Artificial Intelligence In Medicine 147 (2024) 102719

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Research paper

Predicting stroke outcome: A case for multimodal deep learning methods with tabular and CT Perfusion data

Balázs Borsos [a,b,c,1], Corinne G. Allaart [a,b,1,*], Aart van Halteren [a,c]

[a] Vrije Universiteit Amsterdam, De Boelelaan 1105, Amsterdam, 1081 HV, Netherlands
[b] St. Antonius Ziekenhuis, Koekoekslaan 1, Nieuwegein, 3435 CM, Netherlands
[c] Philips Research, Hightech Campus 34, Eindhoven, 5656 AE, Netherlands

ARTICLE INFO

Keywords:
Deep learning
Multimodal data
Acute ischemic stroke
CT perfusion

ABSTRACT

Motivation: Acute ischemic stroke is one of the leading causes of morbidity and disability worldwide, often followed by a long rehabilitation period. To improve and personalize stroke rehabilitation, it is essential to provide a reliable prognosis to caregivers and patients. Deep learning techniques might improve the predictions by incorporating different data modalities. We present a multimodal approach to predict the functional status of acute ischemic stroke patients after their discharge based on tabular data and CT perfusion imaging.

Methods: We conducted experiments on tabular, imaging, and multimodal deep learning architectures to predict dichotomized mRS scores 3 months after the event. The dataset was collected from a Dutch hospital and includes 98 CVA patients with a visible occlusion on their CT perfusion scan. Tabular data is based on the Dutch Acute Stroke Audit data, and imaging data consists of summed-up CT perfusion maps.

Results: On the tabular data, TabNet outperformed our baselines with an AUC of 0.71, while ResNet-10 on the imaging data performed comparably with an AUC of 0.70. Our implementation of the multimodal DAFT architecture outperforms baselines as well as comparable studies by achieving an 0.75 AUC, and 0.80 F1 score. This was achieved with a final model of less than a hundred thousand optimizable parameters, and a dataset less than half the size of reference papers.

Conclusion: Overall, we demonstrate the feasibility of predicting the functional outcome for ischemic stroke patients and the usability of multimodal deep learning architectures for this purpose.

1. Introduction

To this day, acute ischemic stroke (AIS) is one of the leading causes of morbidity and disability worldwide with over 12.2 million new strokes each year [1]. It is estimated that the global cost of stroke is exceeding US$ 721 billion and it remains the second-leading cause of death and the third-leading cause of death and disability combined [1]. Acute ischemic strokes (caused by vessel occlusions usually due to a blood clot, creating a disruption in the blood flow) are the most common globally, accounting for at least 70% of all strokes [2]. Large-Vessel Occlusions (LVOs) amount to 29.3% of these AIS, of which most occur in the anterior circulation [3]. We focus on disruptions caused by LVOs, as they are associated with poor functional outcome and impose a 4.5-fold increase in mortality compared to other types of ischemic strokes [4]. To achieve optimal recovery, it is highly important to have a prompt diagnosis, effective communication, and fast treatment, which also makes it a great candidate for the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques [5]. ML techniques have been shown to achieve significant success in helping medical professionals with fast and accurate diagnoses [6]. Apart from clinical information, which describes the state of the patient as a whole, doctors mainly focus on imaging techniques to get a better understanding of the situation [7]. Traditionally, Non-Contrast Computed Tomography (NCCT) is recommended by the American Heart Association's Stroke Council as the first modality of choice for stroke investigation [8]. In this context, a few studies already looked at the potential of Machine Learning models for predicting clinical outcomes for stroke patients [9,10]. However, these studies mainly experimented with models that are based on available clinical data and NCCT imaging features prior to treatment. Yet, it is limited in capability: deterioration is not visible on images up to 6 h with ischemic stroke [11]. There are now other, more advanced imaging techniques with widespread availability in hospitals to obtain deeper insight into the patient's current situation.

One of these techniques is called Computed Tomography Perfusion (CTP) which is obtained via administering a contrast material to the

* Corresponding author at: Vrije Universiteit Amsterdam, De Boelelaan 1105, Amsterdam, 1081 HV, Netherlands.
E-mail address: c.g.allaart@vu.nl (C.G. Allaart).
[1] These authors contributed equally.

https://doi.org/10.1016/j.artmed.2023.102719
Received 10 January 2023; Received in revised form 12 November 2023; Accepted 14 November 2023
Available online 15 November 2023

patient through intravenous injection and allows measuring of the extent of irreversibly injured brain tissue called the ischemic core, and the potentially salvageable but hypoperfused ischemic penumbra [11]. A favorable outcome is associated with a large ischemic penumbra in the setting of successful recanalization, while it may predict an unfavorable outcome in the absence of recanalization, as a large amount of potentially salvageable brain tissue is going to infarct [12]. A study by Hopyan et al. [13] also showed that an incremental stroke protocol that includes CTP increases diagnostic performance for stroke diagnosis.

Many problem domains are naturally based on multiple modalities, such as medicine [14,15] (e.g., physicians usually make diagnoses based on different kinds of medical images and clinical data), sentiment analysis [16,17] (e.g., in face-to-face interactions, the understanding of sentiment is much more nuanced as we can also interpret body language in addition to spoken language), and many others. Therefore, there has been an increased focus on combining data of different natures recently to create more robust representations and further improve predictive performance. By using information from different representations of the same subject, a more elaborate picture of the problem at hand can be constructed [18].

To improve upon stroke care and to be able to give a reliable prognosis to caregivers and patients, it is beneficial to accurately predict the functional outcome. Utilizing ML techniques not only provides helpful predictions for medical professionals to support clinical decisions but also allows for a more standardized way of diagnosis. In line with this, we propose the use of a multimodal neural network to predict the functional status of ischemic stroke patients based on the modified Rankin Scale score (mRS) [19], which is generally used in clinical trials measuring functional independence and recovery. To provide a fast prognosis, we only use data that is available up until treatment. To achieve this, we investigate the applicability of a novel general-purpose module for computer vision tasks, called the Dynamic Affine Feature Map Transform (DAFT), for the first time on the task of predicting the functional outcome. This module is aimed at effectively combining 3D images with tabular information and has been proven to outperform traditional Convolutional Neural Networks in diagnosis and time-to-dementia prediction regarding Alzheimer's disease [20]. We also plan to evaluate the predictive performance of clinical and imaging characteristics obtained prior to treatment. To do so, first, we propose to implement a state-of-the-art neural network for tabular deep learning called TabNet, which offers interpretability and performance improvements. To the best of our knowledge, no previous studies have tried that before. Secondly, for the imaging modality, we implement a ResNet-type architecture and investigate the utility of transfer learning on our dataset.

2. Related work

Predicting the functional outcome of a stroke patient encompasses a complex problem with many modalities that must be taken into account to arrive at a final prediction. Functional outcome in previous studies is usually measured in Modified Ranking Scale (MrS) scores [21], a scale from 0 (no residual symptoms) to 5 (severe disability), with 6 signifying the patient has since passed away. Due to non-uniform distributions and the difficulty of predicting the functional status, some papers simplified the problem by dichotomizing the mRS. We will look at studies that have aimed to utilize medical imaging (MRI, CT, or PCT), to obtain a prediction of functional outcome, an overview is shown in Table 1. One of the first notable attempts was done by Choi et al. [22], who proposed an ensemble of networks for the tasks of lesion and clinical outcome prediction on the ISLES 2016 dataset. For predicting the mRS score, they combined a CNN and a logistic regression model as an ensemble. They randomly chose between the output of the two models, which resulted in the best performance and ranked second in the challenge. The evaluation measure was the average of absolute errors between the true and predicted scores with a score of 1.10 ± 0.70 for the ensemble model.

