Received: 28 April 2023 | Revised: 9 June 2023 | Accepted: 21 June 2023

CAAI Transactions on Intelligence Technology

DOI: 10.1049/cit2.12291

ORIGINAL RESEARCH

Knowledge-based deep learning system for classifying Alzheimer's disease for multi-task learning

Amol Dattatray Dhaygude[1] | **Gaurav Kumar Ameta**[2] | **Ihtiram Raza Khan**[3] | **Pavitar Parkash Singh**[4] | **Renato R. Maaliw III**[5] | **Natrayan Lakshmaiya**[6] | **Mohammad Shabaz**[7] | **Muhammad Attique Khan**[8,9] | **Hany S. Hussein**[10,11] | **Hammam Alshazly**[12]

[1]Department of Data Science, University of Washington, Seattle, Washington, USA

[2]Department of Computer Science & Engineering, Parul Institute of Technology, Parul University, Vadodara, Gujarat, India

[3]Department of Computer Science, Jamia Hamdard, Delhi, India

[4]Department of Management, Lovely Professional University, Phagwara, India

[5]College of Engineering, Southern Luzon State University, Lucban, Quezon, Philippines

[6]Department of Mechanical Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India

[7]Model Institute of Engineering and Technology, Jammu, J&K, India

[8]Department of Computer Science, HITEC University, Taxila, Pakistan

[9]Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon

[10]Electrical Engineering Department, College of Engineering, King Khalid University, Abha, Saudi Arabia

[11]Electrical Engineering Department, Aswan University, Aswan, Egypt

[12]Faculty of Computers and Information, South Valley University, Qena, Egypt

Correspondence

Mohammad Shabaz.
Email: bhatsab4@gmail.com

Funding information

King Khalid University, Grant/Award Number: GRP/75/44

Abstract

Deep learning has recently become a viable approach for classifying Alzheimer's disease (AD) in medical imaging. However, existing models struggle to efficiently extract features from medical images and may squander additional information resources for illness classification. To address these issues, a deep three-dimensional convolutional neural network incorporating multi-task learning and attention mechanisms is proposed. An upgraded primary C3D network is utilised to create rougher low-level feature maps. It introduces a new convolution block that focuses on the structural aspects of the magnetic resonance imaging image and another block that extracts attention weights unique to certain pixel positions in the feature map and multiplies them with the feature map output. Then, several fully connected layers are used to achieve multi-task learning, generating three outputs, including the primary classification task. The other two outputs employ backpropagation during training to improve the primary classification job. Experimental findings show that the authors' proposed method outperforms current approaches for classifying AD, achieving enhanced classification accuracy and other indicators on the Alzheimer's disease Neuroimaging Initiative dataset. The authors demonstrate promise for future disease classification studies.

KEYWORDS

classification, deep learning

1 | INTRODUCTION

Alzheimer's disease (AD) is a degenerative neurological disorder that is difficult to detect in its early stages. Alzheimer's disease is the most prevalent cause of dementia; however, it can also be caused by other disorders. As it is frequently challenging to define the particular dementia subtype, it is difficult to get an exact estimate of the prevalence of Alzheimer's dementia. The cause of about two-thirds of all dementia diagnoses is commonly acknowledged to be AD. A growing section of society will likely be impacted, according to studies that utilised mathematical models to anticipate the frequency of dementia in the future. These studies took into consideration variables including rising life expectancy, altered mortality rates, and the prevalence of cardiovascular disease. The conventional method of diagnosing AD using magnetic resonance imaging (MRI) scans is laborious and arbitrary. As a result, a more effective and impartial early diagnostic approach is required. On the basis of MRI scans, deep learning (DL) approaches have demonstrated significant promise for categorising AD patients; however, accuracy and constraint reduction must be increased. In order to effectively categorise AD, MCI, and healthy control (HC) participants based on MRI data, this study suggests a DL model that makes use of 3D Convolutional neural networks (CNNs). The model seeks to improve AD diagnostic accuracy and lessen the restrictions of conventional Machine learning (ML) techniques. The research also aims to contribute to the development of effective and objective diagnostic approaches for AD and give insights into the potential of DL techniques for AD diagnosis.

