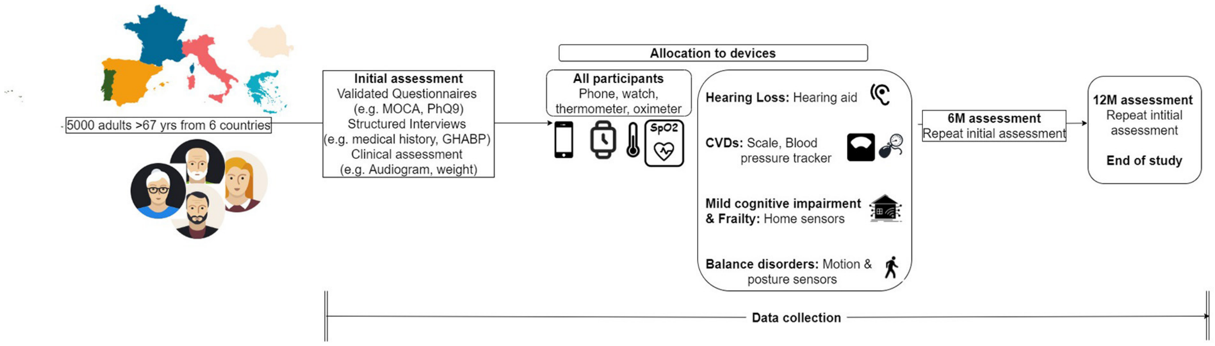
5000 adults >67 yrs from 6 countries
Initial assessment
Validated Questionnaires
(e.g. MOCA, PhQ9)
Structured Interviews
(e.g. medical history, GHABP)
Clinical assessment
(e.g. Audiogram, weight)
Allocation to devices
All participants
Phone, watch, thermometer, oximeter
SpO2
Hearing Loss: Hearing aid
CVDs: Scale, Blood pressure tracker
Mild cognitive impairment & Frailty: Home sensors
Balance disorders: Motion & posture sensors
6M assessment
Repeat initial assessment
12M assessment
Repeat intitial assessment
End of study
Data collection

FIGURE 1
Participants' flow of action.

Materials and methods

Participants

Five thousand elderly participants from six different EU countries will be included in the study. In particular, these six countries are divided into five study groups and 1,000 participants are recruited from each, namely France, Greece, Italy, Romania, and Portugal-Spain. A smaller-scale pilot study with 100 participants is already underway in the island of Madeira. The large-scale project is scheduled to begin in autumn 2022 and run for 24 months. Subjects will be included in the study based on the following eligibility criteria:

1. Age and birth gender: males and females, 67–80 years old.
2. Medical history: at least 2 of the following conditions: cardiovascular diseases (CVDs: hypertension, coronary disease, heart failure), hearing loss, balance disorders, mild depression, mild cognitive impairment, frailty.
3. Cognitive function according to MoCA score: participants with 26–30/30 (no cognitive impairment), and 18–26/30 (mild cognitive impairment) will be included (44). Score lower than 18/30 corresponds to mild dementia which is not addressed in SMART BEAR so those participants scoring $<$ 18/30 will be excluded.
4. Excellent to Moderate level of mobility, which corresponds to be able to perform simple tasks such as walking and jumping independently, with or without the help of a mechanical equipment, for example, a cane.
5. Ability to read.
6. Ability to use the basic functions of a smartphone (answer, call, check a notification, open an application).

Participants who meet the aforementioned criteria but present a severe or life-threatening condition, such as severe depression or high risk of heart failure, will be excluded from the study. All participants willing to provide their informed consent and voluntarily participate in the study will undergo an initial clinical assessment as shown in Figure 1. According to the results of this screening assessment, a specific set of devices and clinical procedures will be allocated to each participant. These devices are being obtained through joint procurement for all six countries and will be the same in terms of type, model, and configuration for all participants.

Participants with hearing loss

We intend to recruit one thousand people with HL to a degree that requires amplification. Participants with a moderate to severe unilateral or bilateral HL, as indicated by their pure tone audiogram, are considered eligible for HAid fitting if their HL negatively impacts their communication ability, cannot be treated surgically, or can be treated but the surgery is contra-indicated for the particular participant. Participants will only be excluded from Fitting if they do not wish to be fitted with a HAid, or if they have profound HL (Pure tone average 0.5–4 kHz $>$ 80 dB), and have not received any benefit from recent previous HAid fitting and use.

Audiological assessment

The same audiometric assessment (Figure 2) will be conducted on all participants with suspected or diagnosed HL by experienced personnel who have undergone additional internal training on every procedure of the clinical protocol by the clinical coordination team of the SMART BEAR. Joint procurement will ensure that the equipment (including HAids) and relevant software will be the same for all countries. Following the audiometric assessment, all participants will be fitted with HAids according to the same fitting protocol. The exact fitting protocol will be defined once the specific model and manufacturer of the HAids is selected during the international procurement procedure as discussed above.

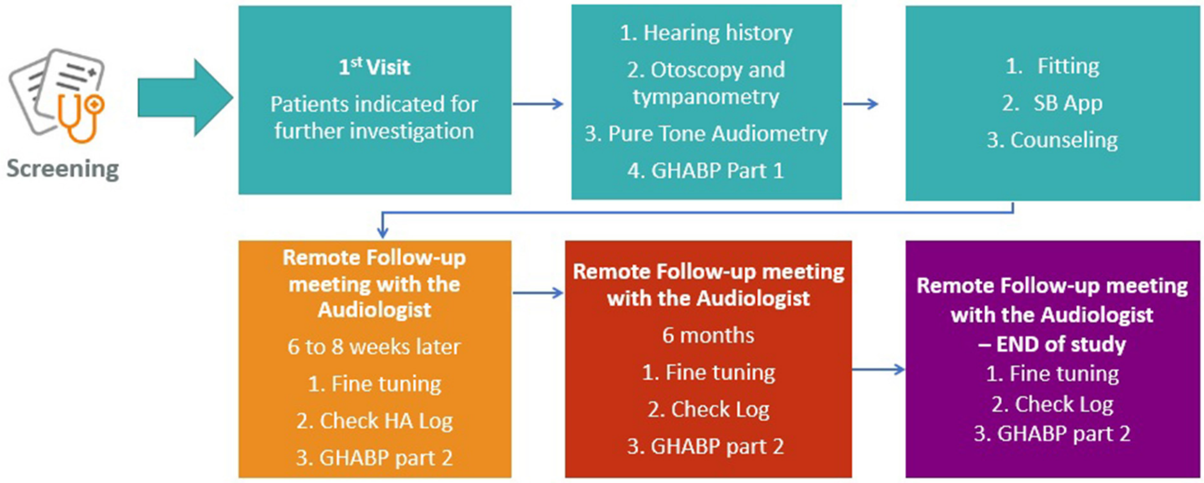

FIGURE 2
Audiological assessment flow of action.

HAids configuration will then be fine-tuned in accordance with the participant's experience level, listening preferences, and language preferences. There will be a predefined HAids program for all participants, other programs may be added based on the judgment of the audiologists and the needs of the participants. Pure tone audiometry will follow the British Society of Audiology[3] guidelines.

In accordance with the SMART BEAR fitting protocol, participants will be monitored for 12 months after they have been fitted with either one or two HAids (same manufacturer, same model). As shown in Figure 2, participants will also have continuous access to remote and face-to-face fine-tuning services provided by the SMART BEAR audiologists. Through the SMART BEAR clinician's dashboard, the audiologists will have access to participants' data and HAids log throughout this period.

Source of data

SMART BEAR is a large-scale multi-centric clinical study that aims to integrate state-of-the-art technology into everyday life of senior citizens with specific comorbidities, composing off-the-shelf and user-friendly devices onto an innovative platform. There are three subsystems in the SMART BEAR architecture as shown in Figure 3, namely the mobile phone application, the SMART BEAR HomeHub, and the SMART BEAR Cloud (SB@Cloud). Data are collected (i) during participants' clinical assessments *via* the clinician dashboard (e.g., anamnestic history, physiological and audiometric measurements), (ii) from all linked portable devices *via* the mobile phone application (e.g., HAid program, heart rate, and steps measurement), and (iii) through the mobile phone application itself (e.g., through questionnaires about their mood, diet, medication adherence and sleep quality). The HomeHub accumulates data from different home-based device sensors, such as weight scales and movement sensors. Finally, SB@Cloud securely stores and analyses the collected data through model and data-driven big data analytics during a 12-month period for each participant.

