A Multi-Modal Deep Learning Approach to the Early Prediction of Mild Cognitive Impairment Conversion to Alzheimer's Disease

Sijan S Rana, Xinhui Ma
Department of Computer Science and Technology
University of Hull
Hull, UK

Wei Pang
School of Mathematical and Computer Sciences
Heriot-Watt University
Edinburgh, UK

Emma Wolverson
Department of Psychological Health, Wellbeing and Social Work
University of Hull
Hull, UK

Abstract—**Mild cognitive impairment (MCI) has been described as the intermediary stage before Alzheimer's Disease – many people however remain stable or even demonstrate improvement in cognition. Early detection of progressive MCI (pMCI) therefore can be utilised in identifying at-risk individuals and directing additional medical treatment in order to revert conversion to AD as well as provide psychosocial support for the person and their family.**

This paper presents a novel solution in the early detection of pMCI people and classification of AD risk within MCI people. We proposed a model, MudNet, to utilise deep learning in the simultaneous prediction of progressive/stable MCI classes and time-to-AD conversion where high-risk pMCI people see conversion to AD within 24 months and low-risk people greater than 24 months. MudNet is trained and validated using baseline clinical and volumetric MRI data (n = 559 scans) from participants of the Alzheimer's Disease Neuroimaging Initiative (ADNI). The model utilises T1-weighted structural MRIs alongside clinical data which also contains neuropsychological (RAVLT, ADAS-11, ADAS-13, ADASQ4, MMSE) tests as inputs.

The averaged results of our model indicate a binary accuracy of 69.8% for conversion predictions and a categorical accuracy of 66.9% for risk classifications.

Index Terms—**Deep learning; Convolutional neural network; Alzheimer's disease; Mild cognitive impairment; ADNI**

I. Introduction

Alzheimer's Disease, like other forms of dementia, affects cognitive ability which is delineated by physical changes to the brain. These changes are characterised by the loss of neurons from a variety of causes, one of which can be attributed to the accumulation of amyloid plaques caused by the breakdown of amyloid-beta 42 proteins between neurons (1). The toxicity built up stagnated communication between neurons, eventually leading to its death. Another protein tau causes a similar effect; excess tau aggregates to form neurofibrillary tangles blocking neuron transport and harming communication between the synapses. Amyloid-beta, being an upstream of tau in AD triggers its conversion from its normal state to toxicity (2). Both proteins can also propagate throughout the brain, causing damage to all regions of the brain – including the hippocampus (3). It is the complex interplay between these two proteins that defines the biological precursors to Alzheimer's disease. There is also evidence to suggest some hereditary factors increase one's risk of dementia – the gene Apolipoprotein E (ApoE4) is found to be a major determinant of risk with the late onset of Alzheimer's (4).

Dementia can be described to have informal stages; typically, people transition from being displaying symptoms of mild cognitive impairment (MCI) at early or late stages before being diagnosed with dementia, most commonly being Alzheimer's (5). Most people do not develop dementia (n = 86 of 1,603 participants) according to a population-based study (6). Many stabilise at MCI (n = 384 participants), however, most reverting to consistent normal cognition (n = 881 participants). MCI can therefore be further categorised into sub-groups; progressive MCI (pMCI) and stable MCI (sMCI), excluding those who demonstrate improvements.

Around 10%-15% of MCI people are reported to later develop Alzheimer's every year over a relatively short observation period (7). Although only a smaller percentage of the entire MCI group convert to AD, the disease is still fatal and contributes to the majority of 87,199 deaths due to dementia in 2017 UK alone – 1 in 4 deaths amongst people aged 75 and over (8).

It is therefore essential that pMCI people are identified from sMCI; limited medical resources and treatment can be utilised effectively diverted to those who are in need the most. It will enable those who are likely to progress to Alzheimer's to make key changes to their lifestyles which, evidence suggests, can minimise their risks (9).

II. Current Methodologies

A. State-of-the-art Method

The current state-of-the-art model (10) achieves a 10-fold cross-validated accuracy of 86% on a train/validation/test split

of 680/72/32 subjects respectively. Their algorithm was aimed at distinguishing pMCI people who convert within 3 years; while also achieving a sensitivity of 87.5% and specificity of 85%. Their multi-modal model achieved these scores by using extensive preprocessing such as template registration and incorporated the use of Jacobian Determinant (JD) from the ADNI baseline MRI scans. The model also used set region of interests (ROIs) which targeted areas of the brain commonly associated with Alzheimer's – the parietal, temporal and frontal lobes.

As part of essential data processing for the analysis of Alzheimer's Disease in Deep Learning (11), MRI registration is achieved by nonlinearly co-registering each scan onto a custom T1 template after additional N4 bias field correction. The state-of-the-art model also used the Montreal Neurological T1 Template (MRI scans registered to MNI-152) to address the co-registration inaccuracies and assess the robustness of their classification methodology (10). After registration, all the non-brain areas were masked out using brain masks generated by Brain Extraction Tool using FSL (12).

Architecture-wise the model uses multiple 3D convolutional (conv3D) layers with 3D max pooling and batch normalisation. The Jacobian Determinant images is fed through their custom conv3D layers which used different filter parameters and the output were added to the MRI conv3D blocks. Each of the model perceptrons have non-linearity controlled by an Exponential Linear Unit (ELU) function. The model's layers are regularised with the use of dropout and L2 – Ridge regularisation, both of which will also be applied in this project's CNN model, as they are essential in reducing overfitting.