To overcome the limitations of standard ML approaches for AD diagnosis, we offer an end-to-end AD classification method based on a deep three-dimensional CNN (CNN). This approach incorporates an attention strategy to improve feature extraction capabilities. Furthermore, it incorporates two auxiliary subtasks, namely Clinical Dementia Rating (CDR) Scale (CDR) score regression and Mini-Mental State Examination (MMSE) score regression, to improve AD classification findings and enable multi-task learning. The CDR Scoring Table provides descriptive anchors to guide the clinician in giving appropriate values based on interview data and clinical judgement. Together with CDR Worksheets, the Table is employed. An overall Global CDR Score may be determined in addition to ratings for each domain utilising a CDR Scoring Algorithm or by applying the established scoring guidelines. The CDR is a copyrighted piece of equipment, and using it calls for a licence. The Global CDR Score helps describe and monitor a patient's degree of impairment or dementia, with 0 denoting normal, 0.5 denoting very mild dementia, 1 denoting mild dementia, 2 denoting moderate dementia, and 3 denoting severe dementia.

The Mini-Mental State Examination (MMSE), a tool for cognitive screening, provides a rapid, reliable measure of the cognitive function. It may be employed to identify cognitive injury, evaluate how severe the impairment is, and monitor cognitive development. In clinical and academic settings, the MMSE is the brief cognitive evaluation that is used the most frequently. The MMSE measures a range of cognitive abilities, such as language and visual creation, orientation, repetition, verbal memory, attention, and calculation. The method begins with standardisation to achieve data enhancement. Then, the attention mechanism is introduced, and the three-dimensional convolutional network is used to obtain features with attention weights, which can get the most beneficial image position information for classification tasks. Finally, to optimise the output of the primary classification task, various fully connected layers (FC) backpropagate iteratively to obtain the classification output, CDR score, MMSE score, and the two auxiliary subtasks.

Convolutional neural network may be used to process images in a way that not only efficiently shrinks the size of a big quantity of data to a smaller amount of data but also efficiently keeps the characteristics of the original image. The input layer, hidden layer, and output layer are all components of CNN networks, much like other neural networks. Following convolution, the dimensional characteristics of the picture continue to split the feature matrix into several single blocks to determine its maximum or average value, which can help reduce the dimensionality of the image, speed up the network's computation, and prevent overfitting. After convolution, the image's dimensional features are still many; therefore, the feature matrix is split into multiple single blocks to determine their maximum or average values. This can help reduce the number of dimensional features, speed up network calculation, and prevent overfitting. The fully connected layer, which is the final layer in the CNN model, computes the final class scores by combining all local features and the feature matrices of each channel into vector representations.

The suggested approach works well on the AD classification problem, and it is a DL model that can efficiently extract medical picture information, as shown by ablation and comparison tests. The comparative experiment demonstrates that the suggested strategy performs better in AD classification than conventional ML techniques. The suggested approach may aid in the creation of AD diagnostic techniques that are more effective and impartial. The only accurate way to tell if someone had AD prior to the early 2000s was through an autopsy, a procedure that is performed after death. Lab and imaging tests are now easily accessible to allow a doctor or researcher to see biological symptoms of the disease or biomarkers in a living person. Currently, results from blood tests should be used in combination with other testing rather than alone to identify dementia. However, these diagnostic tests are still not widely accessible. Teams of researchers funded by the National Institute on Aging are still looking into possibilities for more accurate, less intrusive, and quicker methods of Alzheimer's diagnosis.

Many scholars at home and abroad have carried out classification research for the MRI images of AD patients. Traditional ML methods usually need to select features manually or semi-manually as the basis for classification. Different features will lead to other classification performances, and the choice of the classifier will also be different. For example, literature [1, 2] combined hippocampus and entorhinal cortex volume as features and used Support Vector Machines (SVM) to realise AD classification; literature [3] proposed a multi-level classification method for AD classification where the Gaussian Naive Bayesian classifier is used in the first stage. Then, according to the classification situation in the first stage, some samples are inputted to the SVM classifier or KNN classifier for reclassification. The above classification method may lose vital information in selecting features, which has significant limitations. In recent years, a series of deep and perfect network models such as Alex-Net [4], VGGNet [5], and GoogLeNet [6] have been proposed.