A total of 24 variable and covariates are collected through SMART BEAR HAids, including timestamp of the measurement, environmental noise, and manual program adjustments. Supplementary Table 1 provides a detailed description of each variable and covariate. Several other covariates are also being considered and are shown in Supplementary Table 2. The additional 241 covariates are collected in order to monitor the participants' other comorbidities based on their demographics, biological, environmental, and behavioral characteristics. There is a need to consider the impact of these additional covariates on the outcomes since they have been previously shown to affect to HL and HAid experiences, such as age, occupation, education, family history, mood disorders, cognitive function, diet, glucose levels and medication (3, 45–47). They are also currently being investigated for their correlation to hearing, as in the case of cardiovascular diseases, poorer mobility, frailty, and balance disorders (46, 48, 49). Furthermore, the medical and audiological assessment will also be supplemented by

[3] https://www.thebsa.org.uk/wp-content/uploads/2018/11/OD104-32-Recommended-Procedure-Pure-Tone-Audiometry-August-2018-FINAL.pdf

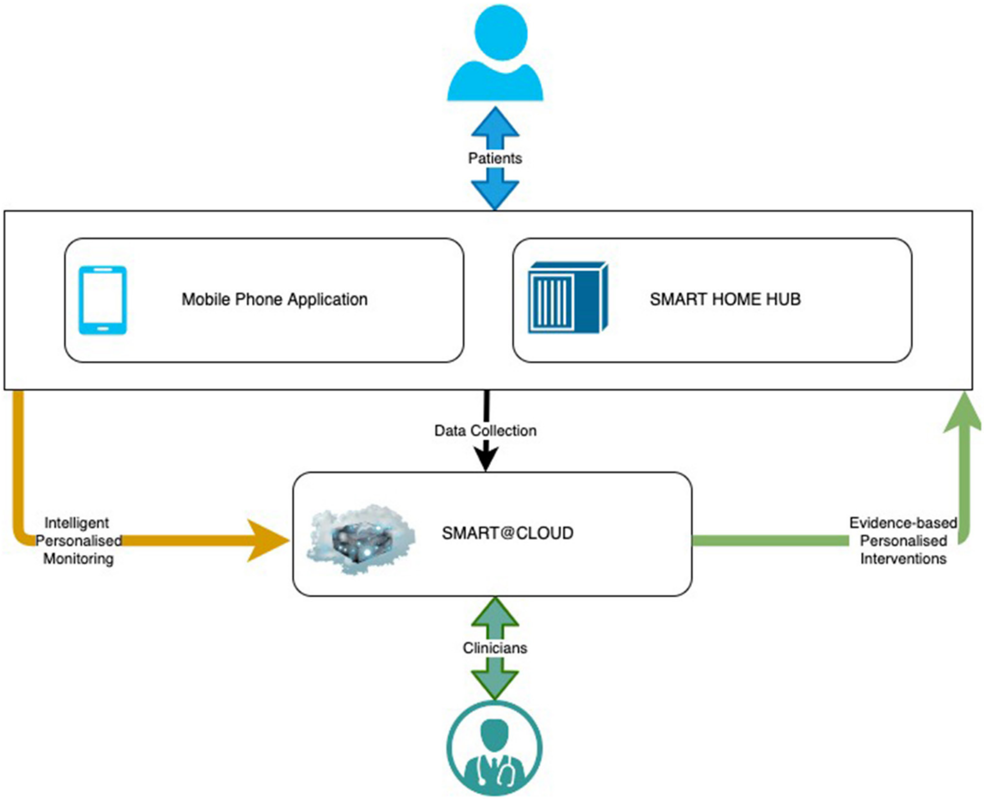

FIGURE 3
The SMART BEAR architecture.

additional sensor data as listed in Supplementary Table 3, such as blood pressure measured by the blood pressure tracker and physical activity measured by the smart watch. These variables are collected as a part of SMART BEAR's commitment to collect a wide range of data which will be explored as a part of data-driven analysis.

Sample size

SMART BEAR is aiming at collecting and analyzing big data—integrating information from many thousands of participants and different data sources. In Big Data, common sample size calculations cannot apply (50). Big data studies need to consider the marginal costs vs. the marginal value of possible sample sizes and include as many participants as possible (51). In SMART BEAR, the maximum number of participants that can be recruited based on available resources and time is 5,000. In accordance with the requirements of the study, this number is considered sufficient for ensuring the impact analysis obtained at the end of the project to be significant. In the case of HL, 200 participants with HL will be recruited from each of the five study groups, creating a sample of 1,000 participants with HL. These participants will then be fitted with either one or two HAids depending on whether one or both ears require amplification. Therefore, the total number of HAids to be used in the planned data collection is estimated between 1,000 and 2,000. The SMART BEAR platform is designed to facilitate the collection of data from a maximum number of 2,000 HAids, in case all participants suffer from bilateral HL. Data collected from up to 2,000 HAids are also considered to be sufficient based on previous experience (50).

Analysis methods

The questions that will be addressed with the proposed framework are based on future events. The prediction model will be used, for example, to predict future HAid usage or future drop-out rate. As a result, the model is fundamentally constructed with participants' historical medical history, HAid usage and habit, as well as the outcomes of medical and audiological assessments. As such, the collected SMART BEAR data are sequential in nature and can be viewed as time series data.

The proposed framework uses an attention-based LSTM (attn-LSTM) as the prediction model and then applies SHAP to interpret the model predictions. More specifically, SHAP is employed to identify those characteristics that influence the model predictions. To enable continuous learning and provision of personalized solutions, the pipeline for the proposed framework is to pre-process the data, hyper-tune the model, train/test the model with the optimal set of hyper-parameters selected from hyper-tuning, and then apply the XAI method. The performance of the prediction models is evaluated using different set of evaluation metrics for classification and regression problems.

Pre-processing the data

The temporal element of the collected data is determined by the Time variable, which records the date and time of the collected variables every 60 s when the SMART BEAR HAids are active in use. In SMART BEAR, clinicians also have the option of choosing how the data are aggregated for different analysis. Due to this, the data frequency is transformed first in order to allow hourly, daily, weekly, monthly, or yearly predictions, depending on the choice of clinician.

Transforming the distribution of the features allows the ML and DL algorithms to converge faster and minimize the weight of any variable with extreme values. *Standardization* and *normalization* are two pre-processing techniques that are particularly important for training an LSTM algorithm, since standardization on the data centers the noise from trend reverse signals and prevents activation functions to saturate (52), whereas normalization prevents the weights of the model being skewed (53).

Ordinal variables will be transformed with ordinal encoding and nominal variables will be transformed with one-hot encoding in order to convert these variables into either binary or multiple values with a numerical form. If the expected outcome variable is categorical then these will be treated label encoding.

Another important pre-processing step is to handle missing data. Several studies regarding data completeness in medical data were reviewed by Chan et al. (54) and found that the percentage of missing values of a variable, such as clinical status, laboratory results, and clinical actions or procedures, can reach as high as 98%. There is a possibility that this phenomenon might also be observed with data collected through SMART BEAR HAids due to connectivity issue and lack of participant adherence. As a result, simply deleting rows with missing values is not feasible for treating missing data, and imputation and model-based approaches should be used instead. There are several types of both imputation and model-based methods. For imputation methods, there are mean, median, zero, linear interpolation, forward, and backward, whereas for model-based methods, there are linear regression, KNN, and Multiple-value Imputation. A generic method was suggested by Salgado et al. (55) for the purpose of evaluating the performance of various methods for handling missing data. To start with, use a sample of the dataset that contains no missing data as ground truth, and then introduce the proportions of missing data at random in increments of say 5%. In the next step, compute the sum of squared errors (SSE) between the ground truth and the reconstructed data, for each method and for each proportion of missing data. Repeat these steps for each method and calculate the average SSE. Lastly, select the method that performed best at the level of missing data in the given dataset.

