Automated Detection of Candidate Subjects With Cerebral Microbleeds Using Machine Learning

Vaanathi Sundaresan[1,2], Christoph Arthofer[1,3,4], Giovanna Zamboni[1,5,6], Robert A. Dineen[3,4,7], Peter M. Rothwell[5], Stamatios N. Sotiropoulos[1,3,4], Dorothee P. Auer[3,4,7], Daniel J. Tozer[8], Hugh S. Markus[8], Karla L. Miller[1], Iulius Dragonu[9], Nikola Sprigg[10], Fidel Alfaro-Almagro[1], Mark Jenkinson[1,11,12†] and Ludovica Griffanti[1,13*†]

[1] Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, Oxford Centre for Functional MRI of the Brain, University of Oxford, Oxford, United Kingdom, [2] Oxford-Nottingham Centre for Doctoral Training in Biomedical Imaging, University of Oxford, Oxford, United Kingdom, [3] NIHR Nottingham Biomedical Research Centre, Queen's Medical Centre, University of Nottingham, Nottingham, United Kingdom, [4] Sir Peter Mansfield Imaging Centre, University of Nottingham, Nottingham, United Kingdom, [5] Wolfson Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom, [6] Dipartimento di Scienze Biomediche, Metaboliche e Neuroscienze, Università di Modena e Reggio Emilia, Modena, Italy, [7] Radiological Sciences, Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom, [8] Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom, [9] Siemens Healthcare Ltd., Research and Collaborations GB & I, Frimley, United Kingdom, [10] Stroke Trials Unit, Mental Health and Clinical Neuroscience, University of Nottingham, Nottingham, United Kingdom, [11] South Australian Health and Medical Research Institute (SAHMRI), Adelaide, SA, Australia, [12] Australian Institute for Machine Learning (AIML), School of Computer Science, The University of Adelaide, Adelaide, SA, Australia, [13] Department of Psychiatry, Wellcome Centre for Integrative Neuroimaging, Oxford Centre for Human Brain Activity, University of Oxford, Oxford, United Kingdom

Cerebral microbleeds (CMBs) appear as small, circular, well defined hypointense lesions of a few mm in size on T2*-weighted gradient recalled echo (T2*-GRE) images and appear enhanced on susceptibility weighted images (SWI). Due to their small size, contrast variations and other mimics (e.g., blood vessels), CMBs are highly challenging to detect automatically. In large datasets (e.g., the UK Biobank dataset), exhaustively labelling CMBs manually is difficult and time consuming. Hence it would be useful to preselect candidate CMB subjects in order to focus on those for manual labelling, which is essential for training and testing automated CMB detection tools on these datasets. In this work, we aim to detect CMB candidate subjects from a larger dataset, UK Biobank, using a machine learning-based, computationally light pipeline. For our evaluation, we used 3 different datasets, with different intensity characteristics, acquired with different scanners. They include the UK Biobank dataset and two clinical datasets with different pathological conditions. We developed and evaluated our pipelines on different types of images, consisting of SWI or GRE images. We also used the UK Biobank dataset to compare our approach with alternative CMB preselection methods using non-imaging factors and/or imaging data. Finally, we evaluated the pipeline's generalisability across datasets. Our method provided subject-level detection accuracy > 80% on all the datasets (within-dataset results), and showed good generalisability across datasets, providing a consistent accuracy of over 80%, even when evaluated across different modalities.

Keywords: cerebral microbleeds, machine learning, susceptibility weighted image (SWI), UK Biobank, subject-level detection, T2*-weighted MRI, structural MRI

1. INTRODUCTION

Cerebral microbleeds (CMBs) represent focal haemosiderin depositions consisting of macrophages in microhaemorraghes (Shoamanesh et al., 2011), and are sometimes surrounded by ischemic areas and gliosis (Gouw et al., 2011). CMBs appear as small focal dark circular lesions on T2*-weighted gradient recalled echo (T2*-GRE) sequences. They usually range from 2 to 10 mm in size, although further subdivision into microbleeds (2–5 mm) and macrobleeds (>5 mm) has been used (Greenberg et al., 2009a). While some of the CMBs might not be visible at all on T2*-GRE images, susceptibility weighted images (SWI) shows more CMBs and they appear more prominently on SWI due to the blooming effect (Greenberg et al., 2009b; Charidimou and Werring, 2011). CMBs can be the early sign of intracerebral haemorrhage (ICH) (Gouw et al., 2011), vascular dementia (Ayaz et al., 2010) and Alzheimer's disease (Gouw et al., 2011; Shoamanesh et al., 2011; Wardlaw et al., 2013). They have been associated with cognitive decline (Werring et al., 2004; Won Seo et al., 2007) and several vascular diseases (Nishikawa et al., 2009; Yates et al., 2014), including lacunar stroke and small vessel diseases (Nannoni et al., 2021). Moreover, CMBs often occur with vascular damage related to cerebral amyloid angiopathy (Gouw et al., 2011). Strong associations have been established between CMBs and risk factors such as age (Roob et al., 2000; Horita et al., 2003; Jeerakathil et al., 2004; Vernooij et al., 2008; Poels et al., 2010; Takashima et al., 2011), hypertension (Roob et al., 2000; Tsushima et al., 2002; Horita et al., 2003; Vernooij et al., 2008; Poels et al., 2010) and white matter damage (Roob et al., 2000).

Recent studies on the predictive value of CMBs for long-term cognitive outcome have shown inconsistent results, therefore the specific role of CMBs in cognitive impairment and neurodegeneration remains unclear (Wardlaw et al., 2013). Hence, it would be useful to observe the prevalence and clinical/demographic associations of CMBs in larger populations. However, exhaustive manual labelling of CMBs is difficult and time consuming, especially in large datasets (e.g., the UK Biobank dataset). Consequently, semi-automated methods (Barnes et al., 2011; Seghier et al., 2011; Kuijf et al., 2012, 2013; De Bresser et al., 2013; van den Heuvel et al., 2016; Morrison et al., 2018) were proposed as a possible solution to reduce false positives. Even though the manual revising step removes the spurious detections effectively, it typically takes at least a few minutes per subject (Kuijf et al., 2013; van den Heuvel et al., 2016; Morrison et al., 2018) and it is not an efficient solution for CMB detection in very large datasets. Hence, it would be highly useful and efficient to develop a fully automated CMB candidate subject preselection method (without involving manual intervention in any stage of the detection pipeline) in the large datasets, and focus on those subjects for manual labelling to facilitate further semi-automated or fully-automated methods with more accurate CMB detection, analysis and characterisation (based on size, shape, location and multiplicity, clustering).

The detection of CMBs, however, even at subject-level, is highly challenging since CMBs occur sparsely, are difficult to detect due to their size, contrast variations and the fact that they are often accompanied by other signs (e.g., haemorrhages).

The SWI modality has been shown to aid in identifying more CMBs (at least >67% Nandigam et al., 2009) compared to T2*-GRE images, since SWI improves CMB contrast. However, the presence of other paramagnetic substances, apart from haemosiderin, causes enhanced appearance of dark structures that resemble CMBs, known as CMB "mimics" (Greenberg et al., 2009b; Charidimou and Werring, 2011). The mimics could be haemorrhagic/paramagnetic such as cavernous malformations, haemorrhagic micrometastases, diffusion axonal injury, small haemorrhages nearer to the infarcts and ICH areas, or non-haemorrhagic such as flow voids, calcifications, motion artefacts, Gibb's ringing artefact and partial volume artefacts at air-bone interfaces (Greenberg et al., 2009b). For description and features of CMB mimics refer to **Supplementary Table 1**.

So far, the proposed automated methods for lesion-level CMB detection used shape descriptors (Bian et al., 2013; Fazlollahi et al., 2014, 2015), intensity and geometric information (Ghafaryasl et al., 2012), and location-based features (Dou et al., 2015). The use of comprehensive features, integrating the geometry, intensity, scale and local image structures from multiple modalities have been shown to improve CMB detection (Ghafaryasl et al., 2012; Dou et al., 2015). Over recent years, with the advent of deep learning, methods using convolution neural networks (CNNs) (Dou et al., 2016; Chen et al., 2018; Liu et al., 2019) and hybrid methods using a combination of CNNs and intensity information (Chen et al., 2015; Dou et al., 2015) have been proposed, occasionally with additional postprocessing steps (Liu et al., 2019). While CNN-based methods provide more accurate CMB detection when compared to conventional methods, they require a large amount of labelled training data. Alternatively, techniques such as semi-supervised and omni-supervised learning (Huang et al., 2018) require more representative labelled CMB instances that would not bias the method towards CMB mimics. While the occurrence of CMBs could be high in disease groups (e.g., small vessel disease), in general, the prevalence of CMBs is low in population-based cohorts (e.g., UK Biobank). In datasets from populations with low prevalence of CMBs, it would be extremely time consuming to look at all subjects manually. A way to automatically preselect a subset that was enriched for CMBs would allow better use of the available manual identification time and lead to better and clinically relevant training datasets for further CMB analyses.

In this work, our aim is to develop a subject-level preselection method that is computationally light, easy to train and scalable to large datasets. Towards this aim, we propose a fully automated method using intensity, shape and location-based features for detecting CMB candidate subjects from large datasets such as the UK Biobank. We evaluated the method on datasets with different image modalities (GRE and SWI) and in the presence of other pathological signs. We then applied the method to the UK Biobank dataset for CMB candidate subject preselection, and compared our method with various alternative methods using non-imaging factors and/or imaging data (CMB count). We finally evaluated the ability of our proposed pipeline to adapt to differences in image characteristics and demographics of the datasets, by training our pipeline on one dataset and testing it on a different one.

2. DATASET DETAILS

In this work we used the following datasets to develop and evaluate the preselection pipeline. The datasets are diverse in their intensity characteristics and are acquired using different protocols. A brief overview of the datasets is provided below.

The Oxford Vascular Study (OXVASC) dataset: The dataset consists of T2*-GRE images from 40 participants from the OXVASC study (Rothwell et al., 2004), who had recently experienced a minor non-disabling stroke or transient ischemic attack. The age range was 35.6–94.8 years, mean age 68.7 ± 15.5 years, median age 67.4 years, and female to male ratio F:M = 15:25. The 2D single-echo T2*-GRE images were acquired using a 3T Siemens Verio scanner with GRAPPA factor = 2, TR/TE = 504/15 ms, flip angle 20°, voxel resolution of $0.9 \times 0.8 \times 5$ mm, with image dimensions of $640 \times 640 \times 25$ voxels. Out of 40 subjects, 20 subjects have CMBs and the corresponding manual segmentations, labelled on T2*-GRE images, are available. The total number of CMBs is 267, mean: 13.3 ± 1.13 CMBs/subject. **Figure 1** shows the histogram of volumes of individual CMBs in the OXVASC dataset.

The Tranexamic acid for IntraCerebral Haemorrhage 2 (TICH2) trial MRI substudy dataset: This is a subset of the MRI dataset used in Dineen et al. (2018) obtained as part of the TICH2 trial (Dineen et al., 2018; Sprigg et al., 2018). The age range was 29–88 years, mean age 64.76 ± 15.5 years, median age 66.5 years, and female to male ratio F:M = 24:26. The dataset consists of images acquired at multiple centres and on multiple MRI platforms with variations in image dimensions and voxel resolutions. MR acquisition parameters for the TICH2 MRI substudy dataset can be found in Dineen et al. (2018). The dataset used in this work consists of 50 SWI images from subjects with spontaneous intracerebral haemorrhage. Out of 50 subjects, 25 subjects have CMBs and manual segmentations for CMBs are available for all 25 subjects. The manual segmentations were labelled on SWI images as either definite or possible according to the microbleed anatomical rating scale (MARS) (Gregoire et al., 2009). Total number of CMBs: 505, mean: 20 ± 32.6 CMBs/subject (the histogram of volumes of individual CMBs is shown in **Figure 1**).