Several papers only used imaging data to predict functional status. Hilbert et al. [23] proposed a data-efficient DL method, based on Structured Receptive Field Neural Networks (SRFNN), a specific type of CNN to reduce model parameters and overfitting on limited data sets. Their model predicts the mRS outcome using full-sized 2D CT images. They reached comparable performance to conventional CNNs while reducing the total number of weights by almost 50%. Osama et al. [24] proposed to use few-shot learning with a parallel multi-parametric feature-embedded Siamese Neural Network (SNN) [25] to predict the mRS score based on the ISLES 2017 challenge dataset [26]. To train the Siamese networks, the MRI images were fed as pairs into 2 parallel CNNs that share the same weights. This results in an averaged accuracy of 0.37 on each mRS class using leave-one-out cross-validation testing. Nishi et al. [27] used only Diffusion-Weighted Imaging (DWI) MRI data of 250 patients as input for training a multioutput CNN, to both predict the dichotomized outcome and segment the ischemic core lesion. Data augmentation was used to improve the model performance due to the small sample size. The reported performance of this method was 0.724 in accuracy and 0.81 in AUC with five-fold cross-validation on the training set while scoring 0.654 and 0.73 respectively for accuracy and AUC on the test set. This shows a big difference between train and test sets, which hints at overfitting, possibly due to the data augmentation. Moreover, these papers only used imaging data, while combining clinical and imaging data can lead to an increase in predictive performance.

Brugnara et al. [28] assessed the predictive performance and relative importance of clinical, multimodal imaging, and angiographic characteristics to predict the clinical outcome after endovascular treatment. They built four Gradient Boosting Classifiers with different types of input features, gradually progressing from clinical and regular CT characteristics through multimodal imaging to even post-interventional parameters. According to their findings, post-interventional parameters had the most predictive power, while CT perfusion had limited importance on top of conventional imaging characteristics. Hamann et al. [29] also investigated whether imaging variables from Magnetic Resonance (MR) scans would improve individual prediction of functional outcome after early (<6 h) endovascular treatment. Their results were also in line with Brugnara et al., outcome prediction only improved slightly when imaging features were added. They made use of Random Forest (RF) classifiers with MR images instead of CTP scans and obtained quantitative imaging features via automatic volume segmentation based on thresholds. While these studies show limited addition of imaging data to clinical data, they do not use DL techniques, thereby showing that DL might be essential to improve functional status predictions with imaging data.

Bacchi et al. [30] did use DL techniques. They used clinical data and Non-Contrast Computer Tomography (NCCT) images to predict the outcome of thrombolysis based on 204 patients. Their best-performing model is a branched deep learning architecture concatenating a CNN branch with a Fully Connected Network (FCN), which reached a score of 0.74 in accuracy, 0.69 in F1 score, and 0.75 AUC. They highlight the need for larger datasets and the incorporation of CT angiograms and perfusion imaging. Samak et al. [31] presented a novel approach to predict the functional status from multimodal data working with the MR CLEAN dataset [32], which contains NCCT images and clinical metadata. Their main contribution is the use of an attention mechanism that extracts features both spatially and channel-wise, resulting in superior performance compared to regular ResNet-type architectures and the results of Bacchi et al. [30]. They reach a 0.62 F1 score and 0.75 AUC in dichotomous mRS scores and 0.35 classification accuracy in individual mRS scores. These papers show that using deep learning techniques on multimodal data can lead to better-performing predictive models.

Table 1
An overview of the best results of previous work on individual mRS score prediction.

Ref.	Model	Img. mod.	Dichotomized			Non-dich.		Dataset
			Acc.	AUC	F1	Acc.	AAE	size
[28]	GBC	CT(P)	0.72	0.75	–	–	–	246
[29]	RF	MRI	–	0.68	–	–	–	222
[23]	SRFNN	2D CT	–	0.69	–	–	–	772
[30]	Naive Late Fusion	CT	0.74	0.75	0.69	–	–	204
[27]	CNN	MRI	0.72	0.81	–	–	–	324
[31]	IMF Block	NCCT	0.77	0.75	0.62	0.35	–	500
[22]	Naive Late Fusion	MRI	–	–	–	–	1.1	35
[24]	SNN	MRI	–	–	–	0.37		43

3. Contributions

In this paper, we propose a method for functional outcome prediction based on the tabular and imaging data, and our main contributions are summarized as:

1. The combination of imaging and tabular clinical data in a novel multimodal deep learning architecture called DAFT [20].
2. The evaluation of three data fusion techniques to arrive at the best performing predictor of functional status.
3. The utilization of a state-of-the-art deep tabular learning network called TabNet on the tabular patient data, that offers interpretability and performance improvements.
4. The investigation of the utility of transfer learning on our dataset.
5. The evaluation of the predictive performance of clinical and CT Perfusion imaging characteristics both individually and combined.

4. Methods

We implemented a network that can combine multimodal features in a novel way based on the Dynamic Affine Feature Map Transform (DAFT) from Pölsterl et al. [20]. To put our results into context, we also introduced different machine learning models for each modality. To predict the functional status, we built models to predict good outcome and treatment success (scores 0–2), or poor outcome and treatment failure (scores 3–6). To evaluate the impact of our proposals and for comparison with previous studies we structured our experiments in the following way:

- First we implemented previously utilized Machine Learning techniques on the clinical data as a baseline. This includes Random Forests [33], Gradient Boosted Classifiers (GBC) [34], and a recently developed DL model for tabular data called TabNet [35].
- Secondly we established an imaging baseline using a ResNet-10 architecture and investigated the utility of transfer learning on this network.
- Thirdly we examined the utility of different data fusion models by providing a baseline with naive late and hybrid fusion, and evaluated the utility of affine transformations based on DAFT. Experiments on a single modality – apart from serving as a benchmark – are also aimed to help and illustrate the process of how we made certain decisions and arrived at the final multimodal model.

4.1. Dataset

We collected patient data from St. Antonius Hospital from October 1, 2018 to December 31, 2019. Clinical data was from the Dutch Acute Stroke Audit (DASA) [36], a national registry with acute ischemic stroke and intracranial hemorrhage patients. In this period, 1003 cases were recorded in the DASA. However, only 799 of them were classified as infarct cases. Further exploration of the patient list revealed that CT perfusion scans were available for 374 patients and only 104 of them had ischemia with both CTP scans, mRS scores, and radiologist reports available. Fig. 1 denotes the composition of the received patient list. This study was approved by the board of St. Antonius Hospital (under number Z20.016) after Medical Ethics Commission Utrecht (MEC-U) issued a non-WMO statement (under number AW22.012/W20.034).

Table 2
Patient characteristics.

Categorical Variables (%)		Numerical Variables (avg(std))	
Referred	14	Door to Needle (min)	14.9 (23.0)
Wake-up	23	Door to Groin (min)	60.9 (85.9)
Gender(m)	51	NIHSS score	11.5 (5.7)
Intravenous thrombolysis	38	Penumbra (volume)	82.1 (64.1)
Intra-arterial thrombolysis	61	Penumbra (%)	8.6 (6.1)
Atrial Fibrillation	14	Core (volume)	32.0 (39.1)
Occlusion in M1	57	Core (%)	3.3 (3.8)
Occlusion in M2	34	Age (years)	73.0 (12.5)
Occlusion in M3	8	Time to hospital (min)	211.2 (388.8)

We chose to use CTP scans for this study, as it has been argued that they would contain predictive information for the functional outcome of CVA [30], yet no other study has investigated deep learning on CTP imaging. We excluded patients who had faulty scans, such as movement during scanning time or incorrect Cerebral Blood Volume readings, which resulted in a dataset of 98 cases. When patients had multiple CTP scans on the same day, here we always extracted the latter ones if they were still before treatment, as most of the time a second scan is requested because of an insufficient first reading. Multiple scanner types were used in different locations, resulting in different reading sizes for patients, which were standardized as described in the data pre-processing section.