In addition, there is the question of how to deal with outliers—"samples that are exceptionally far from the mainstream data" (56). Even with a thorough understanding of the data, outliers can still be difficult to detect (56); however, statistical methods can assist in the identification of them. As standard deviation method is more suited for data with a normal distribution, therefore, it is used after the data have been standardized and normalized. Given the mean and standard deviation of the dataset, z-score can be computed for every ξ_i, which is the number of standard deviations away from the mean, as a way to identify outliers (57). Data points can be declared as outliers if their z-score standard deviation is greater than a predefined threshold. The threshold used in this analysis is three, as it is common practice to identify outliers in data with Gaussian or Gaussian-like distributions.

Lastly, it is important to determine whether there is multicollinearity among the variables. Multicollinearity refers to when there is a lack of orthogonality among two or more variables, and it often creates problems in a regression model (58) because the model results tend to fluctuate significantly when changes are made to independent variables that are highly correlated. In terms of hearing data, multicollinearity is often met among several variables. A typical example is the pure tone thresholds across different frequencies. Pure tone thresholds are measured in frequency bands with each representing a cochlear region, and the neighboring frequencies tend to be highly correlated (59). Moreover, pure tone audiogram also shows a high correlation among the sensitivity of the two ears for each participant when symmetric hearing is present (59). A common method of checking whether the data are multicollinear is to use the Variance Inflation Method (VIF) for each independent variable. In general, a VIF value of 10 indicates weak multicollinearity, and a variable with a higher value is typically considered to have a high correlation with another independent variable (58). A simple way to eliminate high multicollinearity variables is to remove them. However, this may not be feasible in practice. As a result, alternative methods, such as transforming the variables or performing Principal Component Analysis, should be considered instead, depending on the data and the expected outcome. Finally, data will be split into training, validation, and testing sets.

In this conceptual paper, the pre-processing steps discussed here are generic. While these techniques should be considered regardless of the questions to be answered, specific pre-processing methods, such as handling missing data and multicollinearity variables, will only become apparent following the data collection.

Hyper-tuning the model

The model is validated on the validation set during hyper-tuning in order to determine the set of optimal hyper-parameters. The hyper-tuning is performed using the Keras Tuner[4] library to determine the set of optimal hyper-parameters for model trained with TensorFlow[5]. There are many hyper-parameters that need to be determined when training an LSTM model. For this analysis, the number of hidden states in each layer, choice of activation function, learning rate, dropout rate, and batch size are hyper-tuned.

It is imperative to adjust the number of hidden units according to the complexity of the data and select an activation function that is capable of learning the complex relationship in the data. Learning rate is also important because if it is too fast, the model converges too quickly, while if it is too slow, it reaches some local minima. Dropout is a regularization technique while training a DL model, aiming at improving generalization and reducing overfitting. Last but not least, the batch size is the number of samples of training data that will be propagated through the model and should be adjusted accordingly as it impacts the stability of the learning process. Furthermore, the model will also be trained with early stopping in order to prevent overfitting. Early stopping is implemented through a callback function, which monitors the progress of the training, and if no improvements are made during the course of training, the training is terminated early.

Proposed model architecture

The proposed prediction model, attn-LSTM, will be trained on the training set with the set of optimal hyper-parameters from hyper-tuning, and the results are reported by predicting the unseen testing set. Table 1 shows the proposed model architecture of attn-LSTM and hyper-parameters setting for each layer. It should note that the choice of learning rate and batch size is hyper-tuned for the entire model and not for each individual layer.

LSTM (60) is a refined variant of the Recurrent Neural Network that is designed with a feedback architecture such that the current time step prediction is influenced by the network activation from the previous time steps as inputs. LSTM is one of the widely used DL technique for analyzing time series data and is capable of learning long-term time series data as well as short-term time series data (61). The hidden layer inside an LSTM network contains recurrently connected special units called memory cells and their corresponding gate units: input gate, forget gate, and output gate (60) as shown in Figure 4. The input gate is responsible for preventing the memory stored in a memory cell from perturbations by irrelevant inputs. Similarly, the output gate is there so other units are protected from perturbations by currently irrelevant stored memory. To optimize the performance of the LSTM, information that is no longer required by the LSTM is removed in the mechanism of the forget gate.

TABLE 1 Proposed model architecture.

Layer no.	Layer description	Hyper-parameters setting
1	Input layer	N/A
2	LSTM layer	Hidden units are hyper-tuned between 32 and 512. Activation function is hyper-tuned between Sigmoid and Tanh.
3	Self-attention layer	N/A
4	Dropout layer	Dropout rate is hyper-tuned between 0.001 and 0.1.
5	Flatten layer	N/A
6	Output (dense) layer	Regression problems: hidden unit is 1, and activation function is hyper-tuned between ReLu, Sigmoid, and None. Binary classification problem: hidden unit is 2, and activation function is Softmax and Sigmoid.

At each timestep t, the cell takes an input vector, x_t, and produces an output vector, h_t, which also refers to the hidden state of the LSTM. Firstly, the cell needs to determine whether the information from the previous timestep, $t-1$, should be kept or not with the forget gate, f_t. The forget gate takes the input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t*-1}, and produces an output between 0 and 1 where 0 represents "completely forget this information" and 1 represents "completely keep this information". The forget gate, f_t, is calculated as follows:

$$f_t = \sigma(w_x\, x_t + w_h\, h_{t-1} + b)\,,$$

where σ is the sigmoid function, w_x, w_h are the weighting factor, and b is the bias vector. More specifically, the sigmoid function is calculated as:

$$\sigma\,(\mathrm{x}) = \frac{1}{1+\mathrm{e}^{-\mathrm{x}}}.$$

[4] https://keras.io/keras_tuner/

[5] https://www.tensorflow.org/

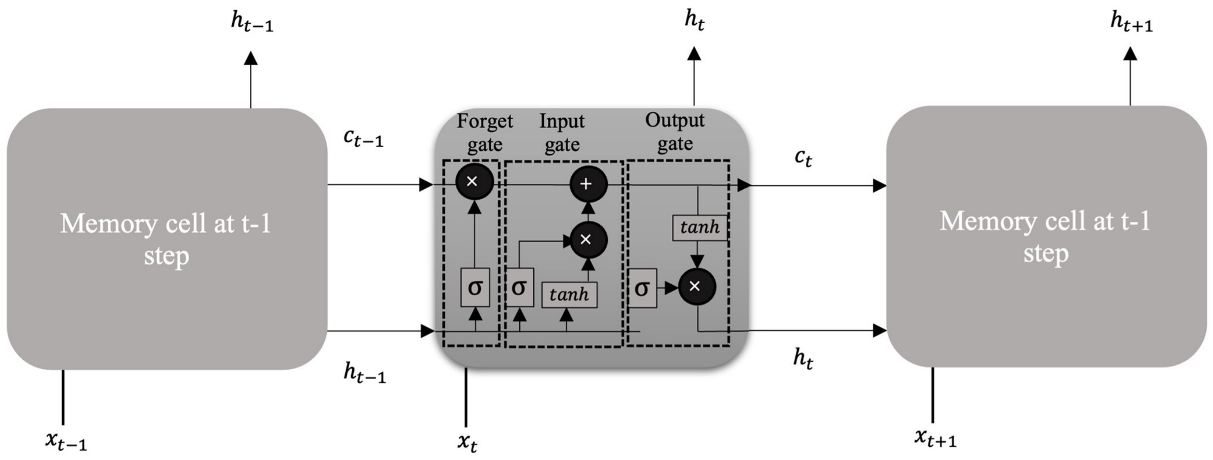

FIGURE 4
An illustration of the LSTM network.