UK Biobank (UKBB) dataset: Out of ≈ 20,000 subjects from the January 2018 release of UKBB, 14,521 had the required data fields (e.g., availability of SWI images and the factors specified below). From these subjects, we randomly selected 180 subjects with age range 46.8–76.8 years, mean age 58.9 ± 9.1 years, median age 58.8 years, and female to male ratio F:M = 86:94. We used SWI images from the selected subjects for our experiments, which were constructed from 3D multi-echo GRE images acquired using a 3T Siemens Magnetom Skyra syngo MR D13 scanner with TR/TE1/TE2 = 27/9.4/20 ms, flip angle 15°, voxel resolution of $0.8 \times 0.8 \times 3$ mm, with image dimensions of $256 \times 288 \times 48$ voxels. Subject-level manual labels were available for the selected subjects, indicating whether each subject was a CMB or a non-CMB subject (only including those graded as definite CMB based on MARS scale). However, lesion-level manual segmentations of CMBs were not available for the dataset. On the UKBB dataset, we also used

several non-imaging factors that are known to be associated with CMBs (given these data were collected as a part of the study), including demographic factors such as age (Greenberg et al., 2009b; Charidimou and Werring, 2011), blood pressure (BP) (Roob et al., 2000; Tsushima et al., 2002; Vernooij et al., 2008; Poels et al., 2010), risk factors such as smoking (Tsushima et al., 2002; Poels et al., 2010), clinical conditions such as white matter damage (Roob et al., 2000) and cognitive decline (Werring et al., 2004; Won Seo et al., 2007). We used the non-imaging factors for a comparison experiment on the dataset (for more details, refer to section 3.4.3). Summary statistics for the shortlisted factors for 180 subjects are: age (range provided above), BP (102/50–202/118 mmHg, mean: 134.7/80.7 ± 42.4/25.3 mmHg, median: 147.5/90.5 mmHg), smoking (smoking:non-smoking = 70:110), white matter hyperintensity (WMH) volume (535–45,186 mm^3, mean: 5,465 ± 4,349 mm^3, median: 2,929 mm^3) and mean reaction time (MRT, as an indicator of cognitive ability) (397–896 ms, mean: 563 ± 96.5 ms, median: 546 ms).

2.1. Data Preprocessing

We reoriented the T2*-GRE and SWI images to match only the axis labels of the standard MNI template by swapping axes as necessary using the fslreorient2std tool (similar to performing 0°, 90° or 180° rotations, but keeping all the original voxel intensities unchanged; e.g., a voxel size of $1 \times 2 \times 1$mm, with a matrix size of $192 \times 256 \times 192$, and axes of LR, IS, AP would become $1 \times 1 \times 2$mm, with $192 \times 192 \times 256$ matrix size and LR, PA, IS axes), in order to aid in further processing. We do not apply any rigid/affine/non-affine transformation of the images to register them to standard space, given the small size of CMBs. We skull stripped the images (T2*-GRE/SWI) using FSL BET (Smith, 2002), followed by bias field correction using FSL FAST (Zhang et al., 2001).
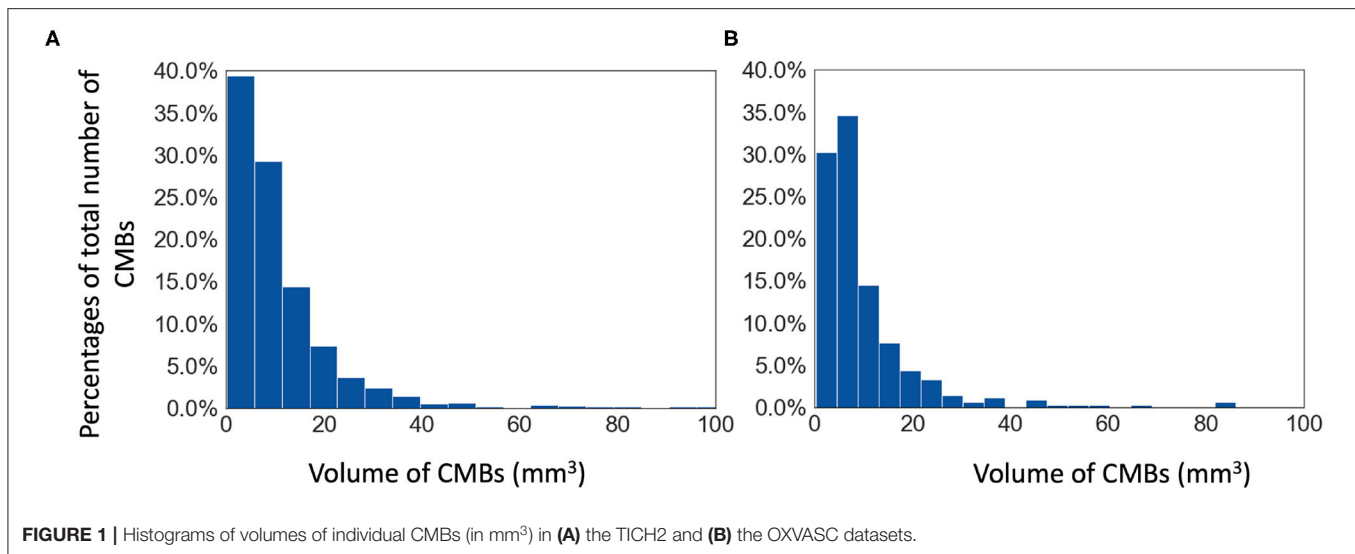
3. PIPELINE FOR DETECTION OF CMB CANDIDATE SUBJECTS

The automated pipeline for CMB candidate subject preselection takes T2*-GRE images or SWI as input and provides a subject-level decision on whether the subject has CMBs or not. The pipeline consists of three steps: 1. removal of blood vessels and sulci, 2. voxel-wise detection of initial CMB candidates, 3. filtering of initial candidates using shape-based attributes.

3.1. Removal of Blood Vessels and Sulci

In the first step, we removed the blood vessels, sulci and other elongated dark structures in the input image to reduce the prevalence of CMB mimics. **Figure 2** shows the extracted features and a few samples of images with blood vessels and sulci removed. We exploited the elongated tubular structure of vessels/sulci to extract the following edge and orientation-based features:

1. **Frangi filters:** Frangi filters (Frangi et al., 1998) use the multiscale second order local structure of the image and denote the degree of vessel-like/edge characteristics at each voxel. We extracted voxel-wise intensity values of Frangi

FIGURE 1 | Histograms of volumes of individual CMBs (in mm^3) in **(A)** the TICH2 and **(B)** the OXVASC datasets.

filter outputs as features (**Figure 2B**). While applying Frangi filters, we adjusted the parameters β_1 and β_2 (controlling the discrimination of lines from blob-like structures and the elimination of background noise, respectively) to avoid detection of CMBs during vessel detection. Based on the results on the training data, we empirically set the values of β_1 and β_2 to 0.9 and 20, respectively.

2. **The eigenvalues of the structure tensor:** The structure tensor is the covariance matrix, at each voxel, consisting of partial derivatives of the gradients (Förstner, 1994). The eigenvalues λ_1 and λ_2 of this matrix indicate the edge strength. We considered the combination of the principal eigenvalue λ_1 and the linearity measure $l = |\lambda_1 - -\lambda_2|/2$ (**Figure 2C**) as voxel-wise features.

We used K-means clustering, an unsupervised learning algorithm, with the above features to classify voxels into 2 classes (vessel vs. background). The detected voxels with vessel-like structures and sulci are used as masks for inpainting. We filled the masked regions with intensity values similar to those in the non-masked neighbourhood (mean of the nearest 3 voxels) to remove linear dark structures from the input image.

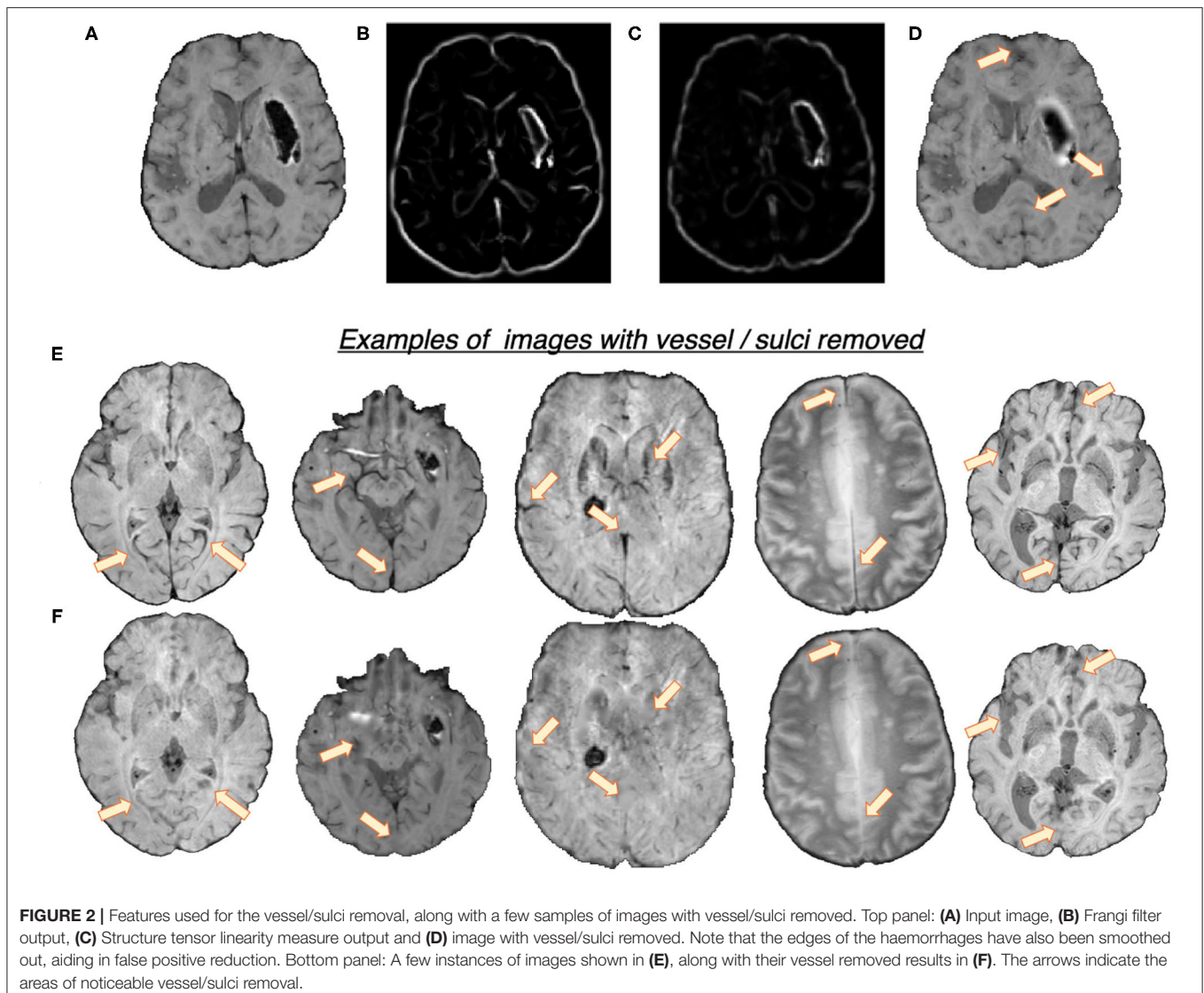## 3.2. Voxel-Wise Detection of Initial CMB Candidates

In the second step, as shown in **Figure 3**, we extracted the following 7 features at each voxel on the vessel-removed image for the initial CMB candidate detection:

1. **Intensity transformations (3 features):** At each voxel, we extracted the intensity value (**Figure 3A**). Additionally, to obtain contrast enhanced intensity features, we normalised the intensity values using standardisation (subtracting the mean intensity within the brain mask and dividing by the standard deviation), extracted the exponential of the intensity (exp(p × intensity), where p = 1) (**Figure 3B**) and applied "contrast limited adaptive histogram equalisation" (CLAHE, Zuiderveld

(1994), using *equalize_adapthist* in *scikit-image* package) with a clip limit of 0.01 on the input image to obtain the CLAHE output at each voxel (**Figure 3C**).

2. **Fast radial symmetry transform (FRST):** We performed FRST (Loy and Zelinsky, 2002). In FRST, at each voxel, the orientation (pointing towards or away from the voxel) and magnitude maps of gradients are calculated at a certain radius. These maps are then used to obtain a voxel-wise symmetry information within the radius. We used 4 different radii: 2, 3, 4 and 6 voxels, and calculated the mean value of all 4 outputs at each voxel as a feature (**Figure 3D**).

3. **The eigenvalues of structure tensor:** We considered the principal eigenvalue λ_1 (since this reduced the noise in background voxels) of the structure tensor at each voxel as a feature (**Figure 3E**).

4. **Gaussian filter:** We smoothed the vessel removed image I with a Gaussian filter with $\sigma = 1.5$ voxels (empirically determined to roughly match the size of CMBs) to get a smoothed image I_S. The difference $I_\delta = I - I_S$ removes the background and highlights the sharper objects and blobs, and hence can be interpreted as a "blobness" measure (**Figure 3F**).

5. **Laplacian of Gaussian filter (LoG):** Since CMBs have well-defined edges, we applied LoG (with $\sigma = 1.5$ voxels), a second order derivative filter for edge detection, on the vessel removed image and obtained the filtered output (**Figure 3G**) at each voxel.