Our clinical dataset included the following features: age, gender, wake-up (whether the patient experienced symptoms at waking up), arterial fibrillation (binary), whether the patient was referred from another hospital, National Institutes of Health Stroke Scale (NIHSS) score at presentation, Time-To-Hospital (TTH), whether treated via intravenous (IVT) or intra-arterial (IAT) thrombolysis, Door-To-Needle time (DTN), Door-To-Groin time (DTG), region of occlusion, the volume and the percentage of the ischemic core and penumbra. We specifically chose these features as they were already part of automated processes in the hospital, and would not require extra data collection by neurologists or radiologists. Details can be found in Table 2.

4.2. Data preprocessing

4.2.1. Tabular data

For all clinical models, continuous variables were standardized by scikit-learn's StandardScaler function. If the Door-To-Needle (DTN) or Door-To-Groin (DTG) features were missing, it was imputed by 0. Other missing values were handled with scikit-learn's IterativeImputer, which is a multivariate imputer that estimates each missing value by modeling them as a function of other features in a round-robin fashion. The function used was Bayesian ridge regression [37].

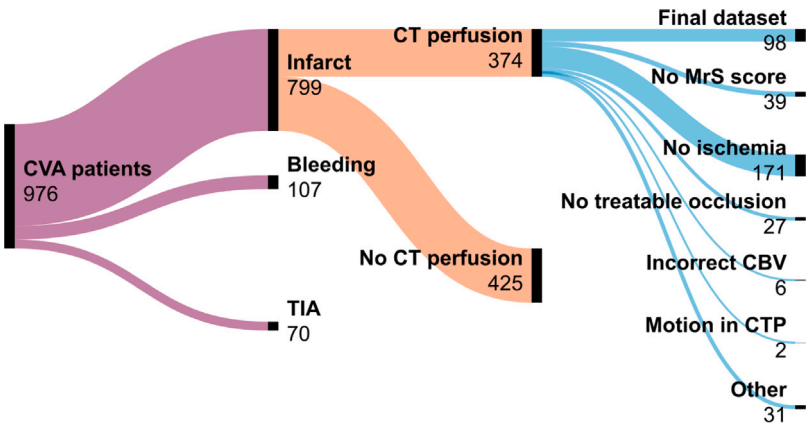

Fig. 1. Flowchart of patient selection.

4.2.2. Imaging data

The CT perfusion dataset was heterogeneous due to the use of different scanners, containing (1) 16 slices repeated 50 times; (2) 16 slices repeated 60 times; and (3) 8 slices repeated 50 times not containing the upper part of the intracranial area, so they were padded to achieve a standard of 16 slices, instead of resampling the voxel space. Each slice had a resolution of 512 × 512 pixels, which we downsampled to 128 × 128, arriving at the final dimension of 128 × 128 × 16.

Raw feature maps As perfusion scans are obtained via administering a contrast agent to the bloodstream and taking repeated readings with a CT scanner, the resulting scan can be dealt with in multiple ways. We treated the time dimension to reduce noise, by obtaining different feature maps by (1) averaging and (2) summing up over the temporal dimension. The first allows us to have a scan with reduced noise levels, while the second results in more defined perfusion levels, by taking the sum of areas where the contrast agent is in higher concentration.

Computed CT Perfusion maps To obtain parametric perfusion maps of cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT), and time-to-peak (TTP), raw CTP scans were post-processed using Philips IntelliSpace Portal 11.0, Brain Perfusion application software [38]. It is capable of automated registration, segmentation, and motion correction and prepares summary maps based on a perfusion delay-sensitive algorithm. All scans were processed by default settings with motion correction and filtering enabled. Ischemic core and penumbra volumes were also extracted for the clinical and multimodal models. Ischemic core was defined as a relative MTT > 150% and an absolute CBV < 2.0 ml/100 g and penumbra as a relative MTT > 150% and an absolute CBV > 2.0 ml/100 g. The images were cropped removing headers and scales. Fig. 2 shows an example of the output of the 6 feature maps (2 raw and 4 computed maps). To try out the utility of combining the 6 maps, we made the parametric computed perfusion maps greyscale (to reduce dimensionality), then all 6 maps were stacked together as channels, resulting in the final input dimension of 6 × 128 × 128 × 16.

4.2.3. Dichotomization of outcomes

The distribution of mRS scores contains large class imbalances. Individual scores should share approximately the same number of instances to be able to train models that could learn to differentiate between them. Up-sampling of minority [39] or down-sampling of majority [40] classes was not possible in our case as the dataset was too limited. Instead, we decided to dichotomize the clinical outcomes by defining a favorable outcome as a patient who has an mRS score between 0 and 2 (requiring no assistance), and an unfavorable outcome as an mRS score above 2 (needing assistance). Fig. 3 shows in detail the distribution of individual and dichotomized outcomes.

4.3. Tabular architectures

To predict the clinical outcome of ischemic stroke patients based on clinical patient characteristics, previous work made use of Multi-Layer Perceptrons and ensemble techniques, such as Random Forests [33] and Gradient Boosted Models [34]. For comparison with previous studies, we also implemented a Random Forest and a Gradient Boosted Classifier. However, instead of a vanilla MLP, we implement a recently developed method called TabNet [35]. TabNet is a novel deep learning method that aims to harvest the power of DL for tabular data with an interpretable multi-step deep tabular data learning architecture based on Transformers [41]. TabNet uses sequential attention to choose which features to reason from at each decision step – essentially mimicking the behavior of decision trees – enabling local and global interpretability and more efficient learning as the learning capacity is used for the most powerful features. Moreover, it can handle raw tabular data without any preprocessing. TabNet translates the local and global interpretability as feature importances.

4.4. Imaging architectures

To establish an imaging baseline, we employ a type of ResNet architecture, but a smaller version of it called ResNet-10, based on the implementation of Pölsterl et al. [20]. It is constructed by removing half of the Residual Units from ResNet-18 making it less complex with fewer parameters to train, which is desirable with a small dataset. As the performance of neural networks is heavily affected by the volume of training data, another approach to improve model performance is by utilizing transfer learning. Networks pre-trained on large datasets can help to converge faster and improve predictive performance and generalizing ability [42], yet it is very challenging to build up a sufficiently large dataset for medical imaging due to privacy concerns. We therefore also implement the ResNet-10 of MedicalNet [43] multiple pre-trained models on 3D medical imaging of up to 23 different medical image segmentation tasks.

4.5. Multimodal architectures

One of the ways to process multimodal data is to use unimodal architectures with a different feature extractor for each modality and combine their representations in the network using multilayer perceptrons (MLP) to create the output predictions. In this situation, data fusion is the unique aspect of learning where information from different data sources needs to be combined. This process may happen right after the input data is presented (*Early fusion*), before the final classification (*Late fusion*), or multiple times in the middle (*Hybrid fusion*) [18]. By using information from different representations of the same subject, a more elaborate picture of the problem at hand can be constructed [18],