The next step is to quantify the importance of the new information with the input gate, i_t:

$$i_t = \sigma(w_x\, x_t + w_h\, h_{t-1} + b),$$

which is also a function of input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t-1}. Then, a new vector named s_t is created which decides if the new information should be stored in the cell state or not. This is done by applying a hyperbolic tangent function, tanh, to the input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t-1}. It is calculated as:

$$s_t = \tanh(w_x\, x_t + w_h\, h_{t-1} + b)\,,$$

and the value of new information is transformed to a value between-1 and 1, where-1 means the new information is subtracted from the cell state and 1 means the new information is added to the cell state. The current cell state, c_t, is finally updated by taking the previous cell state, c_{t-1}, the forget gate, f_t, the input gate, i_t, and s_t into consideration by:

$$c_t = f_t \odot c_{t-1} \; + i_t \odot s_t\,,$$

where $\odot$ is the element-wise product. Then, the output gate, o_t, determines what information from the cell state is going to be the output. The output gate is also a function of input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t-1}, and outputs a value between 0 and 1. It is calculated as follows:

$$o_t = \sigma(w_x\, x_t + w_h\, h_{t-1} + b)\,.$$

Finally, the hidden state, h_t, at timestep t is updated with the current cell state, c_t, and the output gate, o_t, by:

$$h_t = \tanh(c_t) \odot o_t\,.$$

The use of attention-based LSTM was initially designed for natural language processing tasks and has been extended to other areas such as computer vision and time series prediction. The attention mechanism is also inspired by the human biological system, such that humans do not process large amounts of data all at once, but instead selectively focus on certain distinct parts of information (62). Moreover, integrating an attention mechanism into an LSTM model architecture may also enhance the interpretability of the model (63), since the attention mechanism can be used to demonstrate which features are important for predicting a particular outcome. The specific attention mechanism adopted in this framework is the Self-attention similar to the one proposed by Vaswani et al. (64), where the mechanism is relating different positions of a single sequence in order to gain a representation of the sequence.

Vaswani et al. (64) introduced a generalized definition for attention functions in which the inputs of the function consist of three vectors: queries (q), keys (k), and values (v). In practice, the attention function is computed on a set of queries simultaneously and packed into the matrix Q, and similarly the keys and values are packed into the matrix K and V, respectively. The concepts of Q, K, and V were first introduced in the context of NLP, specifically with Encoder-Decoder models. Taking the task of machine translation as an example, the query is derived from the Decoder layers reading the current translated text, whereas the key and value are derived from the Encoder layers reading the original sentence.

However, Self-attention is a special case of the attention mechanism where all of the queries, keys, and values come from the same place, such that $Q = K = V$ (64). The mechanism queries only the inputs to obtain the self-attention, and from the self-attention a new representation of the inputs can be constructed. In this framework, the inputs of the attention function are the sequence of hidden state vectors for all timesteps produced by LSTM, $H = \left(h_1, h_2, \ldots, h_n\right)$, therefore, $H = Q = K = V$.

The next step is to calculate a compatibility score for each hidden state vector in the LSTM. More specifically, it involves scoring the compatibility of each hidden state vector in H against the hidden state vector for which the self-attention is calculated. The specific compatibility score used in this framework is similar to the proposed by Vaswani et al. (64) and calculated as follows[6]:

$$\textit{Compatibility score} = \frac{HH^{\top}}{\sqrt{d_H}},$$

where d_H is the dimension of the sequence of hidden state vectors and it is a dot-product-based compatibility score. For example, the compatibility score of the first hidden state vector, h_1, is calculated by scoring each hidden state vector, $h_2, \ldots, h_n$, against h_1, with $h_1 \cdot h_1^{\top} / \sqrt{d_H}$, $h_1 \cdot h_2^{\top} / \sqrt{d_H}$, …, $h_1 \cdot h_n^{\top} / \sqrt{d_H}$. The other commonly used compatibility score is the additive-based one, where the compatibility score is computed using a single hidden layer feed-forward network. Dot-product-based compatibility scores can be space-efficient and much faster in practice when compared to additive-based compatibility scores (64).

Each compatibility score for each hidden state vector is then sent through to the Softmax function in order to normalize the scores so that all scores are positive and sum to 1. Finally, the output of the self-attention function is calculated as a weighted sum of the hidden state vectors and the compatibility score. The matrix of the output is calculated as follows[7]:

$$\textit{Attention}\,(H) = \textit{softmax}\left(\frac{HH^{\top}}{\sqrt{d_H}}\right) H.$$

Evaluating the model performance

The results of the trained attn-LSTM are reported by predicting the unseen testing set and evaluated using different sets of metrics for classification and regression problems. For classification problems, the evaluation metrics are accuracy, precision, recall, F1 score, and AUC. Accuracy, precision, and recall can be derived from a confusion matrix, and F1 score is the harmonic mean of precision and recall. Each of the metric is calculated as follows:

$$\begin{aligned}
\textit{Accuracy} &= \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}},\\
\textit{Precision} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},\\
\textit{Recall} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},\\
\text{F1 score} &= 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}.
\end{aligned}$$

Finally, AUC measures the area under the ROC curve, which is a graphical representation of how well the model performed and shows the relationship between True Positive Rate and False Positive Rate.

For regression problems, four standard error estimators are used, namely Symmetric Mean Absolute Percentage Error (sMAPE), Mean Absolute Scaled Error (MASE), Mean Absolute Percentage Error (MAPE), and Weighted Average Percentage Error (WAPE). The error estimators are calculated as follows:

$$\begin{aligned}
\textit{sMAPE} &= \frac{200}{N}\sum_{t=1}^{N}\frac{\left|y_i - \tilde{y}_i\right|}{\left|y_i\right| + \left|\tilde{y}_i\right|},\\
\textit{MASE} &= \frac{1}{N}\sum_{t=1}^{N}\frac{\left|y_i - \tilde{y}_i\right|}{\frac{1}{t+N-1}\sum_{j=2}^{t+N}\left|y_j - y_{j-1}\right|},\\
\textit{MAPE} &= \frac{1}{N}\sum_{t=1}^{N}\frac{\left|y_i - \tilde{y}_i\right|}{y_i},\\
\textit{WAPE} &= \frac{\sum_{i=1}^{N}\left|y_i - \tilde{y}_i\right|}{\sum_{i=1}^{N}\left|y_i\right|},
\end{aligned}$$

where y_i is the true value, $\tilde{y}_i$ is the predicted value, and N is the number of data points.

Since sMAPE, MASE, and MAPE are percentage-based error estimators, they are scaled-independent so that they can also be used for comparing prediction performance across different datasets. In addition, all error estimators are symmetric, which means that both positive and negative prediction errors are penalized equally. However, MAPE has the disadvantage that the errors tend to blow-up when the variable values are low, causing the results to be misleading. Thus, WAPE is also applied here since the errors are weighted by the total values.

Explaining the model

SHAP (41), more specifically, Kernel SHAP, is a local, *post-hoc*, and model-agnostic XAI method that can be used for both classification and regression problems. *Post-hoc* interpretation means that the interpretability is created after the model has been constructed (32) and aims to provide an explanation for the black-box models (65). Another method is ante-hoc, in which the decision-making process or the basis of a technique of a model can be understood by humans without additional information (65). Some of the ante-hoc methods

[6] The original notation for the generalized compatibility score in Vaswani et al. (64) is $\frac{QK^{\top}}{\sqrt{d_k}}$.

[7] The original notation for the generalized output of the attention function in Vaswani et al. (64) is $\textit{Attention}\,(Q,\ K,\ V) = \textit{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right) V$.

include LR, Decision Tree, and KNN. Both ante-hoc and *post-hoc* methods can be further divided into two approaches, *Model (Global) Explanation* and *Instance (Local) Explanation.* The Local Explanation approach explains only the model prediction for the single data instance, whereas the Global Explanation approach explains the inner workings of the entire model trained on a dataset. Model-agnostic is a subcategory of *post-hoc* methods, such that it can be applied to a variety of models, whereas model-specific can only be applied to one specific type of model.

SHAP uses the Shapley value from Game Theory to assign importance to each feature. In effect, the feature contributions (Shapley values) are calculated by the marginal contribution of the feature over every feature so that how the model behaves in its absence is analyzed, and then the prediction of the model can be written as the sum of bias and single feature contributions (41). According to Lundberg et al. (79), SHAP belongs to the family of *Additive Feature Attribution Methods,* meaning that the Shapley values are applied to binarised features, where a value of 0 corresponds to an unknown feature value, and a value of 1 corresponds to a feature being observed. The explanation model can be written mathematically as:

$$g\left(z^{'}\right) = \phi_0 + \sum_{i=1}^{M} \phi_i z_i^{'} ,$$

where g is the explanation model of the prediction model, $z^{'} \in \{0, 1\}^M$ where $z'$is the binarised feature and M is the number of binarised input features, ϕ_0 is the model output without binarised inputs, and $\phi_i \in R$ are the Shapley values (41). When compared with the other state-of-the-art explanation approach, LIME (36), SHAP satisfies three crucial properties that LIME does not: Local Accuracy, Missingness, and Consistency (41). Local accuracy requires consistency between the outputs of the explanation model and the prediction model. Missingness requires features missing in the original input to have no impact on the output. Lastly, consistency ensures that the impact of a feature does not decrease as it increases or remains the same.