Although the above methods have achieved good performance in AD classification tasks, they have not effectively used all the information about AD classification: First, although there are many classification algorithms for 3D images in DL, due to the 3D convolutional network computing, the cost is too high, and most AD classification tasks are still 2D convolutional learning after slicing 3D images, which will lead to the loss of crucial structural information; secondly, in the clinical diagnosis of AD, not only MRI images but also other auxiliary information is necessary. However, the additional clinical data is not used in most AD classification tasks at present, which causes a waste of information resources. Alzheimer's disease is a neurological disorder that worsens with time and is challenging to identify in its early stages. The conventional method of diagnosing AD using MRI scans is laborious and arbitrary. As a result, a more effective and impartial early diagnostic approach is required. On the basis of MRI scans, DL approaches have demonstrated a significant promise for categorising AD patients; however, accuracy and constraint reduction must be increased.

Given the above problems, this paper proposes an end-to-end AD classification method based on a deep three-dimensional CNN (CNN), which introduces an attention mechanism to enhance feature extraction capabilities and adds two auxiliary subtasks, namely the CDR Scale (CDR) score regression and MMSE score regression to optimise the AD classification results and realise multi-task learning. For standardisation to achieve data enhancement, the attention mechanism is introduced, and the three-dimensional convolutional network is used to obtain features with attention weights, which can get the most beneficial image position information for classification tasks; In order to optimise the output of the primary classification task, various fully connected layers (FC) backpropagate iteratively to obtain the classification output, CDR score, MMSE score, and the two auxiliary subtasks. It can be seen from the ablation experiment and comparison experiment that the proposed method performs well on the AD classification task, and it is a DL model that can effectively extract medical image information. The study proposes a DL model for AD diagnosis using neuroimaging features that select features automatically and find the optimal network structure. The aim is to fill the research gap by proposing an efficient and effective DL model for AD diagnosis that outperforms existing models, contributing to the early prevention of AD and reducing physicians' burden.

2 | RELATED WORK

ResNet [7] has promoted the development of DL, making it in the medical image. The field of classification has been widely used [8]. However, there is a large gap between medical and natural image classification in the network structure: the latter usually requires a deep network because of the large number of categories; the former only targets specific types (such as tumours) and pays more attention to detailed information. Based on DL, literature [9] modified the network layer based on VGG16 to achieve a better AD classification effect; literature [10] used an ordinary 3D convolutional network and residual convolutional network to extract MRI based on the lightweight principle, several experiments were carried out to determine the network depth and the number of channels per layer; literature [11, 23, 24] constructed a new 3D convolution model HadNet, which fine-tuned hyperparameters through Bayesian optimisation.

The examination of high-dimensional features is the basic basis for the classification techniques employed in AD classification, which may result in the dimensionality curse. Feature selection is an effective way to reduce the number of characteristics by eliminating the irrelevant ones. Currently, AD classification algorithms employ the feature selection techniques, multi-task feature selection, group lasso, and principal component analysis. In multi-modal techniques, it is common to predict the class label using a direct combination of these selected traits from each modality. When adopting the classic multi-modal AD fusion method, the various modalities are generally linearly merged. To describe and capture the relevance across many modalities, sample significance analysis is crucial in the multi-modal fusion process.