The state-of-the-art paper implements the use of residual learning (13) – a technique used in deep learning to help propagate input values of earlier layers into deep layers of the network. Essentially, the problem of deteriorating performance from deeper networks is alleviated with the use of these residuals which in turn allows for deeper and therefore more complex models. The method is based on the architecture of ResNet, a 152-layer state-of-the-art convolutional network (13).

B. All Convolutional Method

Although a simpler solution, the all convolutional method (14) still achieves a binary classification accuracy of 75.1% with their model in the prediction of conversion to AD. It also achieves scores of 74.8% and 75.3% in sensitivity and specificity respectively. Similar to the state-of-the-art model, the all convolutional method utilised cross-sectional data, meaning the prognosis are based on a single T1-weighted MRI scan. Their dataset used a train/validation and testing split of 90% and 10%. Overall, 1,409 subjects were used across ADNI-1, ADNI-2 and ADNI-GO projects.

The all convolutional method proposes a unique model in terms of deep learning architecture by the sole use of convolutional layers, while omitting the use of fundamental deep learning techniques such as max-pooling and batch normalisation. Unlike the ResNet architecture of the previous two approaches as based on – the all convolutional method takes inspiration from the All Convolutional Net which replaces max-pooling operations with normal convolutional layers with increased strides (15). The 14-layered network model also uses RELU which avoids the vanishing gradient problem of deep network architectures (16).

Data preprocessing steps applied also include the normalisation and registration to an MNI space using Diffeomorphic Anatomical Registration Exponentiated Lie Algebra (DARTEL) (17). The segmentation of grey matter, white matter and cerebrospinal fluid tissue to produce probability maps were utilised in the creation of the DARTEL template with modulation using Jacobian Determinants. To solve the problem of limited cross-sectional data, the all convolutional method used augmentation to randomly apply modification to the MRI scans. These augmentations include deformation, deformation and cropping, rotation and flipping, rotation and scaling, thus increasing the amount of available training and validation data.

C. Deep Residual Method

In the deep residual method, they utilised residuals in their network architecture - the method achieved 83.01% test accuracy (18), with sensitivity of 76% and specificity of 87% in the pMCI and sMCI classification task. A train/test split of 90% and 10% were used.

Their preprocessing method used the statistical parameter mapping to segment grey matter brain areas and register the skull-stripped image to the 152 average T1-MNI space. Further smoothing, warping and modulation were applied before training. The deep residual method also conducted a quality analysis correlation check using a threshold from the population mean image to discard outliers which were poorly registered. Only 2 of the 830 subjects did not meet the threshold.

Their extensive data worked with scans across multiple ADNI projects: ADNI-1, ADNI-2, ADNI-GO and ADNI-3. MCI diagnosed subjects that did not convert AD were classified as sMCI while those who converted to AD were classified as pMCI, excluding those with multiple conversions. Only baseline scans of each people were used.

In designing the architecture, they found an optimal layer depth of 15-layers, using 3 residual blocks with each residual block incorporating 2 basic blocks containing 2 convolutional layers each. Similar to the state-of-the-art paper, they utilised batch normalisation, max-pooling and the RELU activation function for the hidden layers. Overfitting was managed with the addition of L2 weight decay (ridge regularisation) at a value of 0.01.

In order to increase the separability of the pMCI versus sMCI classification, the deep residual method also explored the use of domain learning (19). Domain learning is the process of utilising data from a related classification problem. In domains with limited data i.e. AD conversion from MCI, the domain transfer paper (13) proposes that other related domains may

still provide valuable information in better solving the original problem. Therefore, a combination of the most informative features and samples can be extracted from the target and related domains for training. Effectively, a model can be trained with more training data thus better learning the separability of a classification problem.

This method was able to increase the deep residual method test classification accuracy to 83.01%, which is an improvement over their original cross-validated test accuracy of 75.01%. They were able to achieve this by utilising MRI data from an auxiliary domain (Alzheimer's/healthy brain classification).

III. KEY ASPECTS

Analysing some of the current methodologies has been clear in defining some essential techniques better solving the pMCI vs sMCI problem.

Although different in network architectures, many of the models perform similarly averaging test scores above 75% while replicating similar performances when measuring by sensitivity and specificity. The papers above all employ further preprocess from the ADNI pipeline; some practices such as the use of the RELU activation function and MRI registration is seems common practice –especially with the use of the MNI-152 template.

A. Preprocessing MRI

The feature space containing all possible features that could contribute to predicting pMCI conversion is massive. Given that MRI scans follow dimensions similar to $256 \times 256 \times 166$, resulting in $10, 878, 976$ data points per scan. Therefore, the application of deep learning in the problem requires further preprocessing of data to reduce its complexity and increase the extraction of relevant visual features of the progressively impaired brain.

1) Registration: The MNI-152 template space displayed by Figure 1 is used by all of the methods above to register their MRI brain scans. The Montreal Neurological Institute created the T1-weighted MNI-152 space by co-registering 152 normal MRI brain scans to the MNI-305 space. The linearly registered MNI-152 template is adopted by the International Consortium of Brain Mapping to define the standard, replacing the original Talairach atlas (20).

In the related research, ADNI MRI scans are used to register the brains onto the MNI-152 space. The registered images display a voxel representation of $1 \times 1 \times 1$ mm³ with rows, columns and slices resulting $197 \times 233 \times 189$ respectively.