We normalised the above features individually by dividing by their maximum value across the current image and use the normalised features for training a support vector machine (SVM) classifier at the voxel-level, using voxel-wise manual segmentations available for the OXVASC and TICH2 datasets, to obtain a probability map P_{CMB}. We then thresholded it at a global threshold th_{prob} of 0.8. The threshold value was determined empirically (using the training data from the OXVASC and TICH2 datasets), based on the cluster-wise performance metrics with respect to lesion-level manual segmentations, by varying
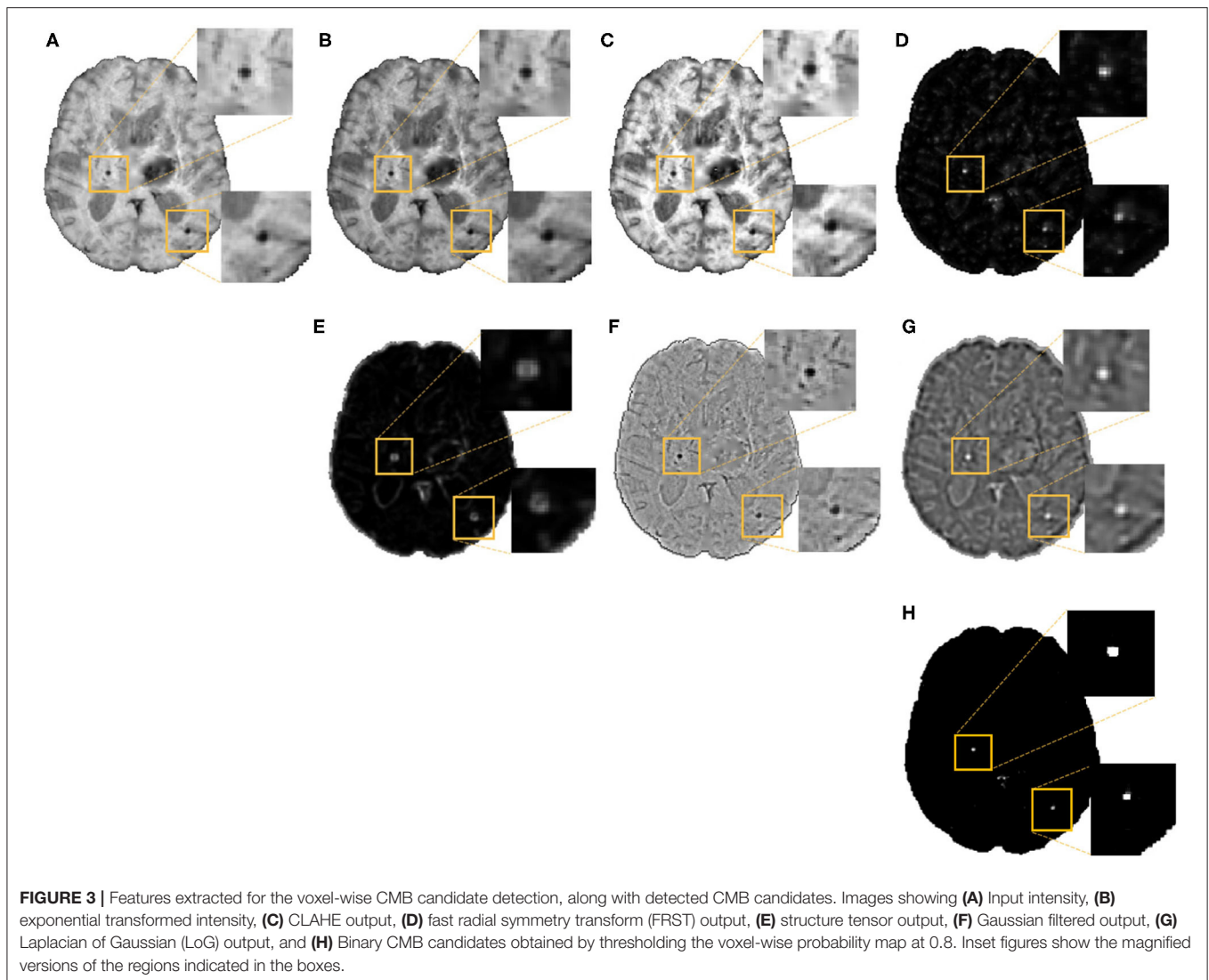
FIGURE 2 | Features used for the vessel/sulci removal, along with a few samples of images with vessel/sulci removed. Top panel: **(A)** Input image, **(B)** Frangi filter output, **(C)** Structure tensor linearity measure output and **(D)** image with vessel/sulci removed. Note that the edges of the haemorrhages have also been smoothed out, aiding in false positive reduction. Bottom panel: A few instances of images shown in **(E)**, along with their vessel removed results in **(F)**. The arrows indicate the areas of noticeable vessel/sulci removal.

th_{prob} within a range [0, 1], increasing in steps of 0.1. For this experiment, we trained the SVM model on the OXVASC dataset and evaluated it on the TICH2 data and vice versa, since evaluating across datasets would provide a more robust threshold value. For each threshold value, we determined the cluster-wise evaluation metrics (specified below in section 3.5), and selected the threshold that provided the best set of cluster-wise performance metrics as the final th_{prob} value.

3.3. Filtering of Initial Candidates Using Shape-Based Attributes

The small dark structures in the image other than CMBs (including noise and stray fragments of blood vessels) are detected as false positives (FPs) in the voxel-level classification. In this step, we used the following shape- and location-based object-level attributes for reducing FP as shown in **Figure 4**:

1. **Volume V_c** of the candidate in mm^3. Candidates with 5 mm^3 < V_c <120 mm^3 were selected as CMBs.
2. **Ellipticity ε_c:** A measure of elongated nature of an object. The value ranges between [0,1] and a sphere has a value of 0. Candidates having ε_c < 0.2 were selected as CMBs.
3. **Solidity S_c:** The ratio between volume of the candidate and its convex volume. The value ranges within [0,1] and an object with its volume equal to that of its convex hull has a value of 1; this criterion removes the curved fragments of blood vessels and intersection of vessels that survived the vessel removal step, since they would have low solidity. A lower threshold value of 0.6 was applied on S_c of candidates to be selected as CMBs.
4. **Diameter D_c** of the candidate in mm. Here, diameter is the distance between the endpoints of the longest line that can be drawn through the candidate. Candidates with diameters 2 mm < D_c < 10 mm were selected as CMBs.

FIGURE 3 | Features extracted for the voxel-wise CMB candidate detection, along with detected CMB candidates. Images showing **(A)** Input intensity, **(B)** exponential transformed intensity, **(C)** CLAHE output, **(D)** fast radial symmetry transform (FRST) output, **(E)** structure tensor output, **(F)** Gaussian filtered output, **(G)** Laplacian of Gaussian (LoG) output, and **(H)** Binary CMB candidates obtained by thresholding the voxel-wise probability map at 0.8. Inset figures show the magnified versions of the regions indicated in the boxes.

We considered a candidate as a CMB if all the above criteria were satisfied. The values for the above criteria were chosen empirically by a trial-and-error method based on the best subject-level performance on the training data from the OXVASC and TICH2 datasets using the metrics specified in section 3.5. The final cluster-wise performance values (providing lesion-level CMB detection performance) after the filtering step are provided in Section 4.1.

Preselection criterion: Finally, a subject is classified as having CMBs if the count of the detected CMB candidates (after the filtering stage) exceeds an empirically set threshold value Th_{NCMB}. For each dataset, we determined the Th_{NCMB} value by measuring the subject-level performance values (specified in section 3.5) at various thresholds within a pre-specified range, and choosing the threshold that provided the best set of performance values as Th_{NCMB} (refer to section 4.2.1). The range of threshold values was set to be slightly higher than the average number of CMBs in the OXVASC and TICH2 datasets since
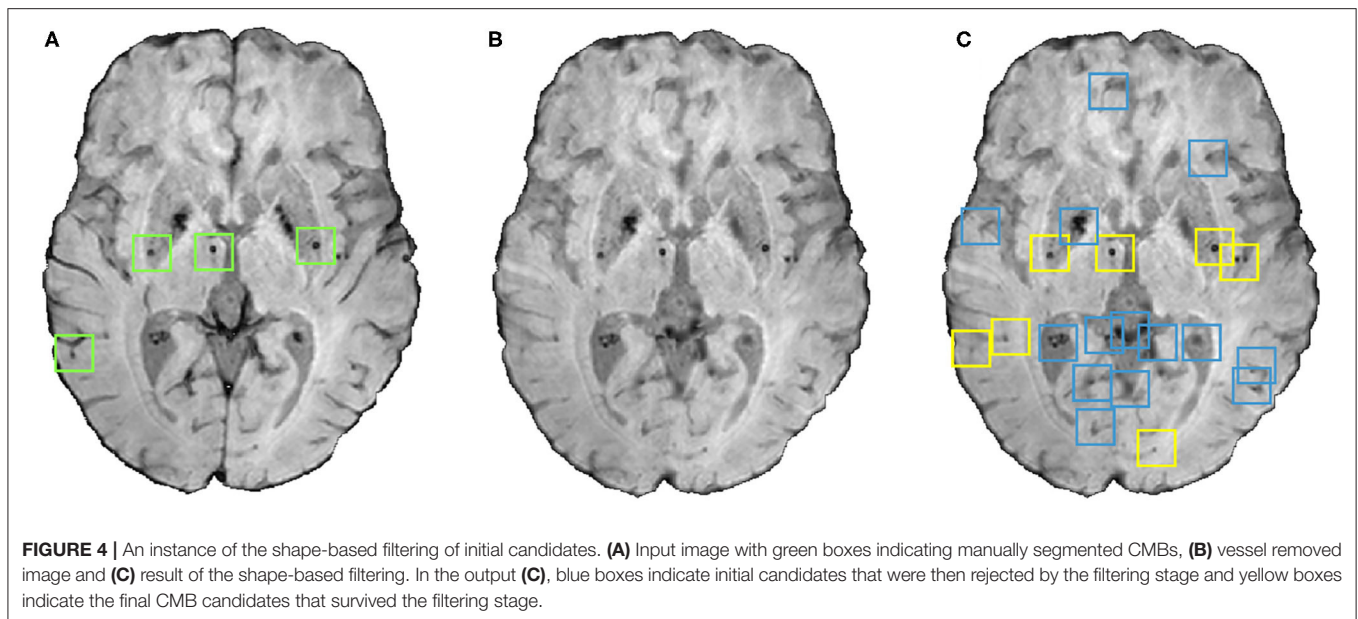
we had to allow for the presence of false positives (e.g., vessel fragments and sulci missed in the vessel removal stage).

3.4. Evaluation of the CMB Candidate Preselection Pipeline

We evaluated the proposed pipeline by training and testing it using different datasets (with different modalities and acquisition characteristics) to study the effect of dataset characteristics on its performance. We performed the following experiments:

3.4.1. Initial Evaluation of the Proposed Preselection Pipeline Within Datasets

We performed leave-one-out validation separately on the OXVASC and TICH2 datasets. We used T2*-GRE images from 40 subjects from the OXVASC dataset and, separately, SWI from 50 subjects from the TICH2 dataset. For the UKBB dataset, we performed 5-fold cross-validation on SWI images from 180 subjects with training/validation/test split of 70/10/20%. Note

FIGURE 4 | An instance of the shape-based filtering of initial candidates. **(A)** Input image with green boxes indicating manually segmented CMBs, **(B)** vessel removed image and **(C)** result of the shape-based filtering. In the output **(C)**, blue boxes indicate initial candidates that were then rejected by the filtering stage and yellow boxes indicate the final CMB candidates that survived the filtering stage.

that, for the cross-validation on the UKBB dataset, we used the SVM model trained on the TICH2 dataset for the initial candidate detection (since voxel-wise manual segmentations were not available for the UKBB dataset). We determined the performance metric values (specified in section 3.5) at different settings of the threshold Th_{NCMB} in order to plot the ROC curve.

3.4.2. Evaluation of the Generalisability of Our Method Across Datasets

We trained our pipeline (SVM classifier in the candidate detection step) on T2*-GRE images from 40 subjects of the OXVASC dataset. We performed hyper-parameter optimisation, as specified in Section 4.1, on the held-out data from the OXVASC dataset (in addition to 40 subjects used for evaluation). We later evaluated it on the SWI from 50 subjects of the TICH2 dataset. Similarly, we trained the pipeline on SWI from the TICH2 dataset and evaluated it on T2*-GRE images of the OXVASC dataset. Finally, we applied both OXVASC-trained and TICH2-trained pipelines on SWI from 180 subjects of the UKBB dataset. We could not train the SVM model on the UKBB dataset due to non-availability of voxel-wise manual labels and hence all training/testing combinations were not possible. For these experiments, we determined the performance of the preselection pipeline using the evaluation measures specified in section 3.5.