Local accuracy is particularly important for providing explanations, as it ensures that the explanation model is less susceptible to adversarial attacks (66). Adversarial attacks refer to when the outputs of a classifier can be manipulated by a small perturbation of an input to conceal the biases of a system. In the study of Slack et al. (67), the authors attempted to fool both LIME and SHAP in order to determine if the feature contributions can be manipulated through the use of biased classifiers. It was found that the SHAP is less vulnerable to adversarial attacks than LIME due its local accuracy property. It is for these reasons that SHAP was chosen over LIME in our framework.

SHAP is a local XAI method that has been used to explain local predictions in many studies. For instance, Lenatti et al. (42) investigated the contribution of specific feature values to an individual prediction based on SHAP values. It is nevertheless also possible to obtain a global SHAP explanation by calculating the mean absolute SHAP values for each feature across the datasets allowing the global importance of each feature and the relative impact of all features over the entire dataset to be determined.

The results of SHAP will therefore be presented in the form of a visualization, in particular, the summary plots[8] will be used where it combines the feature importance with feature effects. The x-axis of the plots represents the SHAP value, or the impact on the model prediction, of each feature, the y-axis lists all the features and ordered according to their importance, and the color depicts the value of the feature from low to high.

In addition to the summary plots proposed to be used here, SHAP values can be analyzed in a variety of ways, including a dependence plot to demonstrate the global interaction effects between features. SHAP values may also be useful for assessing the contribution of features to an incorrect prediction, as demonstrated in the work of Lenatti et al. (42).

Expected outcome and predictors

The objectives of the SMART BEAR project in relations to HL are to answer several questions using the collected SMART BEAR data and the proposed predictive framework that leverages XAI techniques in order to develop a comprehensive profiling of patients with HL. Table 2 summarizes the expected outcome and its associated predictors (characteristics) for each question, and how this framework is applied to each question is discussed below.

As mentioned previously, this is a conceptual paper meaning that the precise details of the pre-processing techniques, optimal hyper-parameters for each question, and the prediction and explanation results will only be available once the study is commenced in autumn 2022.

Q1—Identification of those characteristics that make patients more prone to drop-out and stop using their HAids

The optimal drop-out rate should be less than the general population with HL (7), therefore, the expected outcome for Q1 is to be <45–50% for aged populations. Clinicians have the option of choosing how the data are aggregated in order to determine what the drop-out rate will be in the future in days, weeks, months, or years. In cases where a weekly analysis is required, for example, the average of HL chronicity, degree of HL, and manual adjustments of volume/program, and the sum of time of HAids usage are calculated for each week to

[8] https://shap-lrjball.readthedocs.io/en/latest/generated/shap.summary_plot.html

TABLE 2 A description of the predictive models, their expected outcome, and associated predictors.

Prediction models (PM)	Predictors	Outcome variables	Expected outcome	Value type
Q1	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, manual adjustments of volume/program, overall HAids satisfaction, time, time of hearing aids usage	Dropout	<45–50%	Y/N
Q2	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, time	Time of HAid usage	Adults should use their HAids >10 h a day.	Minutes/day
Q3	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, number of visits, manual adjustments of volume/program, time	GHABP score	Described in detail below.	(Integer)
Q4	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, overall HAids satisfaction, manual adjustments of volume/program, time, time of hearing aids usage	Number of face-to-face sessions	<4 visits to the Audiologist's in the first 6 months.	(Integer)
Q5	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, overall HAids satisfaction, manual adjustments of volume/program, time, time of hearing aids usage	Number of remote sessions	<4 visits to the Audiologist's in the first 6 months.	(Integer)
Q6	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, noise exposure, overall HAids satisfaction, time, time of hearing aids usage	Number of manual changes per day	<3 per day.	(Integer)

convert the data frequency. Apart from handling missing data, outliers, and multicollinearity among the variables, continuous variables such as age, degree of HL, and time of HAids usage are standardized and normalized, nominal variables such as gender are one-hot encoded, and ordinal variables such as HL chronicity, HL type, and manual adjustment of volume/program are ordinal encoded. In addition, the outcome variable is also treated with label encoding, with 1 representing Yes and 0 representing No, for making a binary classification.

Attn-LSTM is then employed to predict whether or not a participant will stop using their HAids in the future and the identification of characteristics that have an impact on this prediction is carried out through SHAP. Finally, the predicted future number of drop-out participants is compared to the general population with HL in order to compute the drop-out rate.

Q2—Identification of those characteristics that make patients more prone to use their HAids sufficiently long during the day

It is recommended that adults should use their HAids for more than 10 hours a day (76). Due to this, data are aggregated to have a daily frequency by default. This is done by taking the average of HL chronicity, degree of HL, manual adjustments of volume/program, and overall HAids satisfaction for each day, and the sum of time of HAids usage for each day in minutes. It should note that, although the data are transformed to have a daily frequency by default, clinicians will still have the option to choose to analyse monthly HAid usage, for example, if required. Similarly to Q1, continuous variables are standardized and normalized, while nominal and ordinal variables are one-hot and ordinal encoded, respectively. Missing data, outliers, and multicollinearity will also be treated with appropriate pre-processing techniques.

As a regression problem, attn-LSTM is used to predict participants' future HAids usage. SHAP is then used to interpret the model prediction to identify which characteristics influence participants to use their HAids more often.

Q3—Identification of those factors augmenting the benefit of patients from using their HAid

The Glasgow Hearing-Aid Benefit Profile (GHABP)[9] is a questionnaire that was designed to assess the operational management for HAid benefit, both at the systematic and clinical levels (15). The questionnaire will assess 4 situations with 6 questions, which are scored with 1 being the best score and 5 being the worst score. Whitmer et al. (77) recruited 1,574 participants and were asked to rate their hearing disability, handicap, HAid use, HAid benefit, HAid satisfaction, and residual (aided) disability with the GHABP questionnaire. The participants were divided into none, unilateral, and bilateral aided users and assessed in the four situations: quiet conversations, TV listening, noisy conversations, and group

[9] https://www.hey.nhs.uk/wp/wp-content/uploads/2020/09/HEY1167-2020-GHABP.pdf

conversations. Their findings regarding the normative GHABP score for HAid benefit will be used as the expected outcome for Q3.

Q3 is also a regression problem as the future GHABP score is predicted with attn-LSTM, and the reasons for this prediction are provided by SHAP. When clinicians require a monthly analysis, for example, the average of the GHABP score, HL chronicity, degree of HL, number of visits, and manual adjustments of volume/program, and the sum of time of HAids usage are calculated for each month to convert the data frequency. For Q3, pre-processing steps are similar to those used for previous questions, where continuous variables such as age, degree of HL, and time of HAids usage are standardized and normalized, nominal variable such as gender are one-hot encoded, and ordinal variables such as GHABP score, HL chronicity, HL type, number of visits, and manual adjustments of volume/program are ordinal encoded.

Q4—Identification of those factors decreasing the number of needed face-to-face sessions with their audiologist for counseling and/or HAid fine tuning, as an indicator of better self-management and optimal initial HAid configuration

The number of face-to-face with the audiologists is suggested to be <4 times in the first 6 months (78). Following this, the data are transformed to have a monthly frequency by default, with the options of analyzing the data at other frequencies still available. Therefore, the average of HL chronicity, degree of HL, number of visits, overall HAids satisfaction, and manual adjustments of volume/program, and the sum of time of HAids usage are calculated for each month. Nominal variables such as gender are one-hot encoded, ordinal variables such as overall HAids satisfaction, HL chronicity, HL type, number of visits, and manual adjustments of volume/program are ordinal encoded, and continuous variables such as age, degree of HL, and time of HAids usage are standardized and normalized.

As a regression problem, the future number of face-to-face sessions is predicted using attn-LSTM, and the characteristics affecting the prediction are investigated with SHAP.

Q5—Identification of those factors decreasing the number of needed remote sessions with their audiologist for counseling and/or HAid fine tuning, as an indicator of better self-management and optimal initial HAid configuration

Similar with Q4, the suggested number of remote sessions with the audiologists is also to be <4 times in the first 6 months (Tecca, 2018). Therefore, the default frequency is also set to be monthly, and attn-LSTM is used to predict the number of remote sessions with the audiologists in future months. SHAP is then used to identify the characteristics that influence participants to request fewer sessions with their audiologist. The pre-processing steps are also in line with Q4.