Image registration is essential in the comparative analysis of medical images – the method allows for the alignment of the different regions in the brain and when paired non-linear algorithms that incorporate affine and deformable transformations; the pathological differences between the sMCI and pMCI are better preserved. The spatial differences, therefore, can be better calculated by appropriate spatial comparisons made by convolution operations in convolutional neural networks.
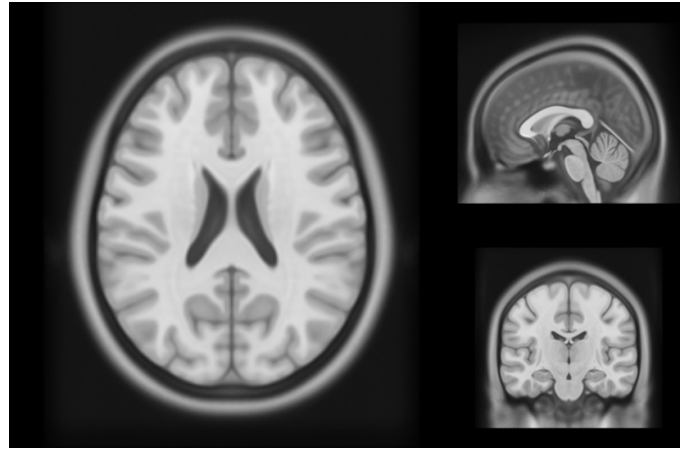

Fig. 1. T1 Weighted MNI-152 (McConnel Brain Imaging Centre, 2014)

2) Skull-stripping: MRI scans represent a complex feature space. The space denotes features that predict pMCI conversion from sMCI can be reduced by discarding irrelevant features within this space, such as the skull and eyes.

Another prevalent preprocessing step taken by all the approaches in the review of current methodologies is the extraction of brain thus removal of the skull – also known as skull-stripping. The deep residual and all convolutional methods both use grey matter extraction to achieve this.

Brain extraction therefore allows the problem classification problem to be more easily separable as only more relevant features exist within this space. Weight optimisation and error propagation can then focus on the spatial differences within these relevant features which will reduce the training time of the model and increase its predictive capacity.

B. Deep Learning Techniques

Current methods show that there is great variability in applying certain techniques and architectures. Residual learning is employed by most of the approaches in Section II. However, each of these methods utilise different CNN architectures as there exist many different parameters. Some of the models such as the deep residual and all convolutional, are deeply layered and large in capacity increase its ability to learn more complex features yet other solutions such as the state-of-the-art model perform better without requiring the depth.

1) Residual Learning: The use of residuals is based on the state-of-the-art convolutional neural network ResNet (13), achieving 1st place on the ILSVRC 2015 classification task. Figure 2 shows this shortcut mapping. The method involves propagating the input values in-between multiple convolutional layers before the activation function is applied. If the desired mapping $h(x)$ produces optimal results for pMCI conversion problem: $h(x)$ can be mapped from the output $f(x) = h(x) - x$ solving the problem as $h(x) = f(x) + x$ which achieves the same results. The benefits of this identity transformation are that the gradient information can be propagated to further layers which alleviates the problem of reduced model perfor-

mance as depth increases from gradient vanishing/exploding. (13).
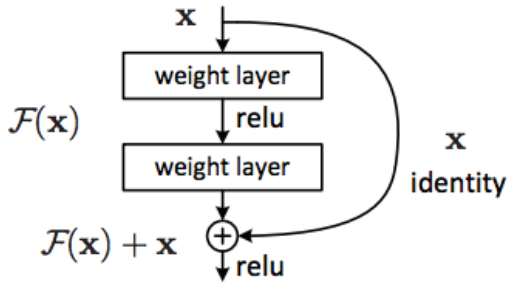

Fig. 2. A Residual block (13)

2) Rectified Linear Unit (RELU): The spatially configured perceptrons are layered to form convolutional neural networks with an aim to model the processes of the brain (21). It is the complex interplay with the activations of neurons within the brain that propagate signals – enabling thinking and actions.

Rectified Linear Unit is the activation function preferred by many CNN models as it mitigates the vanishing gradient problem existed in many deep-layered network architectures. The key property of activation functions is that they are differentiable in order to adjust the weights towards the optimal value in the backpropagation of error. The use of the sigmoid function presents a problem due to its gradient – the partial derivatives of error $w.r.t$ weights calculate the change to the update of the current weights. Each weight update becomes decreases by layer, at times effectively the model no longer updates its weights.

As a ramp function, RELU is able to reduce the vanishing gradient problem as it achieves a larger and constant gradient when compared to the maximum gradient of sigmoid. Another property of RELU is faster convergence (22). It is therefore no surprise that all the proposed models in the current methodologies use RELU or a variation of it (ELU).

3) Batch Normalisation: Batch normalisation is applied to the convolutional layers output in most of the methods above. The method with the lowest accuracy, the all convolutional method, omitted the use of batch normalisation – which may be a reason for its sub optimal performance when compared to other methods.

The problem presented in the training of deep learning models is that it is slowed by the changing distributions of layer inputs – described as internal covariate shift (23).

The process of batch normalisation aims to normalise layer inputs similar to how data are normalised during preprocessing. It also acts as a regulariser and enables greater flexibility in learning rate and initialisation of the model parameters (23).

4) Domain Transfer: The pMCI versus sMCI classification is a challenging prediction problem due to the complexity in unknown factors involved with conversion. This is further exacerbated given the limited data within the domain. This

problem can be alleviated with the use of auxiliary domains that solve a similar problem (19). Both the state-of-the-art method and the deep residual method use auxiliary domains in successfully achieving better classification results for the original conversion problem. This is done by the utilisation of AD/CN classification in extracting informative features for the limited data pMCI/sMCI problem.