3.4.3. Comparison of the Proposed Pipeline With Algorithms Using Imaging/Non-Imaging Factors on the UKBB Dataset

In addition to the CMB lesion count extracted from imaging data, various non-imaging demographic/clinical factors have been associated with the incidence of CMBs, with age and BP being the common ones (Roob et al., 2000; Horita et al., 2003; Vernooij et al., 2008; Poels et al., 2010). Therefore, in this experiment, our objectives are (i) to compare the following category of methods (including the proposed method) using imaging and/or non-imaging information (refer to methods 1–3) and (2) to verify

if using clinical factors alongside the automated pipeline would increase the performance of CMB candidate subject preselection (refer to method 4):

1. **Method 1: demographic/clinical factors:** We considered factors such as age, diastolic BP and systolic BP separately by applying a range of thresholds to each factor to determine the baseline performance of each factor.

2. **Method 2: other non-imaging factors:** While age and BP are the most common factors, several other factors have also been associated with CMBs. Therefore, we considered other non-imaging factors such as smoking, white matter hyperintensity (WMH) volume and mean reaction time (MRT, linked with cognitive ability) in addition to age and BP. We used these 6 factors as features to train a SVM classifier (SVM_{NI}) and a random forest classifier (RF_{NI}). For SVM_{NI}, we used a radial basis function kernel, with tolerance value of 1×10^{-3} and ϵ of 0.1 (used *fitrsvm* in Matlab, 2016b). For RF_{NI}, we used 150 trees, samples at leaf node = 5, mean squared error (MSE) as criteria for splitting and out-of-bag error score set to True in *TreeBagger* command in Matlab (for getting feature importance, see below). The above training hyperparameters were chosen using trial-and-error method on the validation data. We performed 5-fold cross-validation with the same training-validation-test split used in section 3.4.1. We also performed feature ranking to determine the importance of individual features using the out-of-bag (OOB) prediction[1] error metric in RF_{NI}.

3. **Method 3: determining CMB lesion count using imaging-based methods:** Similar to our proposed pipeline, we used

---

[1]In RF, OOB predictions are those obtained on the samples that were left out while randomly choosing the training subsets for individual trees during bootstrap aggregating (bagging). We considered the increase in MSE while permuting the OOB observations across each feature, averaged over all trees in the ensemble and divided by the standard deviation obtained over the trees. Hence, larger increase in MSE indicates more important features.

imaging data (T2*-GRE from OXVASC and SWI from TICH2) and applied another baseline method, where we replaced the SVM-based CMB candidate detection step with a thresholding method to detect initial CMB candidates. We used the same preprocessing steps used in our pipeline (refer to section 2.1) and applied a threshold value at the lower 5th percentile of the intensity value (determined empirically from the intensity histograms) and considered the voxels within the brain mask below the threshold as initial CMB candidates. We compared the 5-fold cross-validation results of our proposed method with the thresholding method.

4. **Method 4: non-imaging factors + CMB lesion count determined from imaging data:** We used a total of 7 features including the 6 factors used in method 2 and the CMB lesion count (determined using the proposed pipeline). Similar to method 2, we trained a SVM (SVM_{NI+I}) and RF (RF_{NI+I}) classifiers and evaluated them using 5-fold cross-validation. In addition, we determined the feature importance from OOB prediction error using RF_{NI+I}. For training the RF and SVM classifiers, we used the same parameters used in method 2.

From all the above methods we determined the performance values (specified in section 3.5) at different thresholds for plotting the ROC curves.

3.5. Performance Evaluation Metrics for CMB Candidate Subject Preselection Pipeline

For the CMB candidate subject preselection pipeline, we used the following measures for evaluation:

- **Subject-level true positive rate (TPR):** For a given dataset D, the subject-level TPR is the number of predicted TP CMB subjects (S_{TP}) divided by the number of true CMB subjects, as given by,

$$\text{subject-level TPR} = \frac{S_{TP}}{(S_{TP} + S_{FN})} \quad (1)$$

where S_{FN} is the number of false negative subjects.

- **Subject-level specificity:** For a given dataset D, the subject-level specificity is the number of predicted true negative subjects (S_{TN}) divided by the number of non-CMB subjects, as given by,

$$\text{subject-level specificity} = \frac{S_{TN}}{(S_{TN} + S_{FP})} \quad (2)$$

where S_{FP} is the number of false positive subjects.

- **Subject-level accuracy:** For a given dataset D consisting of S_N subjects, the subject-level accuracy is given by the number of correctly predicted CMB and non-CMB subjects (S_{TP} and S_{TN}) divided by the total number of subjects,

$$\text{subject-level accuracy} = \frac{S_{TP} + S_{TN}}{S_N} \times 100 \, (\%) \quad (3)$$

We plotted subject-level ROC curves using subject-level TPR and subject-level false positive rate (subject-level FPR = 1 − subject-level specificity) values.

Additionally, although the main focus of this work was the detection of subjects with CMBs rather than the delineation of single lesions, we used the following cluster-wise measures for obtaining an indicative evaluation of the lesion-level performance at the CMB initial candidate detection step.

- **Cluster-level TPR:** The number of true positive CMBs divided by the total number of true CMBs as given by,

$$\text{Cluster-wise TPR} = \frac{TP_{clus}}{(TP_{clus} + FN_{clus})} \quad (4)$$

where TP_{clus} and FN_{clus} are true positive and false negative CMBs, respectively.

- **Average number of false positive clusters per subject (FPavg):** For a given dataset D, FPavg is defined as the ratio of the total number of detected FPs to the number of subjects (or images) in the dataset, as given by,

$$\text{FPavg} = \frac{\text{Total number of FPs}}{\text{Number of subjects in D}} \quad (5)$$

We used 26-connectivity to determine the clusters for obtaining the above metrics. We considered a cluster as a true positive if it overlaps with a ground truth cluster by at least one voxel, while false positive clusters are the ones that have no overlap with any ground truth cluster. We used cluster-wise TPR and FPavg values for plotting a free-response ROC (FROC) curve, which is a plot of cluster-wise TPR vs. the average number of false positives per image/subject.
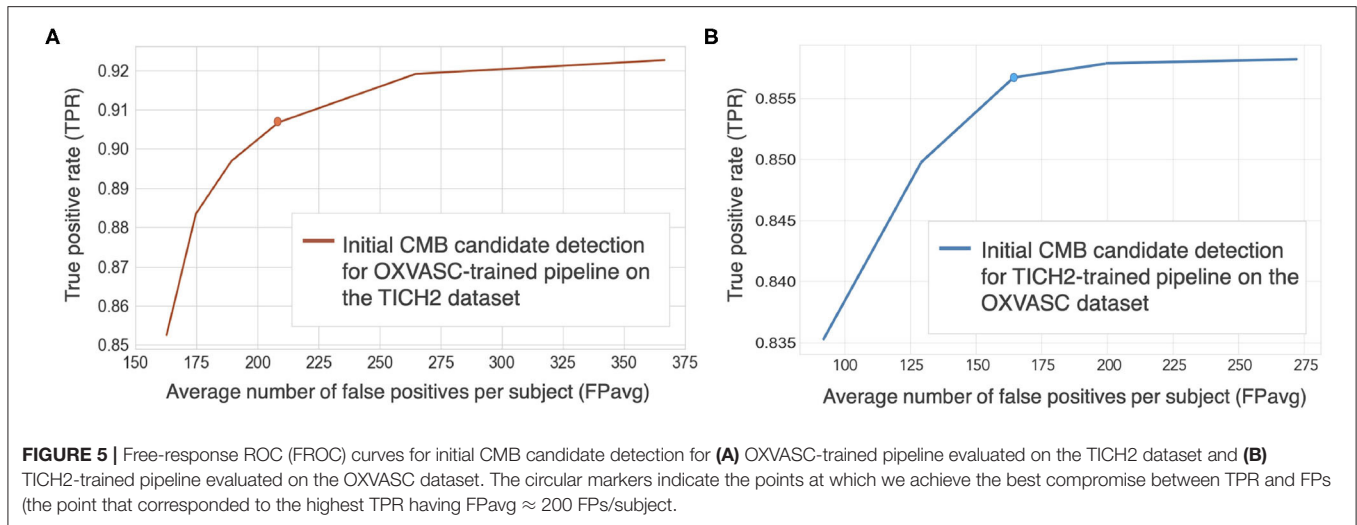
4. RESULTS
4.1. Effect of Threshold Value on Obtaining CMB Candidates

Figures 5A,B show FROC curves for the TICH2 and OXVASC datasets used to determine the threshold value th_{prob} for the voxel-wise initial CMB candidate detection of the OXVASC-trained and TICH2-trained pipelines, respectively. Since we will be applying, at a later stage, the shape-based filtering step (step 3) to further reduce false positives, we prioritised achieving high cluster-wise TPR in this initial CMB candidate detection step (step 2). For both datasets, we achieved the best performance at th_{prob} value of 0.8 with FPavg values close to or less than 200 CMBs/subject. The SVM classifier trained on the OXVASC dataset achieved a cluster-wise TPR of 0.906 with FPavg of 210.6 on the TICH2 dataset, while the SVM trained on the TICH2 dataset achieved cluster-wise TPR of 0.86 with FPavg of 178.7 on the OXVASC dataset.

Cluster-wise results after filtering initial CMB candidates: For the pipeline trained on the OXVASC data and evaluated on the TICH2 data, the pipeline achieved a cluster-wise TPR of 0.88 with FPavg of 20.4 after the final candidate filtering stage. Similarly for the pipeline trained on the TICH2 data and evaluated on the OXVASC data, the pipeline achieved a cluster-wise TPR of 0.85 with FPavg of 12.8 after the final candidate filtering stage.

FIGURE 5 | Free-response ROC (FROC) curves for initial CMB candidate detection for **(A)** OXVASC-trained pipeline evaluated on the TICH2 dataset and **(B)** TICH2-trained pipeline evaluated on the OXVASC dataset. The circular markers indicate the points at which we achieve the best compromise between TPR and FPs (the point that corresponded to the highest TPR having $FP_{avg} \approx 200$ FPs/subject.

4.2. Evaluation of the CMB Candidate Preselection Pipeline

Figures 6, **7** show the ROC curves for leave-one-out validation and our experiments to analyse model generalisability, respectively. **Figure 10** shows the ROC curves for comparison of various methods on the UKBB dataset. The overall results are reported in **Table 1**. Alternatively, subject-level specificity and accuracy values for a specific subject-level TPR value of 95% are reported in **Supplementary Table 3**.