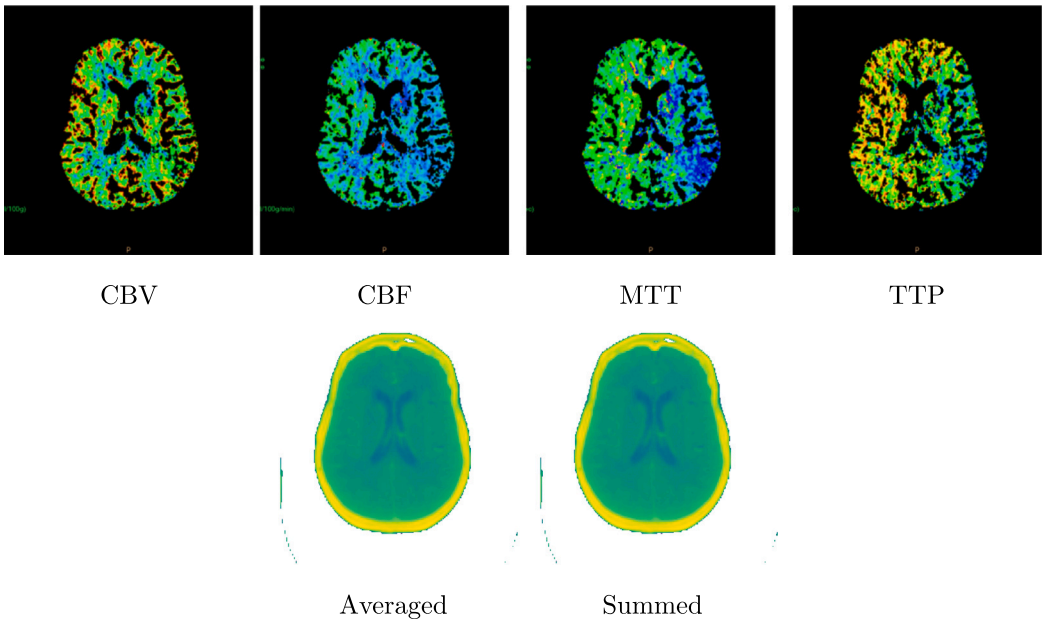

Fig. 2. A random sample from the dataset, showing an example of the parametric perfusion maps. The shown maps are (left to right). Top: cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT) and time-to-peak (TTP). Bottom: Averaged and Summed-up perfusion maps.

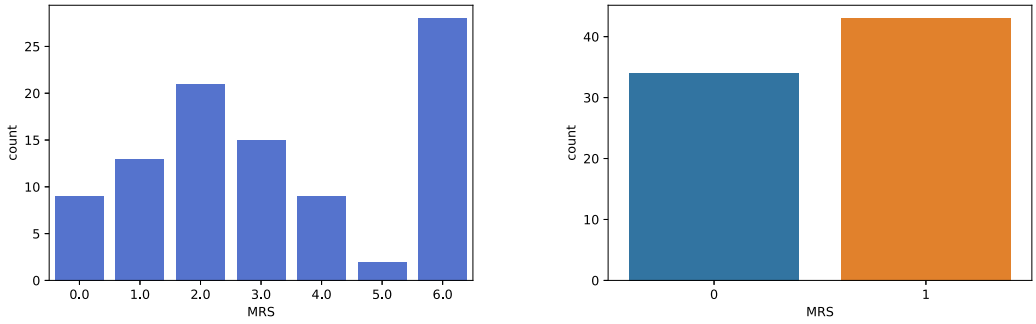

Fig. 3. The distribution of the individual mRS scores, and the favorable (mRS 0–2) and unfavorable (mRS > 2) outcomes.

however, each of these strategies can suffer from shortcomings in their capabilities. Both early and hybrid fusion falls short in utilizing transfer learning effectively [44] while naive late fusion tend to focus on the imaging dimension during training as there is a great dimensionality mismatch between the two data sources, which can result in only marginally better performance than using image data alone [20].

Unlike previous work, where only naive late fusion was applied to predict the functional outcome in multimodal networks, our approach aims to utilize affine transformations to fuse clinical and imaging data in a novel way. We base our network architecture on a recent study by Pölsterl et al. [20], where they successfully utilized a unique network type to diagnose Alzheimer's disease and predict time-to-dementia. They introduced the Dynamic Affine Feature Map Transform (DAFT) block in a ResNet-type architecture. It is a generic module that can be integrated into any CNN and establishes a two-way exchange of information between high-level concepts learned from the 3D image and the tabular biomarkers by dynamically rescaling and shifting the feature maps of a convolutional layer based on clinical data. It is important to have a level exchange of information between the two modalities, as tabular information comprises demographics and summary measures that describe the patient's state as a whole. Therefore, they proposed to affinely transform the output of a convolutional layer in the last residual block, where the network describes the image in terms of high-level rather than primitive concepts, such as edges.

The use of the DAFT block could offer significant improvements, as it can inform the convolution blocks with clinical data such as the occlusion site, so the network can incite or repress high-level features learned from the image in associated brain regions to improve predictive performance. DAFT is computationally efficient, because it does not depend on the number of instances in the dataset, nor the spatial resolution of the feature maps. The transformation is described in [20] in the following way: let $x_i \in \mathbb{R}^P$ denote the for the ith instance of the tabular information and $F_{i,c} \in \mathbb{R}^{DxHxW}$ denote the cth output (feature map) of a convolutional layer based on the ith volumetric image ($c \in \{1, \dots, C\}$). The network learns the Dynamic Affine Feature Map Transform (DAFT), with scale $\alpha_{i,c}$ and offset $\beta_{i,c}$:

$$\mathbf{F}_{i,c} = \alpha_{i,c}\mathbf{F}_{i,c} + \beta_{i,c}, \quad \alpha_{i,c} = f_c(\mathbf{F}_{i,c}, xi), \quad \beta_{i,c} = g_c(\mathbf{F}_{i,c}, xi), \tag{1}$$

where f_c, g_c are arbitrary functions that map the image and tabular data to a scalar. Functions f_c, g_c are modeled by a single auxiliary neural network h_c that outputs one α, β pair, which they refer to as DAFT. DAFT first creates a bottleneck by global average pooling of the feature map, concatenating the tabular data, and squeezing the combined vector by a factor r via a fully connected layer. Next, both vectors are concatenated and fed to an MLP without bias terms that compute the vectors α_i and β_i; following the work of Hu et al. [45]. Linear, sigmoid, and tanh activation functions can be applied to the scale α. We utilize a slightly altered version of the proposed network by adding a dropout layer before the last fully connected layer to combat overfitting caused by our small sample size.

We compare our DAFT-inspired model to a naive late fusion and a hybrid fusion model. Early fusion was not included in the experiments

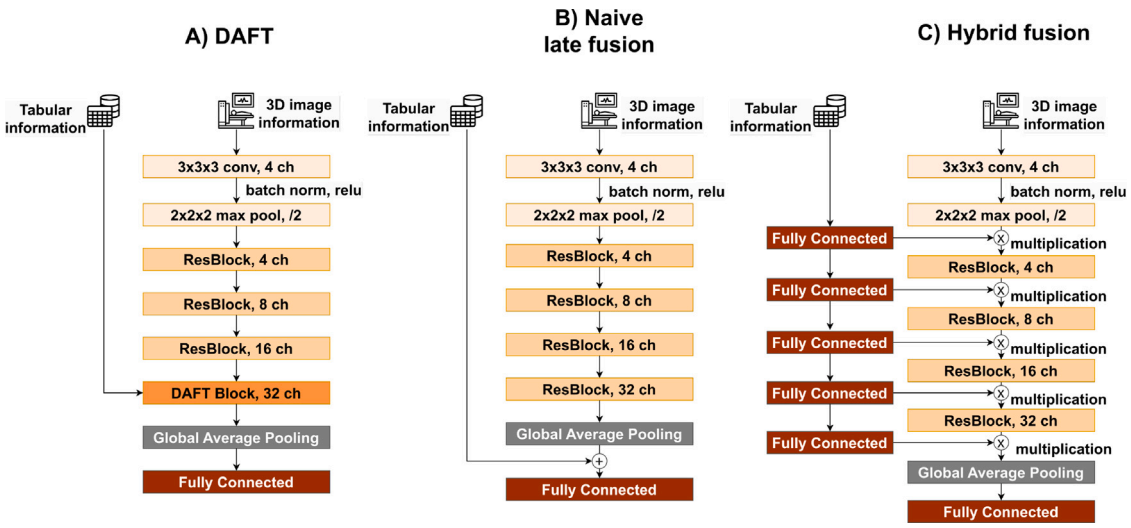

Fig. 4. Implemented multimodal architectures for comparison. A) Dynamic Affine Feature Map Transform (DAFT) based on the work of Pölsterl et al. [20]. B) Naive late fusion model C) Hybrid fusion model.

as it did not seem fruitful based on the work of Pölsterl et al. [20]. A visualization of the implemented architectures can be seen in Fig. 4. We also chose to incorporate the best-performing baselines into the different multimodal architectures, to compare to our vanilla setups. For TabNet, we used the featurized outputs of the pretrainer [35] as input for the tabular branch of all three the multi-modal architectures. The pretrainer functions similar to the standard TabNet, but creates featurized inputs as its outcome. For the pretrainer, we used the same setup and hyperparameters as the best performing regularly-trained TabNet configuration.