IV. METHODS

A. ADNI Data

The data used in training MudNet are collected from the Alzheimer's Disease Neuroimaging Initiative (http://adni.loni.usc.edu/). ADNI's extensive collection of neuroimaging data also includes magnetic resonance imaging (MRI), positron emission tomography (PET), clinical and neurological assessments sampled from 1,821 participants – including MCI, AD and cognitively normal elderly controls.

In the training of MudNet, cross-sectional data from baseline measurements (n = 559 people) were used. The data was a combination of structural MRI and clinical data – containing neurological assessments (RAVLT, ADAS-11, ADAS-13, ADASQ4, MMSE) and demographics. The data were pooled across all ADNI projects (ADNI1/GO/2/3).

Data Preprocessing

1) ADNI Clinical: Clinical data provided by ADNI were also preprocessed. As shown in Tables I and II, these processes include one-hot and label-encoding to transform the data into numerical values to be fed into the model. In addition age, time in education, and neurological assessment scores were z-score normalised. The normalisation of input features allows for faster convergence of gradient decent algorithms (23).

2) Structural MRI: Advanced Normalisation Tools (ANTs) was used in preprocessing the MRI data (24). The ANTs scripts provide many components essential in the preprocessing of volumetric MRI data such as bias field correction and registration while achieving a higher predictive performance when compared to alternatives like FreeSurfer (25). These scripts are also available as Python modules (26).
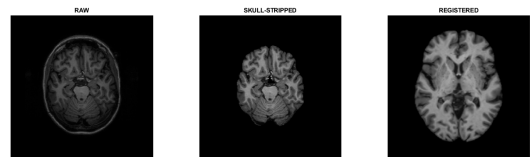

Fig. 3. The preprocessing pipeline – brain extraction and registration

3) N4 Bias Field Correction: Prior to extraction, the T1-weighted MRI scans were prepared by using N4 Bias Field Correction. As variant of the N3 Bias Field Correction, it boasts improved correction of the low frequency intensity non-uniformity within the volumetric data – increasing the accuracy of brain extraction and registration.

TABLE I
PERSONAL AND APOe4 DATA (MEAN±STD FORMAT)

	No. of subjects	Age (years)	Gender (M/F)	Time in education (years)	APOe4		
					0	1	2
pMCI	205	73.7±7.2	113/92	15.8±2.8	82	89	34
sMCI	354	71.9±7.8	197/157	16.1±2.9	218	110	26

TABLE II
NEUROPHYSIOLOGICAL TEST SCORES (MEAN±STD FORMAT)

	CDRSB	ADAS11	ADAS13	ADASQ4	MMSE	RAVLT			
						Immediate	Learning	Forgetting	Perceived Forgetting
pMCI	1.9±0.99	12.4±4.5	20.0±6.2	6.6±2.3	27.1±1.8	29.8±7.8	3.3±2.4	5.1±2.3	73.9±28.6
sMCI	1.3±0.72	8.4±3.8	13.6±5.8	4.5±2.3	28.1±1.7	38.3±11.4	4.9±2.6	4.3±2.5	50.0±31.0

4) Brain Extraction: To extract brain tissue from the MRI data, DeepBrain (https://github.com/iitzco/deepbrain) was used. The convolutional neural network is an alternative method which performs fast brain extraction and mask generation with reasonable accuracy. Brain extraction using this method was performed in less than 2 seconds which is considerably faster (1200x) when compared to BET (Brain Extraction Tool) in FSL (27). The CNN was also tested on a compound dataset which includes ADNI data, achieving state-of-the-art 0.97 Sørensen–Dice coefficient/F1 score.

5) Symmetric Diffeomorphic Image Registration: ANTs provide a novel symmetric image normalisation (SyN) method in image registration (28). The method is reliable in normalising and making anatomical measurements of neurodegenerative brains in volumetric MRI; it achieves strong correlation with volume measurements when compared to expert labelling.

Both masks and brains from extraction were utilised by registering to a common space – the T1 weighted MNI-152 template (Figure 1). Figure 3 visualises this preprocessing pipeline. SyNRA, a variation of SyN was used for the registration. The registration method applies affine and deformation transformations also including fine-matching and further deformation.

6) Fuzzy C-Means-based Intensity Normalisation: Structural MRIs vary in the distribution of intensity values as they do not have a standard scale. Prior to classification, uniformity in the intensity of brain MRIs is applied using fuzzy c-means (29) to calculate a white matter mask of the image and normalise the entire image to the mask's mean. This method is applied through a python module (https://github.com/jcreinhold/intensity-normalization). Additionally, z-score normalisation is applied before training to further ensure the intensities are in normal distribution.

B. Model Architecture Overview

The architecture of the MudNet is based on the state-of-the-art method (10) in identifying AD-converters at their MCI stage. Figure 4 visualises the architecture. Figure 5 expands on the model's invidual layers. MudNet aims to simultaneously solve an additional classification task in predicting the risk of pMCI conversion – converters are classified between high risk ($\leq$ 24 months) and low risk ($>$ 24 months) conversions. It does this by using multi-modality data - MRI and clinical features.

The model was trained for a maximum of 100 epochs using a batch-size of 20 and a learning rate of 0.05. This optimal value was found empirically using Adam optimiser by self-adaptively increasing the learning rate, with the value that produced the lowest validation loss selected.

Early stopping is used to prevent the overtraining of the model. Upon a specified limit of under-performing validation loss, training is automatically stopped i.e. if the validation loss fails to improve within 15 iterations, model training is stopped. Additional optimisation is introduced such as decreasing in the model's learning rate upon the plateauing of validation loss using ReduceLROnPlateau, a TensorFlow module. This allows for further gradient optimisation with smaller gradient updates therefore achieving higher test accuracy.