4.2.1. Initial Evaluation of the Proposed Preselection Pipeline Within Datasets

Figure 6 shows the separate leave-one-out validation results for the OXVASC and the TICH2 datasets. The best set of performance values for each dataset was determined from the knee point on the ROC. From the ROC curves it can be seen that, on the OXVASC dataset, the pipeline achieves the best subject-level performance metric values: TPR = 0.96, specificity = 0.83 and accuracy = 0.90 at the threshold $Th_{NCMB} = 30$ CMBs (and TPR = 0.81, specificity = 0.87 and accuracy = 0.84 at the threshold $Th_{NCMB} = 35$ CMBs). Similarly, on the TICH2 dataset, the pipeline achieves the best subject-level performance values: TPR = 0.91, specificity = 0.81 and accuracy = 0.86 at the threshold $Th_{NCMB} = 35$ CMBs. On performing the 5-fold cross-validation on the UKBB dataset (ROC curve shown by black solid line in **Figure 10**), the proposed pipeline achieved the best subject-level performance with a TPR = 0.91, specificity = 0.86 and accuracy = 0.89 at $Th_{NCMB} = 35$ CMBs.
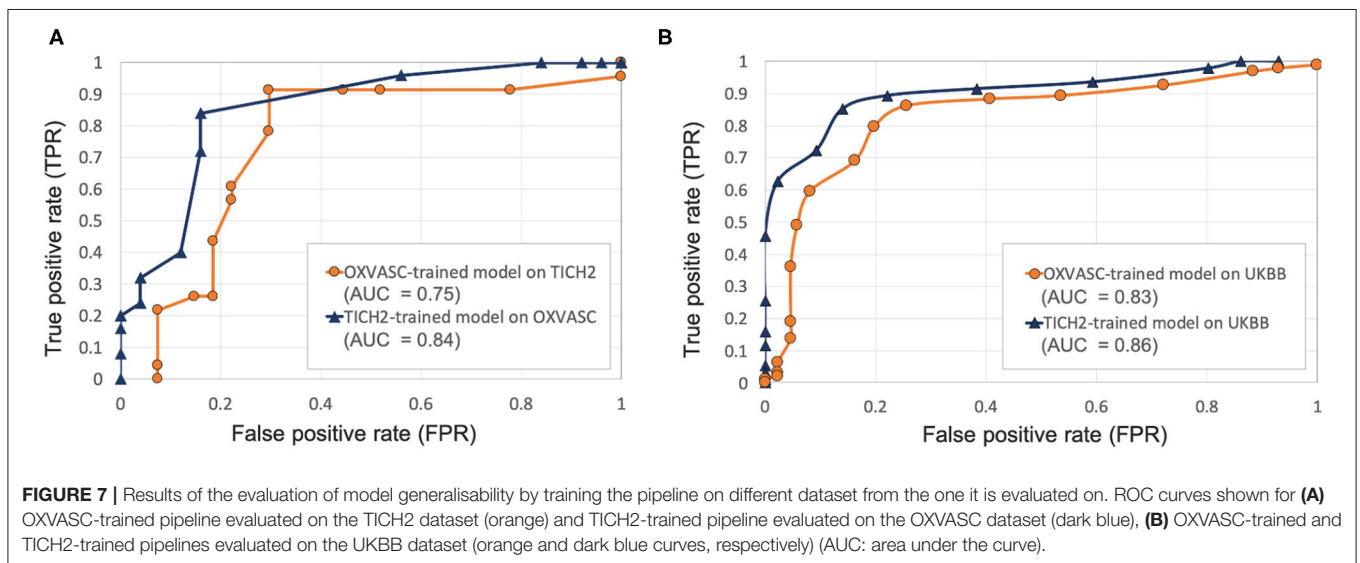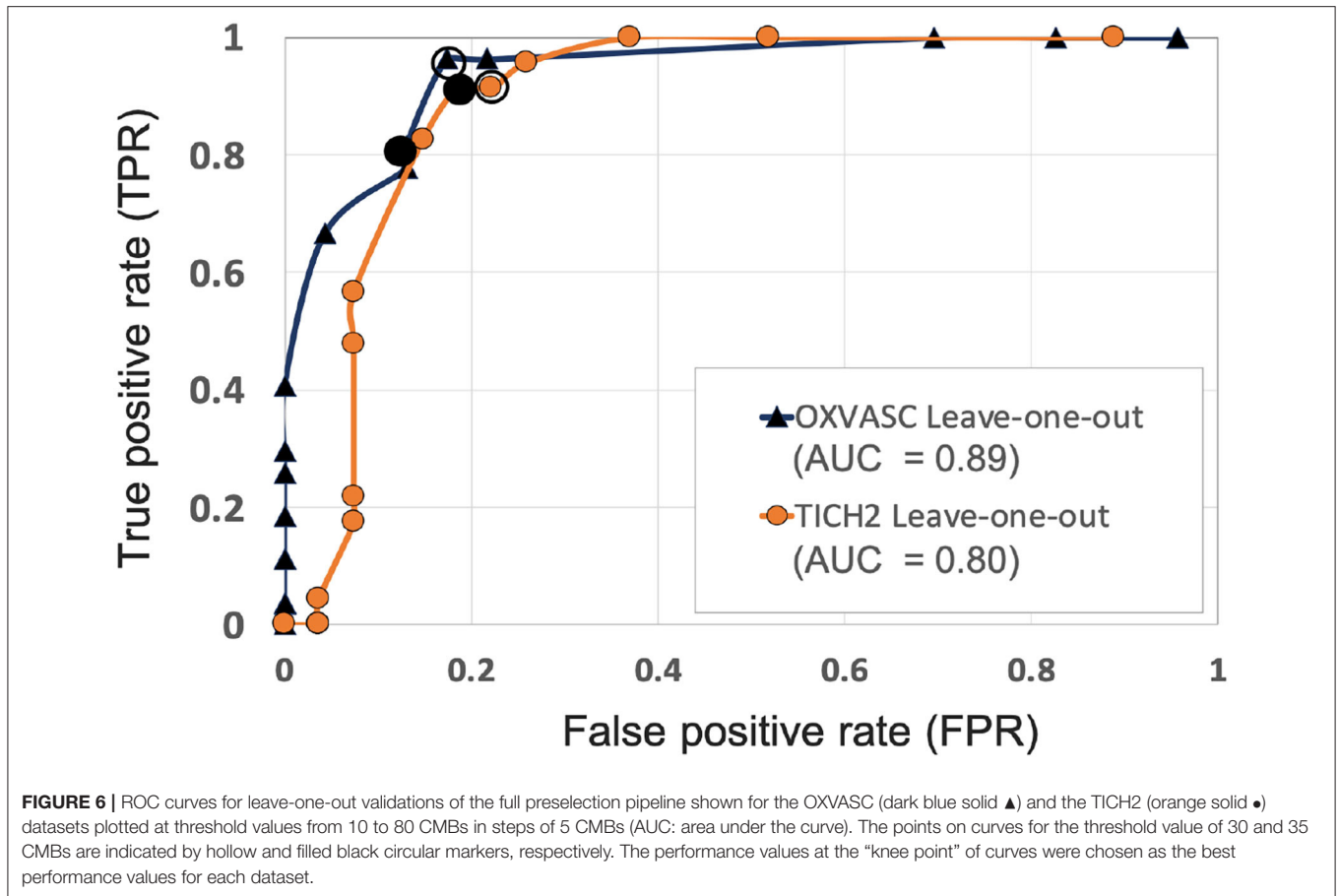
4.2.2. Evaluation of the Generalisability of the Proposed Pipeline Across Datasets

Pipeline trained on OXVASC and evaluated on the TICH2 data: **Figure 7A** shows the ROC curve for the prediction of CMB subjects at various values of the threshold Th_{NCMB} on the number of detected CMBs for individual subjects. We achieved the best performance with Th_{NCMB} set to 35 CMBs. 50 subjects). Despite the presence of haemorrhagic lesions in all subjects, the

pipeline gave a sensitivity of 0.90 and an accuracy of 0.81, with a subject-level specificity of 0.70. **Figure 8** shows a few example cases of correct and incorrect subject-level detections on the TICH2 dataset (manually segmented CMBs are indicated with green boxes). From the Figure, it can be seen that our algorithm correctly predicted the subjects with a high number of CMBs (**Figures 8A,B**). Typically these subjects with high number of CMBs or no CMBs were detected better than subjects with very low number of CMBs (well under Th_{NCMB} value of 30 CMBs, as shown in **Figure 8E**) or subjects having small haemorrhages (**Figures 8C,D**).

Pipeline trained on TICH2 and evaluated on the OXVASC data: **Figure 7A** shows the ROC curve for CMB candidate subject preselection with the TICH2-trained model on the OXVASC dataset. As in the previous case, the model achieved the best performance for a threshold value Th_{NCMB} of 35 CMBs. The model provided a subject-level TPR of 0.83, a subject-level specificity of 0.84 and a subject-level accuracy of 0.83. The number of subjects with high CMB counts (> 50 CMBs) in the training dataset (TICH2) is higher than that in OXVASC. Hence, the TICH2-trained model was more specific in detecting CMB subjects on the OXVASC dataset, achieving a higher specificity of 0.84. We have shown a few examples of subject-level detections in **Figure 9**. As in the previous case, subjects with high CMB counts (**Figures 9A,B**) were correctly predicted as CMB subjects.

OXVASC-trained and TICH2-trained pipelines on the UKBB dataset: We applied both the OXVASC-trained and the TICH2-trained models to the UKBB dataset. As shown in **Figure 7B**, the model trained on the TICH2 dataset provided better performance when compared to the OXVASC-trained model. For the OXVASC-trained model, the pipeline achieved subject-level TPR = 0.80, specificity = 0.80 and accuracy = 0.80 using a threshold of 30 CMBs. For the TICH2-trained model, the pipeline achieved the subject-level TPR = 0.85, specificity = 0.86 and accuracy = 0.85 using a threshold of 35 CMBs. We could not perform the lesion-level analysis on the UKBB dataset (as we did for the OXVASC and TICH2 datasets), since we had only subject-level manual labels.

FIGURE 6 | ROC curves for leave-one-out validations of the full preselection pipeline shown for the OXVASC (dark blue solid ▲) and the TICH2 (orange solid ●) datasets plotted at threshold values from 10 to 80 CMBs in steps of 5 CMBs (AUC: area under the curve). The points on curves for the threshold value of 30 and 35 CMBs are indicated by hollow and filled black circular markers, respectively. The performance values at the "knee point" of curves were chosen as the best performance values for each dataset.

FIGURE 7 | Results of the evaluation of model generalisability by training the pipeline on different dataset from the one it is evaluated on. ROC curves shown for **(A)** OXVASC-trained pipeline evaluated on the TICH2 dataset (orange) and TICH2-trained pipeline evaluated on the OXVASC dataset (dark blue), **(B)** OXVASC-trained and TICH2-trained pipelines evaluated on the UKBB dataset (orange and dark blue curves, respectively) (AUC: area under the curve).

4.2.3. Comparison of the Proposed Pipeline With Algorithms Using Imaging/Non-Imaging Factors on the UKBB Dataset

Figure 10 shows the ROC curves for comparison of various methods used for preselecting CMB candidates from the UKBB

dataset. For the best performance metrics from ROC curves for each method, refer to **Supplementary Section 2**. Out of all the categories, using individual demographic/clinical factors (method 1) provided the worst performance. Among the individual factors, age provided the worst performance while the

TABLE 1 | Performance of the proposed method during the initial evaluation within datasets and during the generalisability experiments across datasets.

		Testing datasets		
		OXVASC	**TICH2**	**UKBB**
Training datasets	OXVASC	Initial leave-one-out evaluation within datasets: (Sec: 3.4.1, results: 4.2.1) Sensitivity = 0.81 Specificity = 0.87 Accuracy = 0.84	Evaluation of method generalisability across datasets: (Sec: 3.4.2, results: 4.2.2) Sensitivity = 0.90 Specificity = 0.70 Accuracy = 0.81	Evaluation of method generalisability across datasets: (Sec: 3.4.2, results: 4.2.2) Sensitivity = 0.70 Specificity = 0.91 Accuracy = 0.81
	TICH2	Evaluation of method generalisability across datasets: (Sec: 3.4.2, results: 4.2.2) Sensitivity = 0.83 Specificity = 0.84 Accuracy = 0.83	Initial leave-one-out evaluation within datasets: (Sec: 3.4.1, results: 4.2.1) Sensitivity = 0.91 Specificity = 0.81 Accuracy = 0.86	Evaluation of method generalisability across datasets: (Sec: 3.4.2, results: 4.2.2) Sensitivity = 0.85 Specificity = 0.86 Accuracy = 0.85

The sensitivity, specificity and accuracy values are reported at subject-level for a threshold value of TH_{NCMB} = 35 CMBs.