4.6. Experimental setup

We used the Receiver Operating Characteristics Area Under the Curve (ROC-AUC) [46] and F1 score [47] metric to compare model performance. AUC is the main characteristic, F1 score is added for context on specificity and sensitivity, as our labels are slightly unbalanced. After preprocessing, we divided the samples into training and testing sets via stratified random sampling to preserve data distribution with a train-validation-test split of 60%-20%–20%. Data of both modalities was standardized.

Our experiments were performed using Python 3.8 with scikit-learn and PyTorch as machine learning frameworks. To train the models, a virtual server was used inside the hospital to prevent data from leaving the hospital's environment, which included 4 Intel Xeon Gold 6254 CPU cores, 24 GB RAM, and a single NVIDIA Tesla T4-4Q GPU with 4 GB VRAM.

To compensate for the small dataset, we used multiple augmentation techniques to combat overfitting. For this, we made use of built-in image transformations from the MONAI [48] framework. Data augmentation was performed on the images, details can be found in A. For all models we performed hyperparameter searches, details of the finetuned hyperparameters and their best configurations can be found in B. For the multimodal models, the versions with TabNet featurized inputs were separately tuned. Moreover, also normalization of the featurized input was attempted as part of the tuning process.

5. Results

5.1. Clinical models

In the clinical modality, we tested three different models to establish a baseline, and to see how we compare to previous studies with our DASA dataset. These results can be seen in Table 3. Our reference RF classifier performed better than the GB classifier on the test set, with better AUC and F1 score. Our proposed method TabNet outperformed the other baselines with a 0.71 in AUC, with a balanced profile in AUC and F1 score. As TabNet also offers interpretability, Fig. 5 illustrates the feature importances on decision-making, both globally and for three local instances. As we can see, age is overwhelmingly important in the global feature importance, while most other features only account for about 5% to 10% of the decision making. It is interesting that for the individual predictions, the feature importance differs from the global feature importance, with age not necessarily being the most important feature. Moreover, TabNets embedding also causes only a few features per prediction to be selected.

Table 3
Model performance on single modality data.

	Model	AUC	F1 score
Clinical modality	GBC	0.61	0.60
	RF	0.65	0.67
	TabNet	**0.71**	0.70
Imaging modality	ResNet-10	0.70	0.52
	MedicalNet	0.62	0.72

5.2. Imaging models

To establish an imaging baseline, we experimented with 2 different models with very similar architectures, Resnet-10 and MedicalNet. We first investigated the effect of including different imaging modalities, as previously described in the data preprocessing section. We trained our ResNet-10 with (1) all of the available perfusion maps (2) only the four computed maps such as CBV, CBF, etc. (3) only the summed up and averaged maps (4) each of the 6 maps on their own. From this, we concluded that using only the summed-up perfusion maps yields the best performance and moved forward using this single imaging modality. Next to the ResNet-10 model from scratch, we fine-tuned the ResNet-10-based MedicalNet. The main difference is in the number of filters in these models. Table 3 shows the results of these experiments. We see an overall better performance than with the clinical modality. Implementing the pre-trained MedicalNet network did not live up to the expectations, as it underperformed in AUC compared to the ResNet-10, though it did have a better F1 score. As MedicalNet did not lead to an improvement in predictive performance, it was not considered further as input for the multimodal architectures.

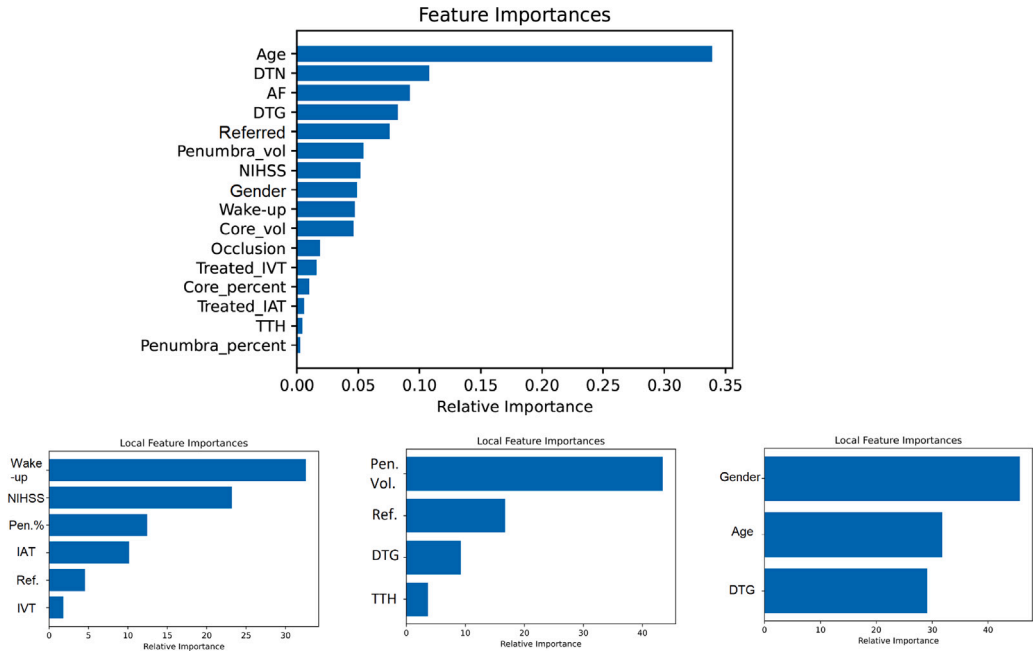

Fig. 5. Feature importances for the TabNet model. Top left: Global feature importance for the total model. Bottom: local feature importance of 3 individual predictions.

5.3. Multimodal models

Finally, to investigate the utility of combining the clinical and imaging modalities, we have trained 3 different models to explore the potential of naive late fusion, hybrid fusion, and dynamic affine transformations (DAFT). We have trained these models with the standard tabular data input, and the featurized TabNet output. We would expect the multimodal models to improve in performance compared to both the separate modalities. However, the results in Table 4 show that the Hybrid fusion model on the standard data falls behind the other modalities. Due to the fact that the training curve of this model comes close to 1 and the validation curve stagnates, this hints at overfitting on the non-augmented tabular data due to our small dataset. Our Naïve late-fusion model comes close to the other modalities but still falls behind with 3% in AUC and 5% in F1 score. We tried to improve the generalizing ability of both models by including dropout layers at different locations in the network, yet they still underperformed compared to the other modalities. For both these models, using the TabNet input increases the performance, to be slightly better than the results from the singular modalities separately.

While these two models show deep learning-based data fusion models on our small dataset as not very attractive, our modified DAFT model shows improvement. With an AUC of 0.75 and F1 score of 0.80, it outperforms our other fusion models, even those that were trained with the TabNet pretrained input. Interestingly, using the TabNet pretrained input led to a comparable AUC, but a worse F1 score. The final DAFT model had less than a hundred thousand optimizable parameters, compared to the pre-trained MedicalNet which had 14.4 million parameters due to the much larger feature map sizes.