1) Model Depth: The model uses multiple convolutional layers (n = 7) with batch normalisation and max pooling operations. Also, layers 3-5 act as a residual block which utilises residual learning. Fully connected dense layers are incorporated to extract complex features within the ADNI clinical dataset; it is also used to produce classifications from the concatenated MRI and clinical outputs.

The final dense layers represent the class outputs of each person – pMCI versus sMCI (n = 2) and Risk (n = 3; for high risk, low risk and no risk).

2) Optimisation Function: The optimisation function to perform gradient decent is Adam – which achieves computa-
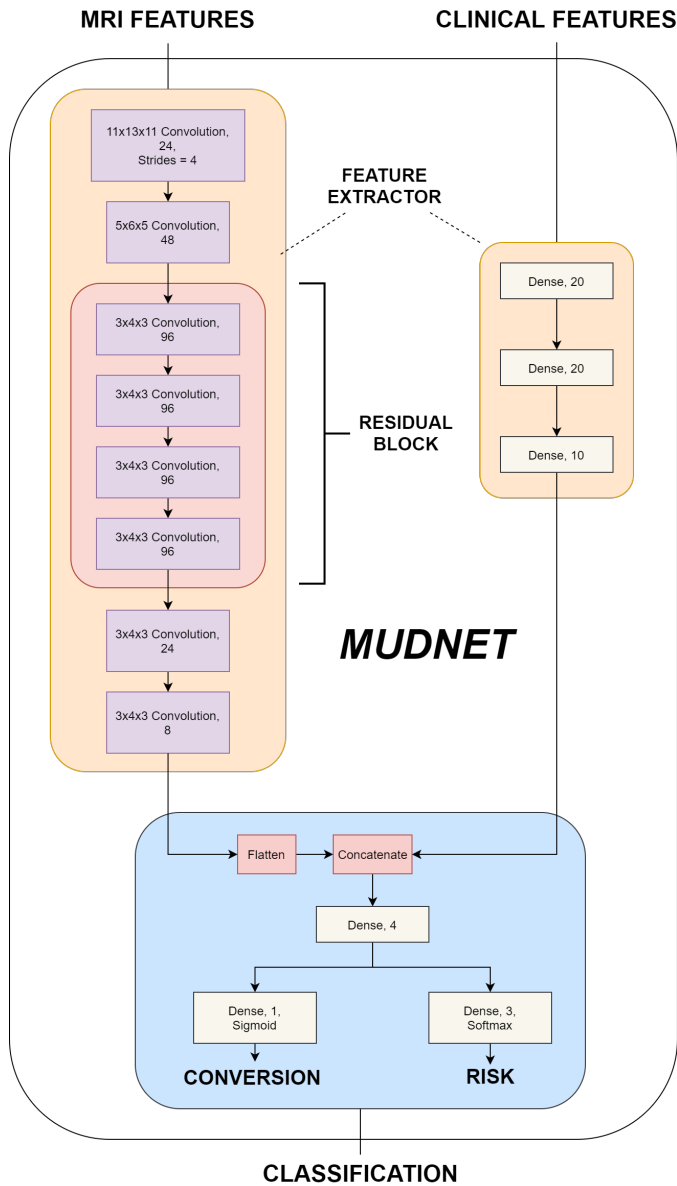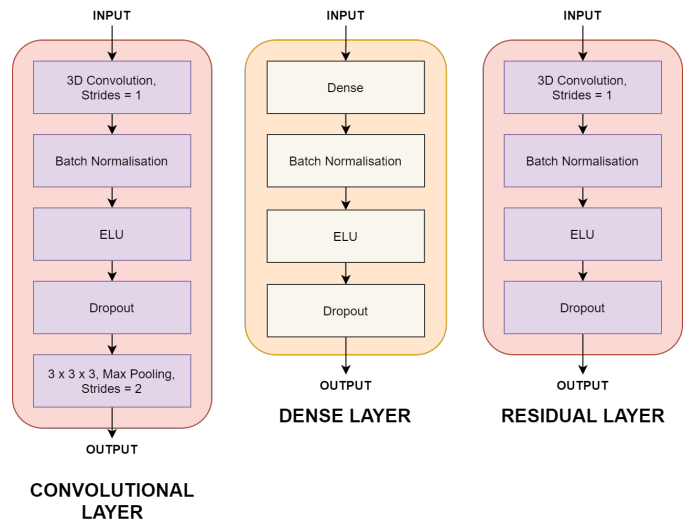
Fig. 4. MudNet architecture

Fig. 5. MudNet layers - convolutional Layers, dense Layers and the residual layers utilising exponential linear unit activation pre-batch normalisation

tionally efficient optimisation of the networks weight parameters whilst being suited to problems require large parameters and data (30).

3) Loss Functions: As the model outputs both binary and multi-class classifications, multiple loss functions are used. Cross-entropy is used in the model to calculate the loss between predicted and truth values in classification problems – binary cross-entropy in conversion (binary output) and categorical cross-entropy in risk (multi-class output).

4) Activation Function: Instead of the popular RELU, a variation – Exponential Linear Unit (ELU) was used instead. Unlike RELU, the function also outputs negative values which allows self-normalisation by pushing the mean unit activation towards zero. It shares the same characteristics of alleviating the vanishing gradient problem whilst simultaneously decreas-

ing training time in deep neural networks and increasing classification accuracies (31). The success of the activation function can also be seen in its use with the state-of-the-art method (10). The output layers, however, require different activation functions. The sigmoid and softmax activations convert their weighted sums into probabilities for activation. The sigmoid function represented by a single dense perceptron will activate for AD converters. Softmax, represented by 3 (n = classes) dense units, combine to produce probability distribution with a summation of 1 – resulting the activation of the single most probable respective class.