Q6—Identification of those factors decreasing the number of manual changes of HAid program, as indication of poor sound quality and bad adaptation of hearing aid configuration to patients' real needs and daily challenges

Although there is no precise definition for the optimal number of manual adjustments of the HAids, clinical experience has shown that fewer than three manual changes per day is considered as acceptable. By default, data are transformed to have a daily frequency in order to predict future daily manual adjustments with attn-LSTM, with SHAP providing information on the characteristics that impact the prediction.

It is also possible for clinicians to select a different data frequency for this analysis if required. The average of HL chronicity, degree of HL, number of visits, overall HAids satisfaction, and manual adjustments of volume and program, and the sum of time of HAids usage are calculated for each day to convert the data frequency. Pre-processing steps also consists of handling missing data, outliers, multicollinearity. As well as transforming continuous variables with standardization and normalization, ordinal variables with ordinal encoding, and nominal variables with one-hot encoding.

As a final point, SHAP values are analyzed with the same principle for all questions. The y-axis on the SHAP summary plot would indicate the most important feature on average for attn-LSTM to predict a certain outcome. The x-axis, along with the color, would show the impact of each feature value on the model prediction. For example, the SHAP values for Q1 may indicate that perhaps Age is the most important feature on average for participants to stop using their HAids. More specifically, younger participants might be less likely to drop out, whereas perhaps participants with a lower HAids usage might be more likely to stop using their HAids. As for Q3, SHAP result might show that perhaps HL type influences future GHABP score the most on average, where participants with a mixed type of HL might be more likely to benefit from their HAids.

Results—Discussion

This paper is a conceptual paper that synthesizes previous work on prediction models in healthcare and audiology (20, 27, 30, 31), and further describes the design and methods of the Big Data research project SMART BEAR with which we are aiming to fill the identified knowledge gaps. To the best of our knowledge, SMART BEAR represents the first research initiative in hearing research aimed at integrating such large and heterogeneous datasets and analyzing them using AI and XAI methods.

According to Mellor et al. (12), many factors beyond the pure tone audiogram should be monitored and dynamically adapted in order to achieve optimal hearing rehabilitation. Prognostic prediction models using audiometric and other lifestyle or medical data may be helpful toward achieving this goal. Education level (68), cognitive performance (69), and performance on speech recognition tests (70) have previously been suggested as potential prognostic factors. Following this, a wide range of data is collected in SMART BEAR as shown in Supplementary materials 2, 3, such as demographics, audiometric data, cognitive status, mental status, habits, and biological gender. Taking advantage of the ability of modern HAids to record their dynamic operation will also enable a relatively low-cost collection of data, such as hours of HAid use, from a large population, while clinical assessment will provide insight into the clinical context of the collected data. Furthermore, instead of assessing patients in a laboratory environment, SMART BEAR is collecting data both at the office and in real life through clinical assessments and smart sensors.

The created and continuously updated data can then be viewed as sequences with temporal elements and contain high-dimensional clinical variables (63). Therefore, collected SMART BEAR data will be analyzed through time-dependent multivariate prediction models that are capable of handling both classification and regression problems while ensuring a high level of accuracy. The XAI method will then be applied in order to explain the model to clinicians so that they will be able to better understand how the model arrives at the predicted results. In this study, attention-based LSTM is proposed to be the prediction model and then using SHAP to interpret the model. The proposed framework introduced in this conceptual paper can also be applied to other comorbidities within the SMART BEAR project.

The findings of this analysis will have implications in clinical practice, health policies and research.

Clinical and research implications

With proper analysis and interpretation of SMART BEAR results, the most accurate patient profile to date can be created for HL patients, allowing it to serve as a valid proxy for anticipated behavior even before the initial HAid fitting session. According to the analysis of synthetic hearing data conducted within the context of the H2020 project EVOTION[10], higher levels of physical activity are associated with longer daily HAid use (43). Therefore, SMART BEAR results also aim to provide a better understanding how physical activity, such as walking, affects HAid experience in order to incorporate physical activity promotion into hearing rehabilitation for different populations. Furthermore, different factors relating to hearing rehabilitation might be identified with different participants. This is shown in the data-driven analysis with the subjective data of 572 HAid users conducted by Sanchez-Lopez et al. (71), where participants with different HL degree preferred different types of hearing rehabilitation. Other factors may include presence of particular comorbidities or different living situations, therefore, the combinations and interactions between the factors will also be examined in SMART BEAR.

The patient profiling proposed by SMART BEAR may be able to assist manufacturers and clinicians in making optimal choices in terms of HAid model and configuration options, or, in future stages, it could create automatic fine-tuning of HAids (12). In this context, after the end of the study, SMART BEAR is considering providing access, upon request, to the de-identified dataset for future exploration. Participants will be fully informed and will provide their consent so access to their de-identified data can be granted in the future for specific scientific purposes. Open Access will be provided for the following SMART BEAR datasets: anonymised data from demographics, questionnaires, interviews, anonymised sensor raw data, video of the protocols for annotation, and anonymised data from basic clinical information for annotation. It is envisaged that this policy will facilitate the use of SMART BEAR's gained knowledge by a range of different stakeholders.

Limitations

All participants in SMART BEAR will be fitted with the same HAid model, following the same fitting protocol, with the use of the same algorithm. Although the fine-tuning and the program selection of the HAids will be based on the needs and preferences of each participant, the fitting of the HAids may not be optimal for every participant when only one universal fitting protocol is used. However, this choice was made since the comparison of programs or algorithms is not in the scope of SMART BEAR, as well as in order to avoid unnecessary heterogeneity or lower quality of the data as a result of systematic errors. This limitation will be taken into account in the interpretation of our results. Moreover, SMART BEAR participants will only be between the ages of 67 and 80, which means that its results cannot be generalized to a population younger than that. Data like hours of usage and changes in programs will be subject to connectivity loss, which is a significant barrier in similar projects (50). The impact of loss of follow-up patients, such as the unavailability of information regarding continuation of usage, is also expected to be low, provided that this percentage will remain in the predicted range (below 20%). Close follow-ups and dedicated helpdesks

[10] https://h2020evotion.eu/

will help minimize these risks, while imputation and model-based approaches will facilitate dealing with missing data, as explained above. Another limitation will be the variation in the population between six different countries with socioeconomic and cultural diversities; however, comparison between study groups is expected to produce useful results. Finally, speech audiometry in quiet or in noise is not part of the SMART BEAR data collection. This is due to the fact that there do not currently exist any universally validated materials that could be used across all six countries and thus in all languages. Speech audiometry, while recognized as having clinical value in fitting choices, does not fall under the scope of SMART BEAR. As an alternative approach to assess HAid benefit, we are aiming to collect other parameters, including real-life data, such as hours of usage and manual changes of programs, as well as interview data, such as the GHABP questionnaire.

It is noteworthy that unlike the evaluation metrics used in this paper to evaluate a prediction model, there are currently no widely accepted objective metrics for evaluating XAI methods. Though the proposed XAI method will be validated by clinicians and medical experts in SMART BEAR, this will only provide a subjective assessment of the XAI method. To this end, existing evaluation metrics for XAI metrics, such as Rosenfield's set (72), should be tested in the future with the collected data in order to obtain both objective and subjective validation. Although SHAP is one of the best known XAI methods, it is often criticized for long computation time and Shapley values do not work if features are correlated (73). As a result, the proposed framework may be unable to deliver what clinicians require in cases where the characteristics to be identified are correlated. Therefore, alternative methods of XAI should be considered in the future. Among them is Attention Mechanism-based XAI methods, such as the one proposed by Choi et al. (74) and Schockaert et al. (75). An attention mechanism-based XAI method can provide an explanation for Recurrent Neural Network or its variants by assigning corresponding values to the importance of the different sub-sequence of the input sequence according to the model and may be more suitable for the proposed prediction model.

Conclusion

SMART BEAR is, to the best of our knowledge, the first big data study whose goal is to integrate heterogeneous and contextualized HAid, medical, societal, and environmental data in order to develop and validate a prognosis framework using AI and XAI methods. The outcomes of the project are expected to benefit multiple stakeholders in the field of Audiology, such as HAid users, manufacturers, clinicians, researchers, and health policy makers, as well as to influence current practice and future research. These outcomes could also improve confidence in integrating AI models in the medical field, particularly with encouraging AI to be used in the medical decision-making process by utilizing XAI methods to enhance its interpretability, transparency, and accountability.