diastolic BP provided better performance, with subject-level TPR = 0.37, specificity = 0.72 and accuracy = 0.54 at a threshold of 93.4 mmHg. For classification based on non-imaging factors (method 2), the RF classifier provided better results compared to the SVM classifier. Using the RF classifier, we obtained subject-level TPR = 0.73, specificity = 0.74, accuracy = 0.74 at a threshold of 0.6. Among imaging-based methods (method 3) used to determine CMB lesion count, the proposed pipeline provided better results (subject-level TPR = 0.91, specificity = 0.86 and accuracy = 0.89) compared to the thresholding method, detecting more CMB lesions in the CMB subjects when compared to the non-CMB subjects, and hence better at identifying subjects containing CMBs. Of all the methods, classification using both non-imaging factors and CMB lesion count (method 4) provided the best performance, especially using the RF classifier, achieving subject-level TPR, specificity and accuracy of 0.95 at a threshold of 0.6.
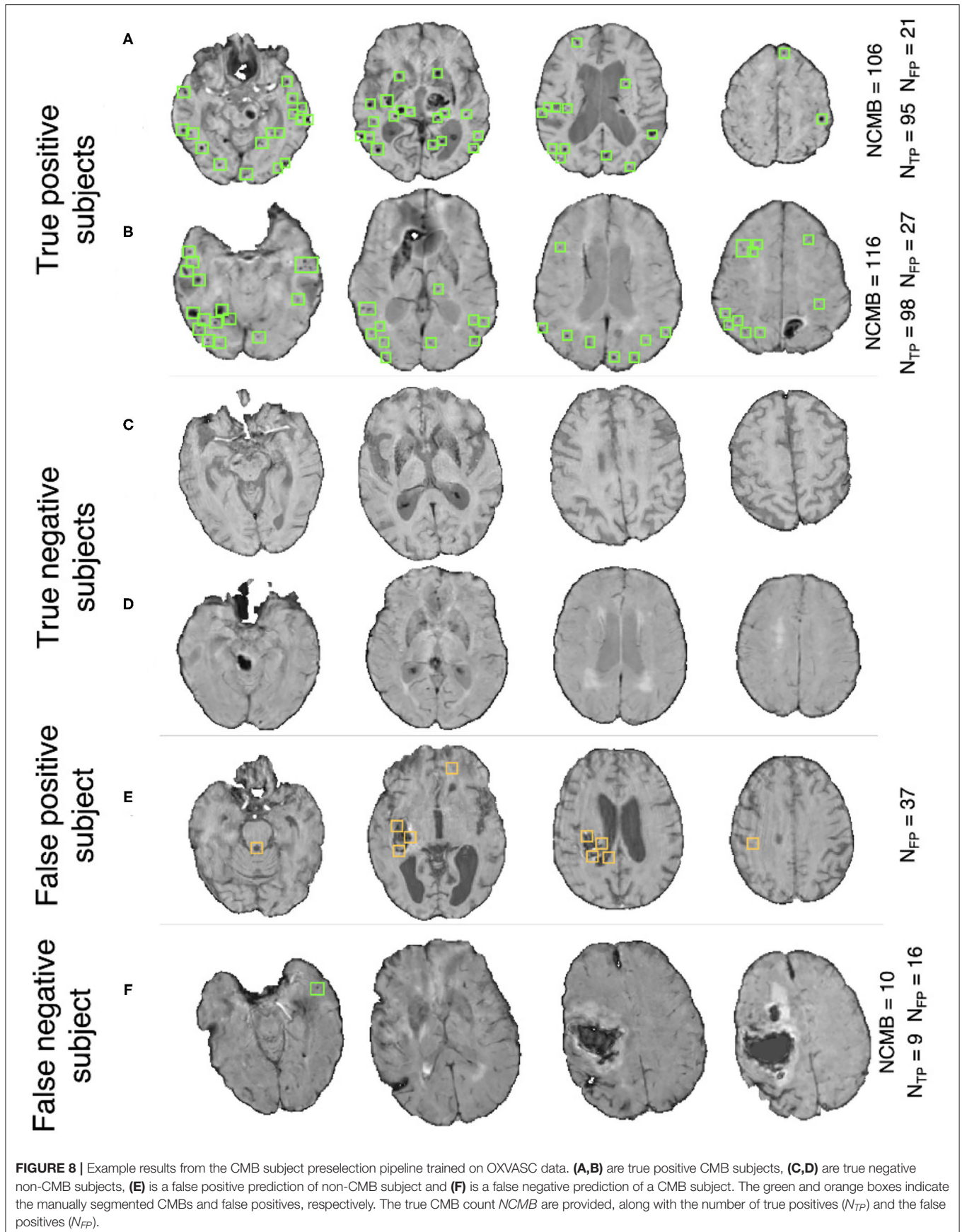
Figure 11 shows the OOB feature importance of various features used in the RF_{NI} and RF_{NI+I} classifiers. For RF_{NI}, WMH volume had the highest importance (imp = 2.1) followed by the BP values (diastolic BP imp = 0.41 and systolic BP imp = 0.42). For RF_{NI+I}, CMB lesion count had the highest feature importance (imp = 2.8) followed by WMH volume (imp = 1.25) and diastolic BP (imp = 0.35). The higher feature importance of diastolic BP compared to age (imp = 0.25) also aligns well with our earlier comparison where diastolic BP provided the best performance among the individual factors. For both RF_{NI} and RF_{NI+I}, smoking was the least important feature with imp = −0.08 and 0.01, respectively.
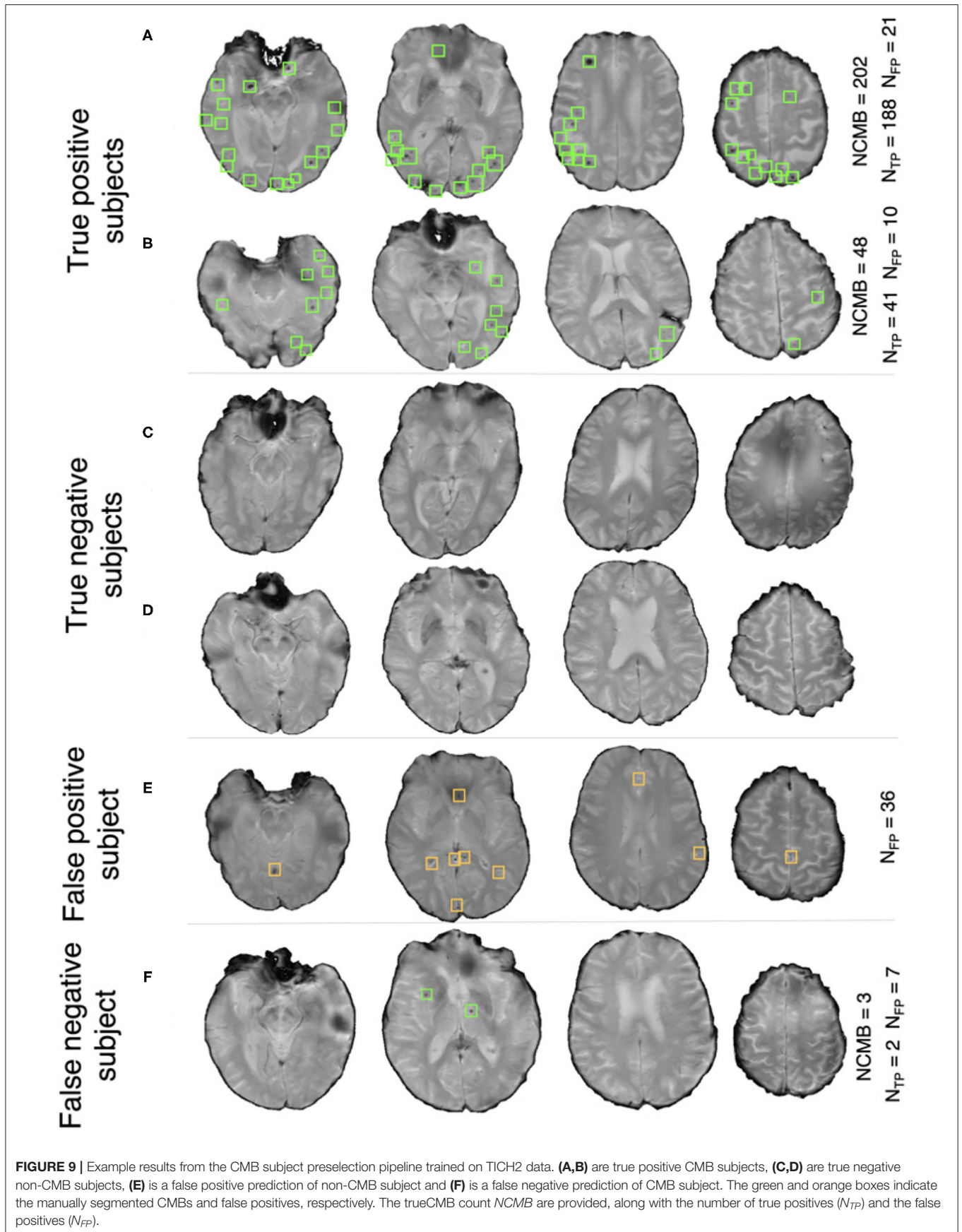
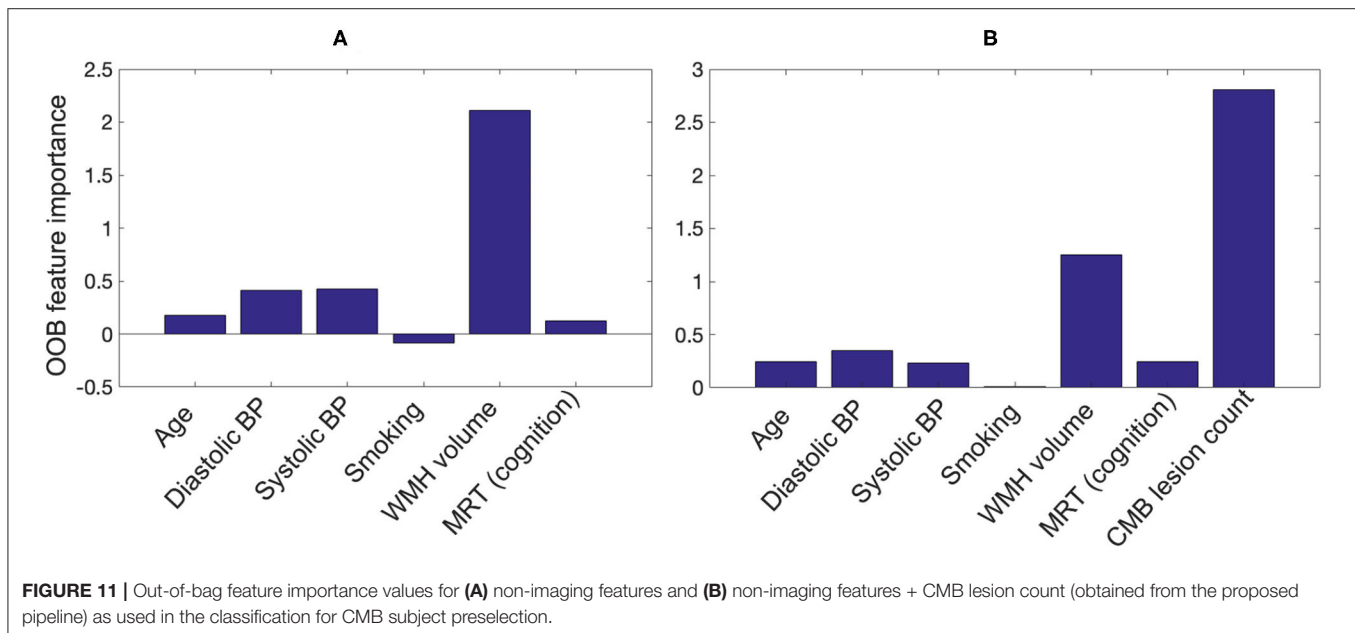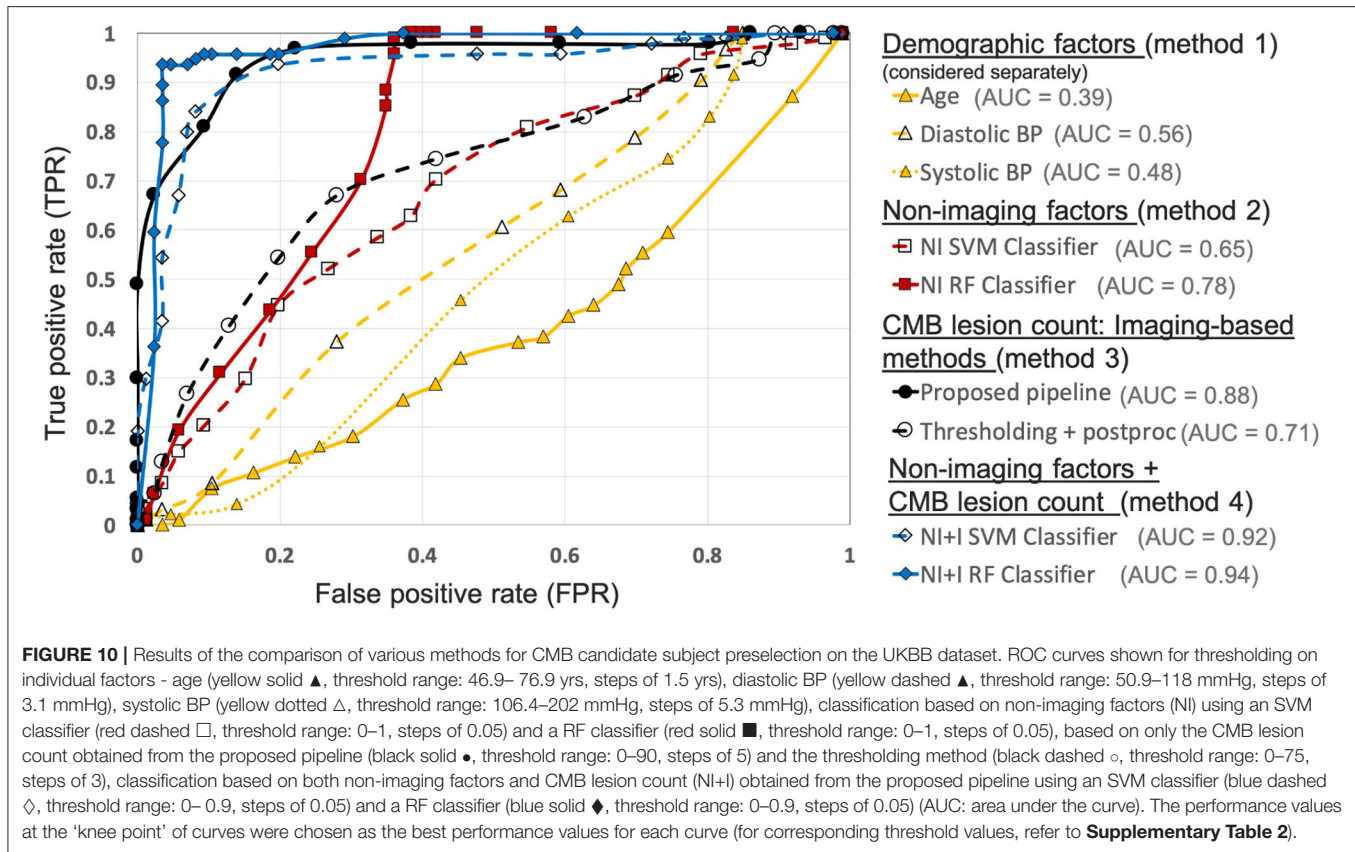## 5. DISCUSSION AND CONCLUSIONS

In this work, we proposed a fully automated pipeline, which is computationally light and takes into account various intensity, shape and anatomy-based characteristics of CMBs, for preselecting CMB candidate subjects from large datasets. We compared various methods involving non-imaging

demographic/clinical factors and CMB lesion count from the imaging data, including our proposed pipeline, on a subset of the UKBB dataset for which we had subject-level manual labels regarding the presence of CMBs. We finally applied our pipeline on different datasets for training and evaluation to validate its generalisability with respect to variations in intensity, acquisition protocols and pathological conditions. Our pipeline provided subject-level accuracy >85% during initial validation on each individual dataset. On applying our pipeline on various datasets, we observed good generalisability across datasets with subject-level accuracy >85% when trained on SWI and applied to SWI/T2*-GRE images, and >80% when trained on T2*-GRE images and applied to SWI.