5) Regularisation: To control overfitting, the model uses both methods of dropout and weight decay. Dropout prevents the overuse of the same connections within the network by using a simple method of randomly dropping these connections. The artificial thinning of overall connections forces the network in utilising different connections, therefore increasing its ability to generalise (32). A value of 0.3 represents the fraction of units to be dropped for the convolutional layers and 0.1 for dense layers. Weight decay (also known as L2/Ridge Regression) is applied to the weighted connections in the kernels of the convolutional layers. This penalises loss using the squared magnitude of the weight coefficients and stabilising its updates. An optimal value of 0.0001 is selected here for all the layers. The value is determined empirically by testing values in factors of 10, choosing the lowest scoring validation loss as the final value.

C. Implementation

MudNet was built in Python (version 3.7.10) using Keras' Functional API (https://www.tensorflow.org/guide/keras/functional?hl=en) which utilises TensorFlow (version 2.2.0) as backend. The model was trained on the University of Hull's VIPER high performance computer (HPC). More specifically, 4 NVIDIA TELSA K40M GPUs were used to

train the model in parallel using TensorFlow's distributed training tool – Mirrored-Strategy. This allows the parallelisation of model training through all four GPUs, drastically decreasing the training time.

D. Performance Evaluation

The evaluation of the MudNet's performance was achieved by using the train-test split method. The dataset (n = 559 MRI scans) is split into a training and testing dataset prior to training using an 80-20 percentage split. Due to missing clinical data, some scans (n = 4 MRI scans) were dismissed. The model's performance is measured by outputting the final metric scores when training reaches an optimum test loss. This process is repeated (n = 10 iterations) with different partitions of train/test datasets to result an average performance.

In total, the dataset contains 63.3% MCI non-converters (n = 354 participants) and 36.7% MCI-converters (n = 205 participants). Of the MCI-converters, 65.4% (n = 134) were high risk individuals when compared to 34.6% (n = 71) who were low risk. Stratified splits were used in grouping the train and test splits, this ensures the balance of classes within both training and test sets so that there are minimal variances in the scores due to class imbalance.

V. CLASSIFICATION TASKS

Along with the clinical input (14-dimensional vector), MudNet is trained on T1-weighted MRI scans of extracted and registered brains. These images were renamed to contain the subject ID, class, and date of scan. This data is then read into the memory stored as a NumPy float array with its respective labels generated in another NumPy array at the same index. Labels that are identified as the truth values for the conversion classification are defined as $y \in \{0, 1\}$. The risk labels are defined as $y \in \{0, 1, 2\}$ representing no conversion, conversion within 24 months, and conversion within greater than 24 months, respectively.

VI. RESULTS

To evaluate MudNet's performance, a test dataset (n = 112 MRI scans) containing 20% of the overall data is used. The test dataset which is separate to the training dataset acts as untrained data to evaluate the model's ability to generalise and produce predictions on data not seen before. In an attempt to balance classes within the two datasets stratification is used. MudNet is trained and evaluated iteratively (n = 10 iterations) in order to produce the model's results displayed in table III.

To summarise the results, MudNet achieves an average accuracies of 69.8% and 66.9% on the test dataset. The model performs better when measured by the area under the curve (AUC) of the receiver operating characteristic (ROC) – achieving values of 0.80 and 0.83 averaged.

When analysing the boxplots in Figures 6 and 7, both conversion and risk classifications problems accuracy and AUC of ROC perform similarly. However, specificity shows a much greater variance in conversion than compared to risk
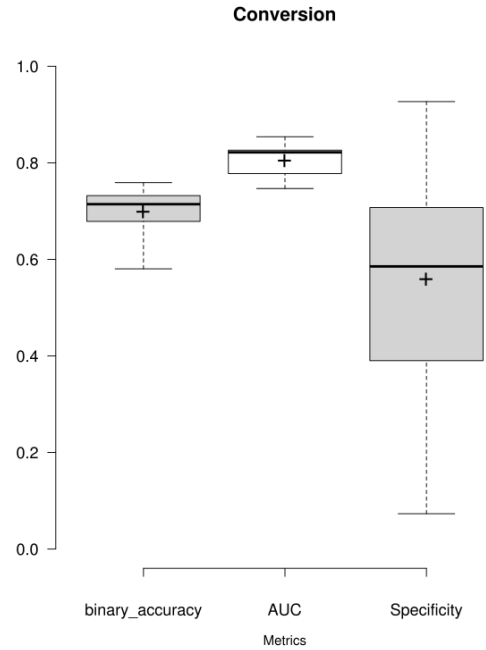
Fig. 6. Conversion classification

achieving 7.3% and 92.3% as its minimum and maximum values.

Overall, the model attains better than random results for both progressive MCI and time-to-ad classification. MudNet is able to identify pMCI from sMCI people 19.8% better when compared to random chance (50%). The model is 33.9% better at classifying pMCI people to their respective time-to-ad classifications when comparing to random chance (33%).