Convolutional neural networks, recurrent neural networks (RNNs), and autoencoders are only a few of the DL models covered in the literature review that have been suggested for diagnosing AD. The authors discuss the advantages and disadvantages of each model and highlight the challenges associated with feature extraction and selection. Organisations may now collect data from the Internet of Things (IoT), social media, and other diverse sources in a variety of forms because of the Internet's quick development. A large-scale dataset with high dimensionality is the outcome of the dataset's dimensions growing at an incredible rate each day. The feature selection method is important in the world of big data today since it helps to reduce the dimensionality and overfitting of the learning process. Different algorithms are better suited to different sorts of data sets than others. As a result, we should apply all relevant models and assess the outcomes. Hyperparameter tuning is the process of selecting the optimal value for each hyperparameter to improve the model's accuracy. To adjust these hyperparameters, you must have a full understanding of their definitions and each one's impact on the model. This process may be repeated with many successful models. Ensemble techniques are the approach that is most commonly utilised in entries that are successful in data science competitions. This approach simply produces better results by integrating the output of numerous unsuccessful models. This may be done using boosting and bagging (Bootstrap Aggregating) in two ways. One of the most crucial data modelling techniques is cross-validation. Before finishing the model, it is advised to test the model on a sample that was not used to train it.

In Ref. [15], the authors introduce a personalised intelligent system that utilises ML techniques to predict and prevent mental health issues by integrating various features such as biometric data, lifestyle, behaviour, and social media activities. Many different learning methods and techniques are said to have been introduced by ML. Examples of frequently used ML techniques include supervised learning and unsupervised learning. With the use of input from labelled data, supervised learning predicts the outcome. Classification and regression issues are particularly well-suited to supervised learning. This learning aims to interpret the data about particular measures. Unsupervised learning stands in contrast to supervised learning, which seeks to interpret the data on its own. There are neither benchmarks nor rules in unsupervised learning. To solve a specific problem, the classifiers are also carefully combined and constructed throughout the ensemble learning process. Ensemble learning is typically used to improve model performance or lessen the chance that poorly performing models will be selected. Furthermore, neural networks and DL have recently grown in favour among ML approaches due to their capacity to address a variety of problems, such as picture identification, audio recognition, and natural language processing. Using these techniques, which are based on the neural networks of the brain, the computers may be able to learn from the observational data.

In Ref. [16], the authors propose a DL model called 3ACL that uses 3D MRI data to classify brain tumours, achieving high accuracy rates and outperforming existing models. By converting the 3D MRI data into 2D space using the 3t2FTS-v2, discriminative 2D-1D pictures with more significance are produced. In the 3t2FTS or 3t2FTS-v1 technique, FOS traits are assessed to create 2D-1D pictures that serve as a 3D voxel's identification. However, in the transition, only FOS traits are taken into account. Additionally, any additional feature extraction techniques and normalisation strategies that might produce reliable 2D-ID pictures by offering discriminatory data are not evaluated. Here, it is clear that the method's inspiration and design may be used to analyse both 3D brain tissue and tumours. Additionally, all versions may be used for the examination of 3D-defined brain illnesses like Alzheimer's as well as malignancies. In this approach, both versions may be used to activate common ML techniques (such as DL and transfer learning) that are based on 2D assessment.

The paper in Ref. [17] provides a thorough analysis of ML algorithms for the early diagnosis of mild cognitive impairment and AD while emphasising positive findings from several research studies. In the early stages of the condition, both pharmaceutical and nonpharmacological therapies are beneficial in lowering cognitive and behavioural symptoms. At this stage, it is important to note that AD should be viewed as a continuum, with people with MCI who may eventually develop AD dementia already having the disease even though the cognitive symptoms have not yet materialised. This makes it crucial to distinguish between MCI individuals who will develop AD dementia and those who will stay stable.

According to Ref. [18], the authors suggest a DL model that performs better than existing algorithms by learning a latent feature representation of the data for classifying AD. Convolutional adversarial autoencoders are proposed in Ref. [19] as a way for improving classification accuracy over existing models and minimising variability in multi-centre AD classification. Using a combination of self-supervised and semi-supervised learning approaches, the authors of [20] suggest a method for classifying medical images that outperforms current models and achieves high accuracy rates. Convolutional autoencoders are a DL technique that the authors suggest employing in Ref. [21] to explore the complex structure of AD and to classify the illness with great accuracy. The authors in Ref. [22] present a model for classifying AD that achieves excellent accuracy rates and outperforms previous algorithms by imputing missing data in multimodal brain pictures. The research study in Ref. [23] suggests an ensemble model for categorising AD that includes many deep networks, yielding greater accuracy than current algorithms. The research study in Ref. [24] presents an ensemble of deep neural networks trained by transfer learning for AD classification, outperforming existing models.