Ethics statement

This is a conceptual paper describing the rationale and design of the large scale H2020 project SMART BEAR. The SMART BEAR protocols have obtained, or are in the process to obtain, ethical approval in all six countries. All participants will have to provide their voluntary consent after oral and written information about the details of the project. General Data Protection Regulation (EU) 2016/679 (GDPR) principles will be implemented in all stages of data collection, storage and sharing.

Author contributions

EI, QS, CK, and TB: conceptualization, methodology: writing—original draft, and writing—review and editing. DK: writing—original draft. CK and TB: supervision. All authors contributed to the article and approved the submitted version.

Funding

SMART BEAR was funded by the European Commission (Grant Agreement No.: 857172/H2020-SC1-FA-DTS-2018-2).

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/fneur.2022.933940/full#supplementary-material

References

1. Shield B. Evaluation of the social and economic costs of hearing impairment. *Hear-It AISBL*. (2006). p. 1–202. Available online at: https://www.hear-it.org/sites/default/files/multimedia/documents/Hear_It_Report_October_2006.pdf

2. Lin FR, Ferrucci L, Metter EJ, An Y, Zonderman AB, Resnick SM. Hearing loss and cognition in the Baltimore Longitudinal Study of Aging. *Neuropsychology.* (2011) 25:763–70. doi: 10.1037/a0024238

3. Dawes P, Emsley R, Cruickshanks KJ, Moore DR, Fortnum H, Edmondson-Jones M, et al. Hearing loss and cognition: the role of hearing aids, social isolation and depression. *PLoS ONE.* (2015) 10:e0119616. doi: 10.1371/journal.pone.0119616

4. Li L, Simonsick EM, Ferrucci L, Lin FR. Hearing loss and gait speed among older adults in the United States. *Gait Posture.* (2013) 38:25–9. doi: 10.1016/j.gaitpost.2012.10.006

5. Saunders JE, Rankin Z, Noonan KY. Otolaryngology and the global burden of disease. *Otolaryngol Clin North Am.* (2018) 51:515–34. doi: 10.1016/j.otc.2018.01.016

6. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet.* (2016) 388:1545–602. doi: 10.1016/S0140-6736(16)31678-6

7. McCormack A, Fortnum H. Why do people fitted with hearing aids not wear them? *Int J Audiol.* (2013) 52:360–8. doi: 10.3109/14992027.2013.769066

8. Saunders GH, Dillard LK, Zobay O, Cannon JB, Naylor G. Electronic health records as a platform for audiological research: data validity, patient characteristics, and hearing-aid use persistence among 731,213 U.S. veterans. *Ear Hear.* (2021) 42:927–40. doi: 10.1097/AUD.0000000000000980

9. Newman CW, Weinstein BE, Jacobson GP, Hug GA. The hearing handicap inventory for adults. *Ear Hear.* (1990) 11:430–3. doi: 10.1097/00003446-199012000-00004

10. Gatehouse S, Naylor G, Elberling C. Linear and nonlinear hearing aid fittings – 2. Patterns of candidature. *Int. J. Audiol.* (2006) 45:153–71. doi: 10.1080/14992020500429484

11. Dillon H. *Hearing Aids*, 2nd ed. New York, NY: Thieme Medical Publishers (2012).

12. Mellor J, Stone MA, Keane J. Application of data mining to a large hearing-aid manufacturer's dataset to identify possible benefits for clinicians, manufacturers, and users. *Trends Hearing.* (2018) 22:233121651877363. doi: 10.1177/2331216518773632

13. Timmer BHB, Hickson L, Launer S. Adults with mild hearing impairment: are we meeting the challenge? *Int J Audiol.* (2015) 54:786–95. doi: 10.3109/14992027.2015.1046504

14. Ferguson MA, Henshaw H. Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. *Front. Psychol.* (2015) 6:556. doi: 10.3389/fpsyg.2015.00556

15. Gatehouse S. Glasgow hearing aid benefit profile: derivation and validation of a client-centered outcome measure for hearing aid services. *J Am Acad Audiol.* (1999) 10:24. doi: 10.1055/s-0042-1748460

16. Wang L, Wang H, Song Y, Wang Q. MCPL-Based FT-LSTM: medical representation learning-based clinical prediction model for time series events. *IEEE Access.* (2019) 7:70253–64. doi: 10.1109/ACCESS.2019.2919683

17. Chakraborty D, Ivan C, Amero P, Khan M, Rodriguez-Aguayo C, Başagaoglu H, et al. Explainable artificial intelligence reveals novel insight into tumor microenvironment conditions linked with better prognosis in patients with breast cancer. *Cancers.* (2021) 13:3450. doi: 10.3390/cancers13143450

18. Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett.* (2020) 471:61–71. doi: 10.1016/j.canlet.2019.12.007

19. Ferroni P, Zanzotto F, Riondino S, Scarpato N, Guadagni F, Roselli M. Breast cancer prognosis using a machine learning approach. *Cancers.* (2019) 11:328. doi: 10.3390/cancers11030328

20. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep.* (2018) 8:3395. doi: 10.1038/s41598-018-21758-3

21. Vasudevan P, Murugesan T. Cancer subtype discovery using prognosis-enhanced neural network classifier in multigenomic data. *Technol Cancer Res Treat.* (2018) 17:153303381879050. doi: 10.1177/1533033818790509

22. Diller GP, Kempny A, Babu-Narayan SV, Henrichs M, Brida M, Uebing A, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. *Eur. Heart J.* (2019) 40, 1069–1077. doi: 10.1093/eurheartj/ehy915

23. Javed Mehedi Shamrat FM, Ghosh P, Sadek MH, Kazi MdA, Shultana S. Implementation of machine learning algorithms to detect the prognosis rate of kidney disease. In: *2020 IEEE International Conference for Innovation in Technology (INOCON).* (2020). p. 1–7.

24. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell.* (2020) 181:1423–33.e11. doi: 10.1016/j.cell.2020.04.045

25. Abdollahi H, Mostafaei S, Cheraghi S, Shiri I, Rabi Mahdavi S, Kazemnejad A. Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck cancer patients: a machine learning and multi-variable modelling study. *Physica Medica.* (2018) 45:192–7. doi: 10.1016/j.ejmp.2017.10.008

26. Zhao Y, Li J, Zhang M, Lu Y, Xie H, Tian Y, et al. Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise: a pilot study. *Ear Hear.* (2019) 40:690–9. doi: 10.1097/AUD.0000000000000649

27. Bing D, Ying J, Miao J, Lan L, Wang D, Zhao L, et al. Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models. *Clin. Otolaryngol.* (2018) 43:868–74. doi: 10.1111/coa.13068

28. Tomiazzi JS, Pereira DR, Judai MA, Antunes PA, Favareto APA. Performance of machine-learning algorithms to pattern recognition and classification of hearing impairment in Brazilian farmers exposed to pesticide and/or cigarette smoke. *Environ Sci Pollut Res.* (2019) 26:6481–91. doi: 10.1007/s11356-018-04106-w

29. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *ArXiv*:1409.0473. (2014). doi: 10.48550/arXiv.1409.047363

30. Park HD, Han Y, Choi JH. Frequency-aware attention based LSTM networks for cardiovascular disease. In: *2018 International Conference on Information and Communication Technology Convergence (ICTC).* (2018). p. 1503–5.

31. Wall C, Zhang L, Yu Y, Mistry K. Deep recurrent neural networks with attention mechanisms for respiratory anomaly classification. In: *2021 International Joint Conference on Neural Networks (IJCNN).* (2021). p. 1–8.

32. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. *J Artif Int Res.* (2021) 70:245–317. doi: 10.1613/jair.1.12228

33. Anderson C. Ready for prime time?: AI influencing precision medicine but may not match the hype. *Clin OMICs.* (2018) 5:44–6. doi: 10.1089/clinomi.05.03.26

34. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. In: *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32 (2021). p. 4793–813.