Table 4
Model performance on multimodal data.

Model	AUC	F1 score
DAFT	**0.75**	**0.80**
Naive late fusion	0.67	0.47
Hybrid fusion	0.68	0.37
DAFT+TabNet	0.74	0.59
Naive late fusion+TabNet	0.72	0.70
Hybrid fusion+TabNet	0.72	0.63

6. Discussion

The results of our paper show that both TabNet and DAFT are suitable options for a clinical prediction model on functional outcome after stroke, depending on what data is available. We compare our results to the literature in Table 5. It is important to highlight that these studies are performed on different datasets. Our proposed method TabNet managed to outperform the baselines with a 0.71 in AUC, with a balanced profile in AUC and F1 score. We can see that our TabNet model performs on par with previous studies, with a better F1 score. Given that our dataset was at most half the size of the other comparable studies, shows that the proposed transformer-based TabNet architecture for the clinical modality is promising in case only clinical data is available. It outperforms both random forests and gradient boosting classifiers on our dataset. These are typically well-performing algorithms on tabular data, as we can see in comparable studies, so this supports the potential of TabNet. It would likely benefit from larger datasets, where even self-supervised pre-training could boost performance.

Moreover, the most important variables found by TabNet (age and door-to-needle time a.o.) match with what is known in the literature [28]. It is important to highlight that TabNet did not outperform the clinical models of the comparative studies. Though the implemented baseline models struggled to achieve comparable performance — GBC being worse with this limited sample size. Moreover, we chose a readily available dataset in DASA to create a more realistic scenario for implementation, but this might have limited our feature set compared to previous studies.

Our imaging models performed similarly to previous papers. Our ResNet-10 model performs comparably or slightly better than the previous studies, apart from [27] seems to outperform it in terms of AUC, but we do not know their F1 score. Transfer learning with MedicalNet did not live up to the expectations, as it underperformed in AUC compared to the baseline ResNet-10 trained from scratch. As the MedicalNet was trained on 23 different segmentation tasks, it is possible that the difference between the pre-training and our classification tasks was too large to improve performance in this case. With the limited data availability, the comparable results could hint at a benefit of CT perfusion scans as opposed to regular Non-Contrast CT imaging.

Table 5
Comparison of our best performing model for each modality to the literature.

	Ref.	Model	AUC	F1 score	Dataset size
Clinical Modality	[30]	FCN	0.61	0.70	204
	[27]	FCN	0.64	–	324
	[29]	RF	0.65	–	222
	[28]	GBC	**0.75**	–	246
	[31]	ClinicDNN	0.70	0.56	500
	Our	TabNet	0.71	0.70	98
Imaging modality	[30]	CNN	0.54	0.72	204
	[27]	CNN	**0.81**	–	324
	[31]	CNN	0.67	0.46	500
	Our	ResNet-10	0.70	0.52	98
Multimodality	[30]	Naive late fusion	**0.75**	0.69	204
	[31]	IMF block	**0.75**	0.62	500
	Our	DAFT	**0.75**	**0.80**	98

If we compare our results with previous multimodality studies in Table 5, our DAFT model managed to surpass comparable studies. While it offers a comparable AUC, the F1 score is a notable 11% better. These results affirm that informing the imaging representation with the overall state of the patient by affinely transforming the feature maps allows the network to differentiate notably better between outcomes.

Intuitively, for a multi-modal model that combines imaging and clinical data the expectation is a better performance than a model based on either of the modalities separately, simply because of the more extensive data. Interestingly, this was not necessarily the case, as the naïve late and hybrid fusion setup did not outperform the imaging baselines or TabNet. This shows that the architecture of the multimodal network is very important: when tabular data was fused with imaging data in the naive late and hybrid fusion models, it tended to overfit on the tabular inputs. Using the featurized TabNet inputs leads to an increase in performance, possibly limiting the overfitting. We also see an increase in performance, compared to both the other modalities as well as the multimodal models in the literature, when using DAFT. This shows a clear advantage for affine transformation-based data fusion compared to models with other fusion techniques or a single modality. Adding the featurized outputs did not lead to a further performance gain. Here, augmenting the imaging modality and using tabular data to affinely transform the representation of these images proved to be a very powerful technique even with a small sample size.

The effect of using CT perfusion does show an added benefit. It is interesting to see that using only the summed-up perfusion maps yielded the best performance, as this would arguably contain less information than using several maps, though it is not unlikely that using too many different maps would lead to overfitting on such a small dataset. We demonstrate that a combination of DAFT with CT perfusion leads to a good performance. It should be investigated whether NCCT scans would lead to a similar performance, as these are done on a wider group of CVA patients.

We believe that utilizing DAFT for predicting the functional outcome of stroke patients is a simple, yet effective technique that could even offer further improvements. As our final models only used the summed-up versions of the CT perfusion scans, no complex preprocessing step is required to arrive at these results. We only used third-party software to get the volume of the affected brain tissue. Furthermore, due to the modular nature of DAFT, it can be adapted to bigger or more complex networks with more capacity to learn high-level features. Additionally, the relatively small size of the network would make implementation in a hospital setup more feasible. Medical centers equipped with the proper data pipelines could implement these models for clinicians in a way that they would be able to see a prediction in their Electronic Health Record (EHR) system as soon as they begin the treatment of the patient. The ability to give a reliable prognosis of recovery could enable hospitals and rehabilitation clinics to have better planning of resources and, more importantly, inform patients and their relatives of the expected functional status after discharge. This allows for better expectation management and it enables the patient to partake in possible shared decision making. Moreover, we noticed that while the global interpretability was consistent with the literature, the local interpretability, which explained the decision for each individual patient, often differed from the global interpretability. This could allow for a more personalized explanation of expected functional outcome.

This study did come with a few limitations. First and foremost, the dataset is smaller compared to other studies. While we expect our predictive performance will go up with extension of our datasets, we cannot show this. With a larger sample size available, it would also give possibility to investigate the applicability of Transformer based imaging models such as ViT, or a variant more suitable for handling smaller datasets. We see how the TabNet outperformed the other models for clinical modality. We had a rather limited spatio-temporal analysis of the CT perfusion, and future studies could incorporate different maps or image denoising [49]. We purposely relied on the available data and software in the hospital, to improve the current process and not create more workload for medical professionals. This also applies to several excluded imaging based-biomarkers, such as the ASPECTS score. These scores are not part of our care process in the hospital. In settings where these biomarkers are collected as part of the tabular data, the clinical model performance could increase.

Moreover, another limitation is the dichotomization of the outcome. While this is similar to previous papers, based on our discussions with neurologists and other medical professionals, a valuable research direction would be to build models that can differentiate between patients better than the favorable and unfavorable recovery. Having 3 or more distinct classes would help a better separation, provide more specific expectations and thereby improve clinical relevance. Looking at the distribution of our data, to predict individual mRS scores such initiatives should collect at least 10 times more data to have enough instances for minority classes. Finally, if clinicians want the best possible prognosis and urgency is not the most important factor, developing models that make use of post-treatment features could further improve predictive power. Including information like treatment success, length of stay, and follow-up CT or CTP scans would likely provide a more elaborate picture of the patients' expected recovery.

7. Conclusion

This study presents a multimodal approach to predict the mRS functional status of acute ischemic stroke patients based on clinical and CT perfusion characteristics. To predict the functional outcome, we built models to predict good outcome and treatment success (scores 0–2), or poor outcome and treatment failure (scores 3–6). In our experiments, our proposed method with the modified DAFT outperforms both other methods and state-of-the-art results by achieving a 75% AUC, and 80% F1 score. It is remarkable that we achieved this, as our final model has less than a hundred thousand optimizable parameters, and was trained with a dataset at least half the size of reference papers. Overall, we

demonstrate the feasibility of predicting the functional outcome for ischemic stroke patients with LVOs.