Recent DL models for AD diagnosis using neuroimaging features include 3D RSA-DNN [26], dual attention multi-instance DL [27], NMF-TDNet [28], 3D CNN-based multi-channel contrastive learning [29], lightweight framework [30], AlzheimerNet [31], DEMNET [32], multi-stream CNN [33], weakly supervised and multimodal SDPN [34], and deep multi-task multi-channel learning framework [35]. Accuracy ranged from 85.6% to 92.96% for these models.

Several research studies have looked into how DL and neuroimaging techniques may be used to diagnose AD and mild cognitive impairment (MCI). Li et al. [36] suggested a deep sparse autoencoder network that performed well when combining MRI and positron emission tomography data. PET is a minimally invasive functional imaging technology that uses positron emitters to provide crucial information about human biology and biochemistry. Because it is a three-dimensional (3D) approach, it calls for detectors with high-detection efficiencies and enhanced spatial resolution. Mehta et al. [37] developed a cascaded DL framework with improved performance on medical imaging datasets including Alzheimer's disease Neuroimaging Initiative (ADNI). Faisal and Kwon [38] created an automated identification technique based on a deep CNN model that performed well on the ADNI-2 dataset. Gamal et al. [39] suggested a set of 3D CNN models for the early detection of AD. Finally, Liu et al. [40] demonstrated an 85.6% accuracy on the ADNI dataset for a multimodal neuroimaging feature learning strategy for multiclass diagnosis of AD. These findings show that DL and neuroimaging approaches have enormous promise for early detection and individualised therapy of brain illnesses such as Alzheimer's and Parkinson's.

For MCI conversion prediction, a single-modal neuroimage may not be sufficient since it only captures a fraction of the data related to brain shrinkage. On the other hand, the multi-modal neuroimage can provide extra supplementary data that can be merged to comprehend the synergy between various neuroimages. Multi-modal fusion, which combines early, late, and mixed fusion approaches, is one of the frontiers of multi-modal ML. Multimodal learning is frequently utilised in the fields of photo categorisation and picture registration. Interest in multi-modal fusion stems from two major advantages. For starters, many modalities seeing the same phenomena might yield more robust predictions. Second, to increase the accuracy of classification findings, complementary information might be collected from several modalities using Sparse Auto Encoder. Deep learning techniques in particular, which have recently made significant strides in AI technology, have shown their positive effects on applications in the fields of genetic medicine and health care. The goal of computer-aided AI approaches like DL models is to allow data-driven algorithms that can, on the whole, help enhance the diagnosis accuracy of AD using neuroimaging and/or genetic data. For example, utilising neuroimaging data like MRI images, CNNs have been employed in DL models to diagnose AD. Since these promising approaches can easily be applied to many different data types, including neuroimaging and genomics, DL models have therefore been proposed to play a major role in the diagnosis and prediction of AD in the next study.

Memory, language, and emotional stability are all affected by the degenerative neurological disorder known as AD. The origin of AD is still unknown, and conventional diagnostic procedures are subjective and time-consuming. To avoid AD's severe stage and stop disease growth, early detection is essential. Alzheimer's disease is frequently diagnosed using MRI; however, conventional ML approaches for MRI classification have a restricted feature selection, which results in inconsistent classification performances. Deep learning models have recently demonstrated encouraging results in the diagnosis of AD; however, accuracy and performance still need to be improved. The literature review discussed different DL models proposed for AD diagnosis, including their advantages and challenges. Several recent studies have explored DL and neuroimaging techniques for diagnosing AD and mild cognitive impairment (MCI). These models achieved accuracies ranging from 85.6% to 92.96%. The paper aims to address the lack of an efficient method for AD diagnosis by proposing a DL model that selects features automatically and finds the optimal network structure. The goal is to contribute to early AD prevention and reduce physicians' burden. Overall, the review highlights the potential of DL and neuroimaging techniques for the personalised treatment of brain diseases such as AD and MCI.