During the initial within-dataset evaluation, the pipeline provided better subject-level results on the OXVASC and UKBB datasets when compared with the results from the TICH2 dataset. This could be due to the fact that the pipeline detected more false positive lesions along the edges of intracerebral haemorrhages in the TICH2 dataset. In addition to the presence of haemorrhages, the increased amount of FPs could be related to the use of SWI. In fact, the best threshold on the CMB lesion count (corresponding to the 'knee point' on the ROC curve) for subject-level preselection is higher for datasets with SWI (the threshold is 35 CMBs for both the UKBB and TICH2 datasets) when compared to the OXVASC dataset with T2*-GRE images (where the threshold is 30 CMBs). This might be because SWI improves not only the contrast of CMBs but also the mimics that could lead to more spurious CMB detections and hence a higher threshold is needed on the CMB lesion count. However, while we chose thresholds that gave the best performance on the datasets, we cannot exclude the possibility of subjects with low number of true CMBs (with low number of false positives) being incorrectly classified as non-CMB subjects. It is worth noting that the threshold (Th_{NCMB}) value of 35 CMBs might not be an optimal threshold value on any unseen dataset. Hence, careful checking and potential experimentation to determine the optimal threshold is recommended when using this method on a

FIGURE 8 | Example results from the CMB subject preselection pipeline trained on OXVASC data. **(A,B)** are true positive CMB subjects, **(C,D)** are true negative non-CMB subjects, **(E)** is a false positive prediction of non-CMB subject and **(F)** is a false negative prediction of a CMB subject. The green and orange boxes indicate the manually segmented CMBs and false positives, respectively. The true CMB count *NCMB* are provided, along with the number of true positives (N_{TP}) and the false positives (N_{FP}).

FIGURE 9 | Example results from the CMB subject preselection pipeline trained on TICH2 data. **(A,B)** are true positive CMB subjects, **(C,D)** are true negative non-CMB subjects, **(E)** is a false positive prediction of non-CMB subject and **(F)** is a false negative prediction of CMB subject. The green and orange boxes indicate the manually segmented CMBs and false positives, respectively. The trueCMB count *NCMB* are provided, along with the number of true positives (N_{TP}) and the false positives (N_{FP}).

FIGURE 10 | Results of the comparison of various methods for CMB candidate subject preselection on the UKBB dataset. ROC curves shown for thresholding on individual factors - age (yellow solid ▲, threshold range: 46.9– 76.9 yrs, steps of 1.5 yrs), diastolic BP (yellow dashed ▲, threshold range: 50.9–118 mmHg, steps of 3.1 mmHg), systolic BP (yellow dotted △, threshold range: 106.4–202 mmHg, steps of 5.3 mmHg), classification based on non-imaging factors (NI) using an SVM classifier (red dashed □, threshold range: 0–1, steps of 0.05) and a RF classifier (red solid ■, threshold range: 0–1, steps of 0.05), based on only the CMB lesion count obtained from the proposed pipeline (black solid ●, threshold range: 0–90, steps of 5) and the thresholding method (black dashed ○, threshold range: 0–75, steps of 3), classification based on both non-imaging factors and CMB lesion count (NI+I) obtained from the proposed pipeline using an SVM classifier (blue dashed ◇, threshold range: 0– 0.9, steps of 0.05) and a RF classifier (blue solid ◆, threshold range: 0–0.9, steps of 0.05) (AUC: area under the curve). The performance values at the 'knee point' of curves were chosen as the best performance values for each curve (for corresponding threshold values, refer to **Supplementary Table 2**).

FIGURE 11 | Out-of-bag feature importance values for **(A)** non-imaging features and **(B)** non-imaging features + CMB lesion count (obtained from the proposed pipeline) as used in the classification for CMB subject preselection.

dataset that has different characteristics (e.g., different modalities, population demographics and scanning characteristics).

On evaluating the generalisability of the proposed pipeline, the performance when the training and testing datasets were acquired with different modalities was slightly lower than the initial cross-validation performance values. Our preselection

pipeline provided results with similar accuracy for different training data, with a slightly better specificity on the OXVASC dataset using TICH2-trained model when compared to the TICH2 dataset using OXVASC-trained model, while providing a lower subject-level TPR. The prediction of CMB candidate subjects was more precise in the OXVASC dataset likely due to

the fact that the pipeline was trained using features extracted from SWI that is highly sensitive to CMBs and susceptibility artefacts. At this point, it is worth noting that there are pros and cons with using T2*-weighted GRE or SWI sequences. In the case of T2*-GRE images (the OXVASC dataset), the contrast between CMBs and background is lower than in SWI (and some CMBs might not be visible at all on T2*-GRE images) and hence the number of detected false positives was low, providing high specificity (85%, when compared to the specificity of 70% in the TICH2 dataset) but low cluster-wise TPR (83% compared to the TPR of 90% in the TICH2 dataset). This shows that SWI sequences (the TICH2 dataset) are highly sensitive to CMBs and hence result in detection of more CMBs, albeit with increased FPs as well. Therefore, choosing training datasets containing a similar modality (especially those having single modalities) to train the pipeline could help in achieving more accurate detection when applied to the large datasets. In fact, on applying OXVASC-trained and TICH2-trained pipelines to the UKBB dataset, the TICH2-trained pipeline provided better results since it was trained on SWI, the modality used in the UKBB dataset. However, it is worth noting that, despite the difference in population and CMB prevalence across the datasets used in this study, our pipeline showed comparable performance, even when trained on a dataset with very different prevalence (e.g., OXVASC, a stroke population, and UKBB, a prospective epidemiological study).

Regarding the factors that could lead to true CMBs being missed, in addition to the effect of various modalities, the preprocessing steps (e.g., brain extraction) and the blood vessel/sulci removal step might also affect CMB detection. For instance, we performed the brain extraction directly on SWI/T2*-GRE images depending on their availability. While this worked in the vast majority of cases, we observed some loss of tissue in the top slices of the brain in very rare cases, which in turn might lead to true CMBs being missed. Improvement in the brain extraction (e.g., by registration of a brain mask obtained from a T1-weighted brain to SWI/T2*-GRE images) could overcome this problem. As for the vessel removal step, a few CMBs were misclassified as a blood vessel or sulci or other elongated hypointense structure and hence was completely or partially painted over with the mean value of neighbouring voxels. While the partially painted over CMBs (thus with reduced contrast, similar to the left-most CMB shown in **Figure 4**) could still be detected, the completely removed CMBs could not, resulting in true CMBs being missed. Another reason for potentially missing true CMBs could be the failure of CMB candidates to pass the criteria on shape-based attributes (section 3.3); for example, if they are so close together that they would be detected as a single cluster. While the chance of two CMBs being very close to each other is very slim, given their size and sparse distribution, we cannot exclude that possibility.

On comparing various methods involving non-imaging and CMB lesion counts from the imaging data, we observed that applying thresholds on the individual demographic factors provided the worst results. This shows that even though factors such as age and hypertension are commonly associated with CMBs, they are not sufficient by themselves to preselect CMB candidate subjects in a given population. This further

supports the need for pipelines such as ours that can also use the imaging data for subject-level preselection, especially in large datasets. Interestingly, among age and BP, the latter provides better preselection (and also has higher feature importance in the RF classifier), despite age being reported as the most commonly used factor (Vernooij et al., 2008; Poels et al., 2010; Takashima et al., 2011). Similarly, among the features used in the RF classifiers (both RF_{NI} and RF_{NI+I}), smoking is the least important feature, even though it has been reported as one of the risk factors for CMBs in various population-level studies (Tsushima et al., 2002; Poels et al., 2010). However, more experimentation would be required on a larger sample size to further establish the relationship between age/smoking status and occurrence of CMBs. Also, higher correlations of non-imaging features (e.g., age/blood pressure) with CMBs might be present in other datasets and hence the reported results might not represent all possibilities.

The proposed method provided better results than the thresholding method, highlighting the utility of shape-based features, in addition to intensity. The best results were obtained for the classifiers using a combination of non-imaging factors and CMB lesion counts, however, with CMB lesion count being the most important feature. The high feature importance of CMB lesion count shows that the proposed pipeline determines a lesion count that is highly useful, despite the number of false positive lesions. Also, the best performance of the combination of features and lesion counts shows that the non-imaging features could be used to improve the results of our proposed pipeline for CMB candidate subject preselection in large datasets, when manual segmentation is unavailable. While the preselected subjects could be used for manual labelling for research purposes, from a clinical point of view the pipeline could be used to flag images as likely to contain CMBs. This could be further used in determining a preliminary CMB or small vessel disease (SVD) score as done in Staals et al. (2014).

Concluding, we proposed a learning-based method for subject-level preselection of CMB candidate subjects in large datasets. The preselected subjects could then be manually segmented and used for further analysis and characterisation of CMBs. Our method provided accurate preselection of CMB candidate subjects on various datasets consisting of T2*-GRE and SWI images with subject-level TPR, specificity and accuracy values >90, >80, and >85%, respectively. Also, our method is computationally efficient, and provided greater performance when compared to other methods using non-imaging factors and thresholding methods for obtaining CMB lesion counts from the imaging data. Our pipeline shows good generalisability across across various datasets providing subject-level accuracy >80%, and even >85% when applied to datasets with the same modality. The future direction of this work would be to improve the detection of CMBs at the lesion-level using deep learning and increase the model generalisability across different modalities. Also, another potential avenue of research would be to provide automated ratings of CMBs using their size and count to provide information that is consistent with clinical rating scales such as MARS scale.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because requests for data from the OXVASC Study will be considered by PR in line with data protection laws. The TICH-2 MRI sub-study data can be shared with bona fide researchers and research groups on written request to the sub-study PI Prof Rob Dineen. Proposals will be assessed by the PI (with advice from the TICH-2 trial Steering Committee if required) and a Data Transfer Agreement will be established before any data are shared. The UK Biobank datasets are available to researchers through an open application via https://www.ukbiobank.ac.uk/register-apply/. Requests to access the datasets should be directed to rob.dineen@nottingham.ac.uk, peter.rothwell@ndcn.ox.ac.uk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Local Ethics Committee (Research Ethics Committee reference number: 05/Q1604/70) for the OXVASC data. Written informed consent was obtained from all participants. UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) to obtain and disseminate data and samples from the participants (http://www.ukbiobank.ac.uk/ethics/), and these ethical regulations cover the work in this study. Written informed consent was obtained from all participants. The TICH-2 trial obtained ethical approval from East Midlands (Nottingham 2) NHS Research Ethics Committee (Reference: 12/EM/0369) and the amendment to allow the TICH2 MRI sub-study was approved in April 2015 (amendment number SA02/15). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

VS: conceptualisation, methodology, software, validation, formal analysis, investigation, visualisation, and writing–original draft. CA: investigation, methodology, software, writing–review, and editing. GZ: supervision, resources, writing–review, and editing. RD and SS: resources, data curation, writing–review, and editing. PR and NS: resources, writing–review, and editing. DA: resources, data curation, investigation, writing–review, and editing. DT and HM: resources, data curation, investigation, writing–review, and editing. KM, FA-A, and ID: methodology, software, writing–review, and editing. MJ: conceptualization, methodology, supervision, writing–review, editing, funding acquisition, and project administration. LG: conceptualisation, methodology, data curation, supervision, writing–review, editing, and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2021.777828/full#supplementary-material

REFERENCES

Ayaz, M., Boikov, A. S., Haacke, E. M., Kido, D. K., and Kirsch, W. M. (2010). Imaging cerebral microbleeds using susceptibility weighted imaging: one step toward detecting vascular dementia. *J. Magn. Reson. Imaging* 31, 142–148. doi: 10.1002/jmri.22001

Barnes, S. R., Haacke, E. M., Ayaz, M., Boikov, A. S., Kirsch, W., and Kido, D. (2011). Semiautomated detection of cerebral microbleeds in magnetic resonance images. *Magn. Reson. Imaging* 29, 844–852. doi: 10.1016/j.mri.2011.02.028

Bian, W., Hess, C. P., Chang, S. M., Nelson, S. J., and Lupo, J. M. (2013). Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images. *Neuroimage Clin.* 2, 282–290. doi: 10.1016/j.nicl.2013.01.012

Charidimou, A., and Werring, D. J. (2011). Cerebral microbleeds: detection, mechanisms and clinical challenges. *Future Neurol.* 6, 587–611. doi: 10.2217/fnl.11.42

Chen, H., Yu, L., Dou, Q., Shi, L., Mok, V. C., and Heng, P. A. (2015). "Automatic detection of cerebral microbleeds via deep learning based 3D feature representation," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)* (Brooklyn, NY: IEEE), 764–767.