Additionally, future research is needed on validating the model performance and investigating when and how to present this information to professionals and relatives. Overall, it is recommended to retrain developed models when more data becomes available and to monitor for data drift in deployed settings. Future studies may alleviate the problem of small, imbalanced datasets with larger, multi-center data collection initiatives where CT perfusion imaging is available. To overcome the challenges of data sensibility, this project – and many others in the medical field – would also be a great candidate for implementing distributed machine learning methods, such as federated learning [50] across horizontally or vertically partitioned datasets.

Declaration of competing interest

None

Acknowledgments

We sincerely want to thank Paul van der Nat, Lea Dijksman, Pieter Hilkens and Lucianne Langezaal from the St. Antonius Hospital for their contributions. This publication is part of the project Enabling Personalized Interventions (EPI) with project number 628.011.028 of the research programme Commit2Data-Data2Person, which is (partly) financed by the Dutch Research Council (NWO), Netherlands.

Appendix A. Data augmentation details

To compensate for the small dataset, we used multiple augmentation techniques to combat overfitting. For this, we made use of built-in image transformations from the MONAI [48] framework. After finetuning, data augmentation was performed on the images with the following configurations.

- RandRotate(range_x = 0, range_y = 0, range_z = 1, prob = 0.5)
- RandZoom(min_zoom = 0.9, max_zoom = 1.1, prob = 0.4)
- RandGaussianNoise(mean = 0, std = 0.01, prob = 0.5)
- Rand3DElastic(sigma_range = (5,7), magnitude_range = (50,140), padding_mode = 'zeros', prob = 0.5)
- RandAdjustContrast(gamma = (1,1.5), prob = 0.5), for the multimodal setup only.

Appendix B. Hyperparameters

For all models we performed hyperparameter search. The finetuned hyperparameters and their best configurations are as follows:

- **Random Forest**: *maximum depth* = 5, *max features* = log2, *minimum samples per leaf* = 1, *minimum samples per split* = 10, *n_estimators* = 100
- **Gradient Boosting**: *learning rate* = 0.1, *criterion* = friedman_mse, *maximum depth* = 2, *max features* = log2, *n_estimators* = 500
- **TabNet**:*Learning rate* = 0.002, *optimizer* = Adagrad, *decaying of learning rate* = 0.95 every 20 steps, *scheduler* = PyTorch's *StepLR*, *momentum* = 0.005. *n_steps* = 3, *gamma* = 1. Gated Linear Units: *n_shared* = 2, *n_independent* = 5. Width: *prediction layer* = 8, *attention embedding layer* = 30, *epochs* = 100
- **Resnet-10**: *Learning rate* = 0.001 (decreased by multiplying it by 0.1 and 0.05 when the number of epochs reaches 60% and 90% respectively), *optimizer* = AdamW, *weight decay* = 0.0001, *batch size* = 15, *epochs* = 100.
- **MedicalNet**: Same optimization and augmentation settings as ResNet-10, but *feature maps* = 512, all layers in training mode
- **Naive late fusion**: epochs = 50, learning rate = 0.001, decay rate = 0.0001, optimizer = AdamW, batchsize = 15
- **Naive late fusion+ TabNet**: *epochs* = 50, *learning rate* = 0.001, *decay rate* = 0.0001, *optimizer* = AdamW, *batchsize* = 15, *featurized input normalization* = None
- **Hybrid fusion**: *epochs* = 50, *learning rate* = 0.001, *decay rate* = 0.0001, *optimizer* = AdamW, *batchsize* = 15, bottleneck factor = 5
- **Hybrid fusion+Tabnet**: *epochs* = 50, *learning rate* = 0.001, *decay rate* = 0.0001, *optimizer* = AdamW, batchsize = 5, bottleneck factor = 5, *featurized input normalization* = None
- **Daft**: *Bottleneck factor* = 5, *Location FiLM block* = 0, *scaling* = enabled, *shifting* = enabled, *scaling activation* = linear, *probability dropout layer* = 0.3, *epochs* = 50, *learning rate* = 0.001, *decay rate* = 0.0001, *optimizer* = AdamW, *batchsize* = 15
- **Daft+Tabnet**: *Bottleneck factor* = 7. *Location FiLM block* = 0, *scaling* = enabled, *shifting* = enabled, *scaling activation* = linear, *probability dropout layer* = 0.3, *epoch* = 50, *learning rate* = 0.0005, *decay rate* = 0.0001, *optimizer* = AdamW, *batchsize* = 10, *featurized input normalization* = None

References

[1] Feigin VL, Brainin M, Norrving B, Martins S, Sacco RL, Hacke W, et al. World stroke organization (WSO): Global stroke fact sheet 2022. Int J Stroke Off J Int Stroke Soc 2022;17:18–29. http://dx.doi.org/10.1177/17474930211065917, URL https://pubmed.ncbi.nlm.nih.gov/34986727/.

[2] Inamdar MA, Raghavendra U, Gudigar A, Chakole Y, Hegde A, Menon GR, et al. A review on computer aided diagnosis of acute brain stroke. Sensors 2021;21:8507. http://dx.doi.org/10.3390/S21248507, https://www.mdpi.com/1424-8220/21/24/8507/htm, https://www.mdpi.com/1424-8220/21/24/8507.

[3] Lakomkin N, Dhamoon M, Carroll K, Singh IP, Tuhrim S, Lee J, et al. Prevalence of large vessel occlusion in patients presenting with acute ischemic stroke: a 10-year systematic review of the literature. J Neurointerventional Surg 2019;11(3):241–5.

[4] Zhu W, Churilov L, Campbell BC, Lin M, Liu X, Davis SM, et al. Does large vessel occlusion affect clinical outcome in stroke with mild neurologic deficits after intravenous thrombolysis? J Stroke Cerebrovasc Dis 2014;23(10):2888–93.

[5] Shlobin NA, Baig AA, Waqas M, Patel TR, Dossani RH, Wilson M, et al. Artificial intelligence for large-vessel occlusion stroke: A systematic review. World Neurosurg 2022. http://dx.doi.org/10.1016/j.wneu.2021.12.004.

[6] Wiens J, Shenoy ES. Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. Clin Infect Dis 2018;66:149–53. http://dx.doi.org/10.1093/CID/CIX731, URL https://academic.oup.com/cid/article/66/1/149/4085880.

[7] Mouridsen K, Thurner P, Zaharchuk G. Artificial intelligence applications in stroke. Stroke 2020;51(8):2573–9.

[8] Adams Jr HP, Del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the early management of adults with ischemic stroke: a guideline from the American heart association/American stroke association stroke council, clinical cardiology council, cardiovascular radiology and intervention council, and the atherosclerotic peripheral vascular disease and quality of care outcomes in research interdisciplinary working groups: the American academy of neurology affirms the value of this guideline as an educational tool for neurologists. Stroke 2007;38(5):1655–711.

[9] Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS One 2014;9:e88225. http://dx.doi.org/10.1371/JOURNAL.PONE.0088225, URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0088225.

[10] Os HJV, Ramos LA, Hilbert A, Leeuwen MV, Walderveen MAV, Kruyt ND, et al. Predicting outcome of endovascular treatment for acute ischemic stroke: Potential value of machine learning algorithms. Front Neurol 2018;9:784. http://dx.doi.org/10.3389/FNEUR.2018.00784/BIBTEX.

[11] Bivard A, Levi C, Spratt N, Parsons M. Perfusion CT in acute stroke: A comprehensive analysis of infarct and penumbra. Radiology 2013;267:543–50. http://dx.doi.org/10.1148/RADIOL.12120971, URL https://pubs.rsna.org/doi/abs/10.1148/radiol.12120971.