Alzheimer's disease accounts for 60%–80% of dementia cases, making it the most prevalent cause of dementia. Alzheimer's disease begins with moderate cognitive impairment (MCI) and eventually deteriorates into a neurodegenerative type of dementia. It interferes with thinking, damages brain cells, causes memory loss, and makes it difficult to do simple activities. Alzheimer's disease is a multifaceted, degenerative neurological brain disease. Alzheimer's disease is more likely to develop in MCI patients than in the general population. The lack of a safe therapy for AD at this time is the largest problem facing scientists in the field. Nevertheless, modern AD treatments can reduce symptom severity or halt the onset of new ones. Therefore, it is crucial to catch AD early when it is still in the prodromal stage. Accurate and early AD identification is achieved via the use of computer-aided systems, reducing the high care costs associated with AD patients—costs that are predicted to climb sharply. Two types of features, namely region of interest (ROI)-based features and voxel-based features, are frequently used in early AD diagnosis using standard ML approaches.

While many research studies have concentrated on the application of DL and neuroimaging techniques for the diagnosis of AD and mild cognitive impairment (MCI), there is a gap in the literature regarding the incorporation of various features such as biometric data, lifestyle, behaviour, and social media activities to develop personalised intelligent systems for the prediction and prevention of mental health issues. Machine learning is an AI strategy that uses a variety of techniques to help an algorithm learn. Supervised, unsupervised, and DL are the three "learning" methods most frequently utilised in the healthcare industry. Other ML techniques include reinforcement learning and semi-supervised learning, which combine supervised and unsupervised learning. In these techniques, the algorithm functions as an agent in an interactive environment and learns by making mistakes and gaining rewards for its accomplishments. ML techniques do not differentiate between samples and populations; instead, they find informational patterns in data that may be used to forecast outcomes for specific patients. The descriptive component of statistics is comparable to ML, but the inferential component—which is the heart of statistics—is distinct since it employs only samples to form conclusions about the population from which the sample was taken. Modern ML methods are superior to conventional statistical methods because they can find complicated (non-linear), high-dimensional interactions that might help with predictions. Furthermore, given their excellent accuracy rates in other medical picture classification tasks, DL models with lightweight frameworks and multimodal SDPNs need to be investigated for their potential in AD diagnosis. Finally, given that they have demonstrated promising outcomes in other medical picture classification tasks, deep multi-task and multi-channel learning frameworks need to be investigated for their potential in AD diagnosis.

3 | PROPOSED WORK

The absence of an effective approach for diagnosing AD is discussed in the study report. Based on medical imaging, the conventional AD diagnosis is arbitrary and time-consuming. There are drawbacks to ML techniques, such as human feature selection that might result in the loss of crucial data. Alzheimer's disease is the most prevalent type of dementia; however, there are currently no treatments that can stop the condition. The dearth of reliable disease endpoints and/or biomarkers contributes to the scarcity of effective medications. Utilising ML to analyse EEG is one method to get past many of the limitations of the current diagnostic modalities. The problems include the necessity for very large sample sizes and exact EEG source localisation employing high-density devices, as well as the lack of automation and unbiased artefact removal. The use of irrational feature selection in black-box ML approaches such as deep neural networks is another problem. The paper suggests a DL approach for diagnosing AD that uses neuroimaging characteristics to automatically choose features and determine the best network layout. By putting out a DL model for AD diagnosis that performs better than current models, contributes to AD early prevention, and lessens the load on doctors, the goal is to close the research gap.