VII. DISCUSSION

In this paper, a convolutional neural network – MudNet is developed to discriminate mild cognitively impaired people who convert to Alzheimer's Disease, from those who stabilised at the condition. MudNet also simultaneously predicts the time-to-conversion of Alzheimer's Disease, classifying people into classes that define conversions within a 24-month period (high risk) or greater (low risk) as well as no risk (sMCI). This was achieved using preprocessed T1-weighted MRI scans. Clinical measurements from neurophysiological tests were also utilised as well as other data regarding time in education, gender, age and APOe4 genetic status data. The data collected are baseline measurements meaning they are recorded from people's first visits. The use of cross-sectional data attempts to simulate diagnosis conditions for a person's first visit.

MudNet is built using different aspects of current performant methodologies. The literature review summarises the some of the advances made in current research for early AD prediction in terms of the preprocessing employed and the

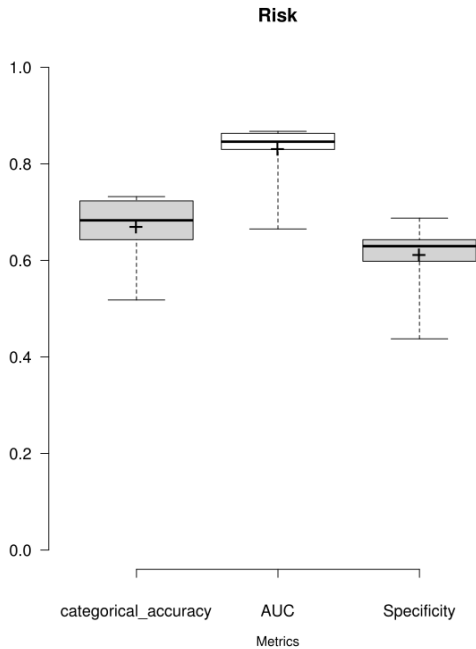| Test result	Accuracy		AUC (ROC)		Specificity	
	Conversion	Risk	Conversion	Risk	Conversion	Risk
1	0.6786	0.6161	0.7858	0.8261	0.7073	0.5625
2	0.5804	0.5179	0.7470	0.6651	0.9268	0.4375
3	0.7321	0.7143	0.8351	0.8673	0.6829	0.6250
4	0.7143	0.6696	0.7467	0.8336	0.5854	0.6071
5	0.6607	0.6518	0.8262	0.8298	0.0732	0.6429
6	0.7143	0.7248	0.7777	0.8451	0.5854	0.6429
7	0.7411	0.6964	0.8184	0.8632	0.5122	0.6339
8	0.7143	0.7321	0.8253	0.8562	0.3171	0.6875
9	0.7589	0.7232	0.8540	0.8673	0.3902	0.6696
10	0.6875	0.6429	0.8251	0.8467	0.8049	0.5982
Mean	0.69822	0.66891	0.80413	0.83004	0.55854	0.61071
	69.8%	66.9%	0.80	0.83	55.9%	61.1%

Fig. 7. Risk classification

used domain learning reported an increase in their validation accuracies. MudNet utilises some of these methods; batch normalisation and the use of residuals in training parts of the network with identity mappings are some of the methods also employed to the model proposed in the paper.

The model developed achieves validation accuracies of 69.8% and 66.9% in progressive MCI conversion and risk class predictions. Specificity measures the true rate of negative examples; MudNet achieves scores of 55.9% and 61.1% in conversion and risk. These results are sub-par to current research standards and for practical medical use but solidify that the application of deep learning has potential predictive power in detecting early individuals at risk in converting to Alzheimer's Disease. However, MudNet also aims to classify pMCI cases to their time-to-AD classes – which is a further step and a difficult problem to tackle simultaneously when compared to the existing related current methodologies. The model shows further potential when measured by its AUC performance – it achieves 80% and 83% for conversion and risk predictions respectively. When evaluated with specificity, the performance of the model is not ideal, especially in the conversion problem. A specificity of 55.9% would expose 44.1% to misclassifying as stable people.

In respect to the papers aims and objectives – the development of MudNet meet the purpose of the research in its ability to detect the conversion of progressive MCI people from stable people. The model also meets its secondary objective in more successfully solving the problem of predicting the time-to-AD class of these progressive people. However, the validation results of MudNet suggests further work is required in developing the model's architecture and for the hyper-parameter optimisation. As the data is limited and the preprocessing pipeline within current standards (11), feature extraction and therefore the performance can only be improved with adjustments in these stages.

In a domain with limited data, the use of available data should be maximised. One method to better solve the problem

deep learning techniques utilised. The existing related research established certain methodologies in the pMCI versus sMCI prediction such as image registration of the MRI data in their preprocessing pipeline. Other methods which can help explain the successful results achieving over 80% in cross-validation accuracies and AUCs (ROC) can be attributed to the use of domain learning in extracting the most informative auxiliary features from a similar domain I.e. Alzheimer's Disease versus cognitively normal classification. All related research that

of progressive MCI identification and its time-to-AD prognosis could be domain learning. In the literature review of current methodologies, domain learning has already seen positive impact when regarding the model performance and its increase in papers use the method. The training of the model's weights in recognising features between auxiliary AD and non-AD classes could not increase the performance of the model in the original problem but serve to reduce training time with faster convergence.

Brain segmentation is another strategy that could be employed in achieving better performance. The segmentation of the brain regions (temporal, parietal, prefrontal, occipital lobes) can allow the use of parallel 3D convolutional layers to better extract features specific to these regions in order to reduce the complex feature space. A smaller feature space should allow for the easier finding of informative features.

Also, the same data can be utilised in producing additional features. FSL provides a tool – SIENA SIENAX which allows for the longitudinal and cross-sectional analysis of structural changes within the brain (27). Using single or multiple MRI scans, the rate of neurodegeneration can be measured via volume changes which could be useful feature in providing the MudNet improved predictive performance for both pMCI versus sMCI and time-to-AD classifications.