[12] Leiva-Salinas C, Patrie JT, Xin W, Michel P, Jovin T, Wintermark M. Prediction of early arterial recanalization and tissue fate in the selection of patients with the greatest potential to benefit from intravenous tissue-type plasminogen activator. Stroke 2016;47:397–403. http://dx.doi.org/10.1161/STROKEAHA.115.011066, URL https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.115.011066.

[13] Hopyan J, Ciarallo A, Dowlatshahi D, Howard P, John V, Yeung R, et al. Certainty of stroke diagnosis: Incremental benefit with CT perfusion over noncontrast CT and CT angiography 1. Radiology 2010;255:142–53. http://dx.doi.org/10.1148/radiol.09091021.

[14] Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: A survey. Pervasive Mob Comput 2018;51:1–26.
[15] Huang B, Yang F, Yin M, Mo X, Zhong C. A review of multimodal medical image fusion techniques. Comput Math Methods Med 2020;2020.
[16] Chandrasekaran G, Nguyen TN, Hemanth D J. Multimodal sentimental analysis for social media applications: A comprehensive review. Wiley Interdiscip RevData Min Knowl Discov 2021;11(5):e1415.
[17] Soleymani M, Garcia D, Jou B, Schuller B, Chang S-F, Pantic M. A survey of multimodal sentiment analysis. Image Vis Comput 2017;65:3–14.
[18] Sleeman IV WC, Kapoor R, Ghosh P. Multimodal classification: Current landscape, taxonomy and future directions. 2021, arXiv preprint arXiv:2109.09020.
[19] Swieten JCV, Koudstaal PJ, Visser MC, Schouten H, Gijn JV. Interobserver agreement for the assessment of handicap in stroke patients. Stroke 1988;19:604–7. http://dx.doi.org/10.1161/01.STR.19.5.604, URL https://pubmed.ncbi.nlm.nih.gov/3363593/.
[20] Pölsterl S, Wolf TN, Wachinger C. Combining 3d image and tabular data via the dynamic affine feature map transform. In: International conference on medical image computing and computer-assisted intervention. Springer; 2021, p. 688–98.
[21] Hop JW, Rinkel GJ, Algra A, van Gijn J. Quality of life in patients and partners after aneurysmal subarachnoid hemorrhage. Stroke 1998;29(4):798–804.
[22] Choi Y, Kwon Y, Lee H, Kim BJ, Paik MC, Won JH. Ensemble of deep convolutional neural networks for prognosis of ischemic stroke. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol. 10154 LNCS, Springer, Cham; 2016, p. 231–43. http://dx.doi.org/10.1007/978-3-319-55524-9_22, URL https://link.springer.com/chapter/10.1007/978-3-319-55524-9_22.
[23] Hilbert A, Veeling BS, Marquering HA. Data-efficient convolutional neural networks for treatment decision support in acute ischemic stroke. MIDL; 2018.
[24] Osama S, Zafar K, Sadiq MU. Predicting clinical outcome in acute ischemic stroke using parallel multi-parametric feature embedded siamese network. Diagnostics 2020;10:858. http://dx.doi.org/10.3390/DIAGNOSTICS10110858, URL https://www.mdpi.com/2075-4418/10/11/858.
[25] Koch G, Zemel R, Salakhutdinov R, et al. Siamese neural networks for one-shot image recognition. In: ICML deep learning workshop, vol. 2. Lille; 2015.
[26] Winzeck S, Hakim A, McKinley R, Pinto JA, Alves V, Silva C, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. Front Neurol 2018;9:679. http://dx.doi.org/10.3389/FNEUR.2018.00679/BIBTEX.
[27] Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, et al. Deep learning-derived high-level neuroimaging features predict clinical outcomes for large vessel occlusion. Stroke 2020;1484–92. http://dx.doi.org/10.1161/STROKEAHA.119.028101, URL https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.119.028101.
[28] Brugnara G, Neuberger U, Mahmutoglu MA, Foltyn M, Herweh C, Nagel S, et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. Stroke 2020;3541–51. http://dx.doi.org/10.1161/STROKEAHA.120.030287, URL https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.120.030287.
[29] Hamann J, Herzog L, Wehrli C, Dobrocky T, Bink A, Piccirelli M, et al. Machine-learning-based outcome prediction in stroke patients with middle cerebral artery-M1 occlusions and early thrombectomy. Eu J Neurol 2021;28:1234–43. http://dx.doi.org/10.1111/ENE.14651.
[30] Bacchi S, Zerner T, Oakden-Rayner L, Kleinig T, Patel S, Jannes J. Preliminary investigation deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: A pilot study. Academic Radiol 2020;27:e19–23. http://dx.doi.org/10.1016/j.acra.2019.03.015.
[31] Samak ZA, Clatworthy P, Mirmehdi M. Prediction of thrombectomy functional outcomes using multimodal data. Commun Comput Inf Sci 2020;1248 CCIS:267–79. http://dx.doi.org/10.1007/978-3-030-52791-4_21/TABLES/3, URL https://link.springer.com/chapter/10.1007/978-3-030-52791-4_21.
[32] Jansen IG, Mulder MJ, Goldhoorn RJB. Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (MR CLEAN Registry). BMJ 2018;360. http://dx.doi.org/10.1136/BMJ.K949, URL https://www.bmj.com/content/360/bmj.k949.
[33] Breiman L. Random forests. Mach Learn 2001;45(1):5–32. http://dx.doi.org/10.1023/A:1010933404324, URL https://link.springer.com/article/10.1023/A:1010933404324.
[34] Mason L, Baxter J, Bartlett P, Frean M. Boosting algorithms as gradient descent. Adv Neural Inf Process Syst 1999;12.
[35] Arik SÖ, Pfister T. Tabnet: Attentive interpretable tabular learning. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35. 2021, p. 6679–87.
[36] Kuhrij LS, Wouters MW, van den Berg-Vos RM, de Leeuw F-E, Nederkoorn PJ. The Dutch acute stroke audit: benchmarking acute stroke care in the Netherlands. Eur Stroke J 2018;3(4):361–8.
[37] Tipping ME. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res 2001;1(Jun):211–44.
[38] Philips. IntelliSpace portal. 2019, URL https://www.philips.nl/healthcare/product/HC881103/intellispace-portal-11.
[39] Tran T, Le U, Shi Y. An effective up-sampling approach for breast cancer prediction with imbalanced data: A machine learning model-based comparative analysis. PLoS One 2022;17(5):e0269135.
[40] Taneja S, Suri B, Kothari C. Application of balancing techniques with ensemble approach for credit card fraud detection. In: 2019 international conference on computing, power and communication technologies. GUCON, IEEE; 2019, p. 753–8.
[41] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst 2017;2017-December:5999–6009. http://dx.doi.org/10.48550/arxiv.1706.03762, URL https://arxiv.org/abs/1706.03762v5.
[42] Tammina S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. Int J Sci Res Publ (IJSRP) 2019;9(10):143–50.
[43] Chen S, Ma K, Zheng Y. Med3D: Transfer learning for 3D medical image analysis. 2019, http://dx.doi.org/10.48550/arxiv.1904.00625, arXiv preprint arXiv:1904.00625.
[44] Joze HRV, Shaban A, Iuzzolino ML, Koishida K. MMTM: Multimodal transfer module for CNN fusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 13289–99.
[45] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7132–41.
[46] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 1997;30(7):1145–59.
[47] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45(4):427–37.
[48] MONAI Consortium. MONAI: Medical Open Network for AI. 2022, URL https://github.com/Project-MONAI/MONAI.
[49] Hu D, Zhang Y, Liu J, Luo S, Chen Y. DIOR: Deep iterative optimization-based residual-learning for limited-angle CT reconstruction. IEEE Trans Med Imaging 2022;41(7):1778–90.
[50] Yang Q, Liu Y, Cheng Y, Kang Y, Chen T, Yu H. Federated learning. Synth Lect Artif Intell Mach Learn 2019;13(3):1–207.