Chen, Y., Villanueva-Meyer, J. E., Morrison, M. A., and Lupo, J. M. (2018). Toward automatic detection of radiation-induced cerebral microbleeds using a 3d deep residual network. *J. Digit. Imaging* 32, 766–772. doi: 10.1007/s10278-018-0146-z

De Bresser, J., Brundel, M., Conijn, M., Van Dillen, J., Geerlings, M., Viergever, M., et al. (2013). Visual cerebral microbleed detection on 7T MR imaging: reliability and effects of image processing. *Am. J. Neuroradiol.* 34, E61–E64. doi: 10.3174/ajnr.A2960

Dineen, R. A., Pszczolkowski, S., Flaherty, K., Law, Z. K., Morgan, P. S., Roberts, I., et al. (2018). Does tranexamic acid lead to changes in MRI measures of brain tissue health in patients with spontaneous intracerebral haemorrhage? protocol for a MRI substudy nested within the double-blind randomised controlled TICH-2 trial. *BMJ Open* 8:e019930. doi: 10.1136/bmjopen-2017-019930

Dou, Q., Chen, H., Yu, L., Shi, L., Wang, D., Mok, V. C., et al. (2015). "Automatic cerebral microbleeds detection from MR images via independent subspace analysis based hierarchical features," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Milan: IEEE), 7933–7936.

Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., et al. (2016). Automatic detection of cerebral microbleeds from MR images via 3d convolutional neural networks. *IEEE Trans. Med. Imaging* 35, 1182–1195. doi: 10.1109/TMI.2016.2528129

Fazlollahi, A., Meriaudeau, F., Giancardo, L., Villemagne, V. L., Rowe, C. C., Yates, P., et al. (2015). Computer-aided detection of cerebral microbleeds in susceptibility-weighted imaging. *Compute. Med. Imaging Graph.* 46, 269–276. doi: 10.1016/j.compmedimag.2015.10.001

Fazlollahi, A., Meriaudeau, F., Villemagne, V. L., Rowe, C. C., Yates, P., Salvado, O., et al. (2014). "Efficient machine learning framework for computer-aided detection of cerebral microbleeds using the radon transform," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (Beijing: IEEE), 113–116.

Förstner, W. (1994). "A framework for low level feature extraction," in *European Conference on Computer Vision* (Stockholm: Springer), 383–394.

Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). "Multiscale vessel enhancement filtering," in *International Conference on Medical IMAGE Computing and Computer-Assisted Intervention* (Boston, MA: Springer), 130–137.

Ghafaryasl, B., van der Lijn, F., Poels, M., Vrooman, H., Ikram, M. A., Niessen, W. J., et al. (2012). "A computer aided detection system for cerebral microbleeds in brain MRI," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)* (Barcelona: IEEE), 138–141.

Gouw, A. A., Seewann, A., Van Der Flier, W. M., Barkhof, F., Rozemuller, A. M., Scheltens, P., et al. (2011). Heterogeneity of small vessel disease: a systematic review of MRI and histopathology correlations. *J. Neurol. Neurosurg. Psychiatry* 82, 126–135. doi: 10.1136/jnnp.2009.204685

Greenberg, S. M., Nandigam, R. K., Delgado, P., Betensky, R. A., Rosand, J., Viswanathan, A., et al. (2009a). Microbleeds versus macrobleeds: evidence for distinct entities. *Stroke* 40:2382. doi: 10.1161/STROKEAHA.109.548974

Greenberg, S. M., Vernooij, M. W., Cordonnier, C., Viswanathan, A., Salman, R. A.-S., Warach, S., et al. (2009b). Cerebral microbleeds: a guide to detection and interpretation. *Lancet Neurol.* 8, 165–174. doi: 10.1016/S1474-4422(09)70013-4

Gregoire, S., Chaudhary, U., Brown, M., Yousry, T., Kallis, C., Jäger, H., et al. (2009). The Microbleed Anatomical Rating Scale (MARS): reliability of a tool to map brain microbleeds. *Neurology.* 73, 1759–1766. doi: 10.1212/WNL.0b013e3181c34a7d

Horita, Y., Imaizumi, T., Niwa, J., Yoshikawa, J., Miyata, K., Makabe, T., et al. (2003). Analysis of dot-like hemosiderin spots using brain dock system. *No Shinkei Geka.* 31, 263–267. Available online at: https://europepmc.org/article/med/12684979

Huang, R., Noble, J. A., and Namburete, A. I. (2018). "Omni-supervised learning: scaling up to large unlabelled medical datasets," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Granada: Springer).

Jeerakathil, T., Wolf, P. A., Beiser, A., Hald, J. K., Au, R., Kase, C. S., et al. (2004). Cerebral microbleeds: prevalence and associations with cardiovascular risk factors in the Framingham Study. *Stroke* 35, 1831–1835. doi: 10.1161/01.STR.0000131809.35202.1b

Kuijf, H. J., Brundel, M., de Bresser, J., van Veluw, S. J., Heringa, S. M., Viergever, M. A., et al. (2013). Semi-automated detection of cerebral microbleeds on 3.0 T MR images. *PLoS ONE* 8:e66610. doi: 10.1371/journal.pone.0066610

Kuijf, H. J., de Bresser, J., Geerlings, M. I., Conijn, M. M., Viergever, M. A., Biessels, G. J., et al. (2012). Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform. *Neuroimage* 59, 2266–2273. doi: 10.1016/j.neuroimage.2011.09.061

Liu, S., Utriainen, D., Chai, C., Chen, Y., Wang, L., Sethi, S. K., et al. (2019). Cerebral microbleed detection using susceptibility weighted imaging and deep learning. *Neuroimage* 198, 271–282. doi: 10.1016/j.neuroimage.2019.05.046

Loy, G., and Zelinsky, A. (2002). "A fast radial symmetry transform for detecting points of interest," in *European Conference on Computer Vision* (Copenhagen: Springer), 358–368.

Morrison, M. A., Payabvash, S., Chen, Y., Avadiappan, S., Shah, M., Zou, X., et al. (2018). A user-guided tool for semi-automated cerebral microbleed detection and volume segmentation: evaluating vascular injury and data labelling for machine learning. *Neuroimage* 20, 498–505. doi: 10.1016/j.nicl.2018.08.002

Nandigam, R., Viswanathan, A., Delgado, P., Skehan, M., Smith, E., Rosand, J., et al. (2009). MR imaging detection of cerebral microbleeds: effect of susceptibility-weighted imaging, section thickness, and field strength. *Am. J. Neuroradiol.* 30, 338–343. doi: 10.3174/ajnr.A1355

Nannoni, S., Ohlmeier, L., Brown, R. B., Morris, R. G., MacKinnon, A. D., Markus, H. S., et al. (2021). Cognitive impact of cerebral microbleeds in patients with symptomatic small vessel disease. *Int. J. Stroke.* doi: 10.1177/17474930211012837. [Epub ahead of print].

Nishikawa, T., Ueba, T., Kajiwara, M., Fujisawa, I., Miyamatsu, N., and Yamashita, K. (2009). Cerebral microbleeds predict first-ever symptomatic cerebrovascular events. *Clin. Neurol. Neurosurg.* 111, 825–828. doi: 10.1016/j.clineuro.2009.08.011

Poels, M. M., Vernooij, M. W., Ikram, M. A., Hofman, A., Krestin, G. P., van der Lugt, A., et al. (2010). Prevalence and risk factors of cerebral microbleeds: an update of the Rotterdam scan study. *Stroke* 41(10_suppl_1):S103–S106. doi: 10.1161/STROKEAHA.110.595181

Roob, G., Lechner, A., Schmidt, R., Flooh, E., Hartung, H.-P., and Fazekas, F. (2000). Frequency and location of microbleeds in patients with primary intracerebral hemorrhage. *Stroke* 31, 2665–2669. doi: 10.1161/01.STR.31.11.2665

Rothwell, P., Coull, A., Giles, M., Howard, S., Silver, L., Bull, L., et al. (2004). Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). *Lancet* 363, 1925–1933. doi: 10.1016/S0140-6736(04)16405-2

Seghier, M. L., Kolanko, M. A., Leff, A. P., Jäger, H. R., Gregoire, S. M., and Werring, D. J. (2011). Microbleed detection using automated segmentation (MIDAS): a new method applicable to standard clinical MR images. *PLoS ONE* 6:e17547. doi: 10.1371/journal.pone.0017547

Shoamanesh, A., Kwok, C., and Benavente, O. (2011). Cerebral microbleeds: histopathological correlation of neuroimaging. *Cerebrovasc. Dis.* 32, 528–534. doi: 10.1159/000331466

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062

Sprigg, N., Flaherty, K., Appleton, J. P., Salman, R. A.-S., Bereczki, D., Beridze, M., et al. (2018). Tranexamic acid for hyperacute primary intracerebral haemorrhage (TICH-2): an international randomised, placebo-controlled, phase 3 superiority trial. *Lancet* 391, 2107–2115. doi: 10.1016/S0140-6736(18)31033-X

Staals, J., Makin, S. D., Doubal, F. N., Dennis, M. S., and Wardlaw, J. M. (2014). Stroke subtype, vascular risk factors, and total MRI brain small-vessel disease burden. *Neurology* 83, 1228–1234. doi: 10.1212/WNL.0000000000000837

Takashima, Y., Mori, T., Hashimoto, M., Kinukawa, N., Uchino, A., Yuzuriha, T., et al. (2011). Clinical correlating factors and cognitive function in community-dwelling healthy subjects with cerebral microbleeds. *J. Stroke Cerebrovasc. Dis.* 20, 105–110. doi: 10.1016/j.jstrokecerebrovasdis.2009.11.007

Tsushima, Y., Tanizaki, Y., Aoki, J., and Endo, K. (2002). MR detection of microhemorrhages in neurologically healthy adults. *Neuroradiology* 44, 31–36. doi: 10.1007/s002340100649

van den Heuvel, T., Van Der Eerden, A., Manniesing, R., Ghafoorian, M., Tan, T., Andriessen, T., et al. (2016). Automated detection of cerebral microbleeds in patients with traumatic brain injury. *Neuroimage* 12, 241–251. doi: 10.1016/j.nicl.2016.07.002

Vernooij, M., van der Lugt, A., Ikram, M. A., Wielopolski, P., Niessen, W., Hofman, A., et al. (2008). Prevalence and risk factors of cerebral microbleeds: the Rotterdam Scan Study. *Neurology* 70, 1208–1214. doi: 10.1212/01.wnl.0000307750.41970.d9

Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., et al. (2013). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* 12, 822–838. doi: 10.1016/S1474-4422(13)70124-8

Werring, D. J., Frazer, D. W., Coward, L. J., Losseff, N. A., Watt, H., Cipolotti, L., et al. (2004). Cognitive dysfunction in patients with cerebral microbleeds on T2*-weighted gradient-echo MRI. *Brain* 127, 2265–2275. doi: 10.1093/brain/awh253

Won Seo, S., Hwa Lee, B., Kim, E.-J., Chin, J., Sun Cho, Y., Yoon, U., et al. (2007). Clinical significance of microbleeds in subcortical vascular dementia. *Stroke* 38, 1949–1951. doi: 10.1161/STROKEAHA.106.477315

Yates, P., Villemagne, V., Ellis, K., Desmond, P., Masters, C., and Rowe, C. (2014). Cerebral microbleeds: A review of clinical, genetic, and neuroimaging associations. *Front. Neurol.* 4:205. doi: 10.3389/fneur.2013.00205

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57. doi: 10.1109/42.906424

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. *Graphics Gems* 474–485. doi: 10.1016/B978-0-12-336156-1.50061-6