In summary, MudNet, a convolutional neural network is developed to simultaneously solve the problems of detecting of AD converters from non-converters early and identifying high risk converters (conversion in $\leq$ 24 months) from low risk converters (conversion in $>$ 24 months). In these tasks the model achieves averaged (n = 10) validation accuracies of 69.8% and 66.9% outperforming random chance but did not achieve similar results from current research (80%+). The model attempts solve two complex problem simultaneously while potentially lacking the architecture depth to do so. MudNet could see improvements with its predictive capabilities if a singular problem was focused or extra depth is added to the network architecture. Further methods of improving the model may also include limiting class samples to ensure a global balance of classes within the dataset so that the availability of data does not cause bias.

VIII. ACKNOWLEDGEMENT

REFERENCES

[1] J. Weller and A. Budson, "Current understanding of alzheimer's disease diagnosis and treatment," *F1000Research*, vol. 7, p. 1161, 2018.

[2] G. S. Bloom, "Amyloid- and tau," *JAMA Neurology*, vol. 71, no. 4, p. 505, 2014.

[3] "What happens to the brain in alzheimer's disease?" [Online]. Available: https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease

[4] N. Zhong and K. Weisgraber, "Understanding the basis for the association of apoe4 with alzheimers disease: Opening the door for therapeutic approaches," *Current Alzheimer Research*, vol. 6, no. 5, p. 415–418, 2009.

[5] B. J. Kelley and R. C. Petersen, "Alzheimers disease and mild cognitive impairment," *Neurologic Clinics*, vol. 25, no. 3, p. 577–609, 2007.

[6] M. Ganguli, Y. Jia, T. F. Hughes, B. E. Snitz, C.-C. H. Chang, S. B. Berman, K. J. Sullivan, and M. I. Kamboh, "Mild cognitive impairment that does not progress to dementia: A population-based study," *Journal of the American Geriatrics Society*, vol. 67, no. 2, p. 232–238, 2018.

[7] A. J. Mitchell and M. Shiri-Feshki, "Temporal trends in the long term risk of progression of mild cognitive impairment: a pooled analysis," *Journal of Neurology, Neurosurgery Psychiatry*, vol. 79, no. 12, p. 1386–1391, 2008.

[8] 2017. [Online]. Available: https://www.gov.uk/government/publications/death-in-people-aged-75-years-and-older-in-england-in-2017/death-in-people-aged-75-years-and-older-in-england-in-2017

[9] "Can alzheimer's disease be prevented?" [Online]. Available: https://www.alz.org/alzheimers-dementia/research_progress/prevention

[10] S. Spasov, L. Passamonti, A. Duggento, P. Liò, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimers disease," *NeuroImage*, vol. 189, p. 276–287, 2019.

[11] A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect alzheimers disease from neuroimaging: A systematic literature review," *Computer Methods and*

Programs in Biomedicine, vol. 187, no. 105242, Nov 2019.

[12] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *NeuroImage*, Aug 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/21979382

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv.org*, Dec 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[14] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks," *NeuroImage: Clinical*, Dec 2018.

[15] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *CoRR*. [Online]. Available: https://arxiv.org/pdf/1412.6806.pdf

[16] X. Glorot, A. Bordes, and B. Bengio, "Deep sparse rectifier neural networks," *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, p. 315–323, 2011.

[17] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, p. 95–113, 2007.

[18] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, "Deep residual learning for neuroimaging: An application to predict progression to alzheimer's disease," *Journal of Neuroscience Methods*, vol. 339, p. 108701, 2020.

[19] B. Cheng, D. Zhang, and D. Shen, "Domain transfer learning for mci conversion prediction," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012 Lecture Notes in Computer Science*, p. 82–90, 2012.

[20] "About the mni space(s)." [Online]. Available: https://www.lead-dbs.org/about-the-mni-spaces/

[21] F. Rosenblatt, "Perceptron simulation experiments," *Proceedings of the IRE*, vol. 48, no. 3, p. 301–309, 1960.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, p. 84–90, 2017.

[23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv*. [Online]. Available: https://arxiv.org/pdf/1502.03167.pdf

[24] N. J. Tustison, P. A. Cook, A. Klein, G. Song, S. R. Das, J. T. Duda, B. M. Kandel, N. V. Strien, J. R. Stone, J. C. Gee, and et al., "Large-scale evaluation of ants and freesurfer cortical thickness measurements," *NeuroImage*, vol. 99, p. 166–179, 2014.

[25] B. Fischl, A. Dale, S. Martin, and G. Doug. [Online]. Available: https://surfer.nmr.mgh.harvard.edu/

[26] stnava, "Antspy." [Online]. Available: https://github.com/ANTsX/ANTsPy

[27] S. M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P. Matthews, A. Federico, and N. D. Stefano, "Accurate, robust, and automated longitudinal and cross-sectional brain change analysis," *NeuroImage*, vol. 17, no. 1, p. 479–489, 2002.

[28] B. Avants, C. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, p. 26–41, 2008.

[29] J. C. Reinhold, B. E. Dewey, A. Carass, and J. L. Prince, "Evaluating the impact of intensity normalization on mr image synthesis," *Medical Imaging 2019: Image Processing*, 2019.

[30] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*. [Online]. Available: https://arxiv.org/pdf/1412.6980.pdf

[31] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *CoRR*. [Online]. Available: https://arxiv.org/pdf/1511.07289.pdf

[32] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.