35. Schlegel U, Arnout H, El-Assady M, Oelke D, Keim DA. Towards a rigorous evaluation of XAI methods on time series. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW).* (2019) 4197–201. doi: 10.1109/ICCVW.2019.00516

36. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?": explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016). 1135-1144.

37. Sarp S, Kuzlu M, Wilson E, Cali U, Guler O. The enlightening role of explainable artificial intelligence in chronic wound classification. *Electronics.* (2021) 10:1406. doi: 10.3390/electronics10121406

38. Malhi A, Kampik T, Pannu H, Madhikermi M, Framling K. Explaining machine learning-based classifications of in-vivo gastral images. In: *2019 Digital Image Computing: Techniques and Applications (DICTA).* (2019). p. 1–7.

39. Das D, Ito J, Kadowaki T, Tsuda K. An interpretable machine learning model for diagnosis of Alzheimer's disease. *PeerJ.* (2019) 7:e6543. doi: 10.7717/peerj.6543

40. Gu D, Su K, Zhao H. A case-based ensemble learning system for explainable breast cancer recurrence prediction. *Artif Intell Med.* (2020) 107:101858. doi: 10.1016/j.artmed.2020.101858

41. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems.* (2017). p. 4768–4777.

42. Lenatti M, Moreno-Sánchez PA, Polo EM, Mollura M, Barbieri R, Paglialonga A. Evaluation of machine learning algorithms and explainability techniques to detect hearing loss from a speech-in-noise screening test. *Am J Audiol.* (2022) 1–19. doi: 10.1044/2022_AJA-21-00194 [Epub ahead of print].

43. Saunders GH, Christensen JH, Gutenberg J, Pontoppidan NH, Smith A, Spanoudakis G, et al. Application of big data to support evidence-based public health policy decision-making for hearing. *Ear Hear.* (2020) 41:1057–63. doi: 10.1097/AUD.0000000000000850

44. Nasreddine ZS, Phillips NA, BÃ©dirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x

45. Carpenter MG, Campos JL. The effects of hearing loss on balance: a critical review. *Ear Hear.* (2020) 41 (Suppl. 1):107S—19S. doi: 10.1097/AUD.0000000000000929

46. Oishi N, Shinden S, Kanzaki S, Saito H, Inoue Y, Ogawa K. Influence of depressive symptoms, state anxiety, and pure-tone thresholds on the tinnitus handicap inventory in Japan. *Int J Audiol.* (2011) 50:491–5. doi: 10.3109/14992027.2011.560904

47. Samocha-Bonet D, Wu B, Ryugo DK. Diabetes mellitus and hearing loss: a review. *Ageing Res Rev.* (2021) 71:101423. doi: 10.1016/j.arr.2021.101423

48. Manson J, Alessio H, Cristell M, Hutchinson KM. Does cardiovascular health mediate hearing ability? *Med Sci Sports Exerc.* (1994) 26:866–71. doi: 10.1249/00005768-199407000-00009

49. Simões JFCPM, Vlaminck S, Seiça RMF, Acke F, Miguéis ACE. Cardiovascular risk and sudden sensorineural hearing loss: a systematic review and meta-analysis. (2022) *Laryngoscope*. doi: 10.1002/lary.30141 [Epub ahead of print].

50. Dritsakis G, Kikidis D, Koloutsou N, Murdin L, Bibas A, Ploumidou K, et al. Clinical validation of a public health policy-making platform for hearing loss (EVOTION): protocol for a big data study. *BMJ Open.* (2018) 8:e020978. doi: 10.1136/bmjopen-2017-020978

51. Nayak B. Understanding the relevance of sample size calculation. *Indian J Ophthalmol.* (2010) 58:469. doi: 10.4103/0301-4738.71673

52. Sethia A, Raut P. Application of LSTM, GRU and ICA for stock price prediction. In: *Information and Communication Technology for Intelligent Systems*. Singapore: Springer (2019). p. 479–487.

53. Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock price prediction using LSTM, RNN and CNN-sliding window model. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).* (2017). p. 1643–7.

54. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev.* (2010) 67:503–27. doi: 10.1177/1077558709359007

55. Salgado CM, Azevedo C, Proença H, Vieira SM. Missing data. In: *Secondary Analysis of Electronic Health Record*s, MIT Critical Data, editor (New York, NY: Springer International Publishing) (2016). p. 143–62.

56. Kuhn M, Johnson K. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Boca Raton, FL: CRC Press (2019).

57. Ilyas IF, Chu X. *Data Cleaning*. New York, NY: ACM (2019).

58. Alin A. Multicollinearity. *Wiley Interdiscip Rev Comput Stat.* (2010) 2:370–4. doi: 10.1002/wics.84

59. Coren S. Summarizing pure-tone hearing thresholds: the equipollence of components of the audiogram. *Bull Psychon Soc.* (1989) 27:42–4. doi: 10.3758/BF03329892

60. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

61. Preeti BR, Singh RP. Financial and non-stationary time series forecasting using LSTM recurrent neural network for short and long horizon. In: *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT).* (2019). p. 1–7.

62. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing.* (2021) 452:48–62. doi: 10.1016/j.neucom.2021.03.091

63. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. *Adv Neural Inf Process Syst.* (2016) 9:29. doi: 10.48550/arXiv.1608.05745

64. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in Neural Information Processing systems 30 (NIPS 2017)*. Long Beach, CA: Curran Associates (2017).

65. Zhang Y, Weng Y, Lund J. Applications of explainable artificial intelligence in diagnosis and surgery. *Diagnostics.* (2022) 12:237. doi: 10.3390/diagnostics12020237

66. Janizek JD, Sturmfels P, Lee S-I. Explaining explanations: axiomatic feature interactions for deep networks. *J Mach Learn Res.* (2021) 22:1–54.

67. Slack D, Hilgard S, Jia E, Singh S, Lakkaraju H. Fooling LIME and SHAP. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* (2020). p. 180–6.

68. Fuentes-López E, Fuente A, Valdivia G, Luna-Monsalve M. Does educational level predict hearing aid self-efficacy in experienced older adult hearing aid users from Latin America? Validation process of the Spanish version of the MARS-HA questionnaire. *PLoS ONE.* (2019) 14:e0226085. doi: 10.1371/journal.pone.0226085

69. Meister H, Rählmann S, Walger M, Margolf-Hackl S, Kießling J. Hearing aid fitting in older persons with hearing impairment: the influence of cognitive function, age, and hearing loss on hearing aid benefit. *Clin Interv Aging.* (2015) 10:435. doi: 10.2147/CIA.S77096

70. Davidson A, Marrone N, Wong B, Musiek F. Predicting hearing aid satisfaction in adults: a systematic review of speech-in-noise tests and other behavioral measures. *Ear Hear.* (2021) 42:1485–98. doi: 10.1097/AUD.0000000000001051

71. Sanchez-Lopez R, Dau T, Whitmer WM. Audiometric profiles and patterns of benefit: a data-driven analysis of subjective hearing difficulties and handicaps. *Int J Audiol.* (2022) 61:301–10. doi: 10.1080/14992027.2021.1905890

72. Rosenfeld A. Better metrics for evaluating explainable artificial intelligence. In: *20th International Foundation for Autonomous Agents and Multiagent Systems (AAMAS '21).* (2021), 45–50.

73. Molnar, C. (2020). *Interpretable Machine Learning*. Available online at: https://www.lulu.com/ (accessed July 02, 2022).

74. Choi KS, Choi SH, Jeong B. Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. *Neuro Oncol.* (2019) 21:1197–209. doi: 10.1093/neuonc/noz095

75. Schockaert C, Leperlier R, Moawad A. Attention mechanism for multivariate time series recurrent model interpretability applied to the ironmaking industry. *arXiv[Preprint].arXiv*:2007.12617 (2020).

76. Laplante-Lévesque A, Nielsen C, Jensen LD, Naylor G. Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport. *J Am Acad Audiol.* (2014) 25:187–98. doi: 10.3766/jaaa.25.2.7

77. Whitmer WM, Howell P, Akeroyd MA. Proposed norms for the glasgow hearing-aid benefit profile (Ghabp) questionnaire. *Int J Audiol.* (2014) 53:345–51. doi: 10.3109/14992027.2013.876110

78. Tecca JE. Are post-fitting follow-up visits not hearing aid best practices? *Hear. Rev.* (2018) 25:12–22.

79. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, CA: Curran Associates (2017). p. 4766–75.