The proposed AD classification method is shown in Figure 1. This model mainly includes a basic network module, an attention module, and a multi-task learning module. The primary network module is used for feature extraction, including a total of 5 convolutional blocks (Block); the attention module is integrated into the midst of the main network to improve the network's ability to gather enough visual data; the multi-task learning module is introduced at the end of the primary network to supplement the knowledge of two additional tasks. Convolutional blocks are the fundamental building blocks of a CNN, which is used to recognise pictures. It consists of one or more convolutional layers that are used to take the characteristics of the input picture and extract them. One or more pooling layers are frequently employed to minimise the spatial dimensions of the feature maps while preserving the most crucial data after the convolutional layers. During one or more convolutional layers, filters are applied to the input picture to extract features like edges, textures, and forms. The feature maps are then downsampled by applying one or more pooling layers to the output of the convolutional layers. Due to the smaller spatial dimensions of the feature maps, the computational efficiency of the network is improved, and the likelihood of overfitting is decreased. To predict or categorise the image, one or more fully connected layers are subsequently applied to the output of the pooling layers. In the first phases of a CNN, convolutional blocks are utilised when the main objective is to extract characteristics from the input picture. In the network, they are frequently repeated numerous times, with each block collecting increasingly detailed information from the output of the one before it. The network can learn increasingly detailed information because of its hierarchical structure, which enhances picture recognition ability.

3.1 | Basic network

Brain MRI images are three-dimensional images. Many previous studies were based on the two-dimensional slice data of brain MRI images. Still, the work of obtaining 2D slice data is time-consuming, and easy to lose much important information. In order to effectively categorise AD, MCI, and HC participants based on MRI data, this work proposes a DL approach that makes use of 3D CNNs. The model seeks to improve AD diagnostic accuracy and lessen the restrictions of conventional ML techniques. The study also aims to contribute to the development of effective and impartial AD diagnostic approaches and to give insights into the possibilities of DL techniques for AD diagnosis. The absence of an effective approach for diagnosing AD is discussed in the study report. Based on medical imaging, the conventional AD diagnosis is arbitrary and time-consuming. There are drawbacks to ML techniques, such as human feature selection that might result in the loss of crucial data. The study proposes a DL model for AD diagnosis using neuroimaging features that select features automatically and find the optimal network structure. The aim is to fill the research gap by proposing an efficient and effective DL model for AD diagnosis that outperforms existing models, contributing to the early prevention of AD and reducing physicians' burden.

Based on previous research experience, this paper builds a three-dimensional volume. The three-dimensional convolutional network was first proposed by literature [12, 25] for human action analysis and recognition. Compared with the two-dimensional convolution, the three-dimensional convolution can better capture the sequence information and retain more. In this paper, the C3D network [13, 26, 27] is selected as the primary network to retrieve the features of images. The C3D network is a classic general-purpose network, first used for behaviour recognition and video feature extraction—one of the networks. Owing to a large amount of calculation of 3D MRI images and the apparent difference between medical image datasets and action recognition datasets, this paper improves its network structure to apply C3D to AD classification more quickly and effectively. The network structure is prescribed in Table 1.

The primary network retains five convolutional blocks of C3D, and each block contains a pooling layer to downsample the image, and the pooling layer can expand the receptive field. The experiments in the literature [13] prove that for the 3D convolutional network, all the convolution layers use a $3 \times 3 \times 3$ small convolution kernel to achieve the best results; so all convolution layers in the proposed primary network use a $3 \times 3 \times 3$ convolution kernel. Except for the pooling layer size and the step, the size is $1 \times 2 \times 2$. The size and measure length of all subsequent pooling layers are set to $2 \times 2 \times 2$; so if the input size of Block 1 is (N, C, D, H, and W), the output size of the final Block 5 is (N, C', D/16, H/32, and W/32), where N is the count of batches, C and C' are the count of channels, and D, H, W, are three-dimensional image information. The number of data sets is usually tiny; so the batch normalisation layer (Batch Normalisation, BN) is very suitable for medical image tasks. Its role is to accelerate network convergence and prevent overfitting. This paper adds a BN layer to the primary network, significantly improving the network performance and speeding up the training speed. After Block 5, the fully connected layer integrates the features, outputs the probability of different categories, and then uses the type corresponding to the maximum probability as the classification result.

For feature extraction in three subtasks and the primary classification task, the convolution layers are employed. A

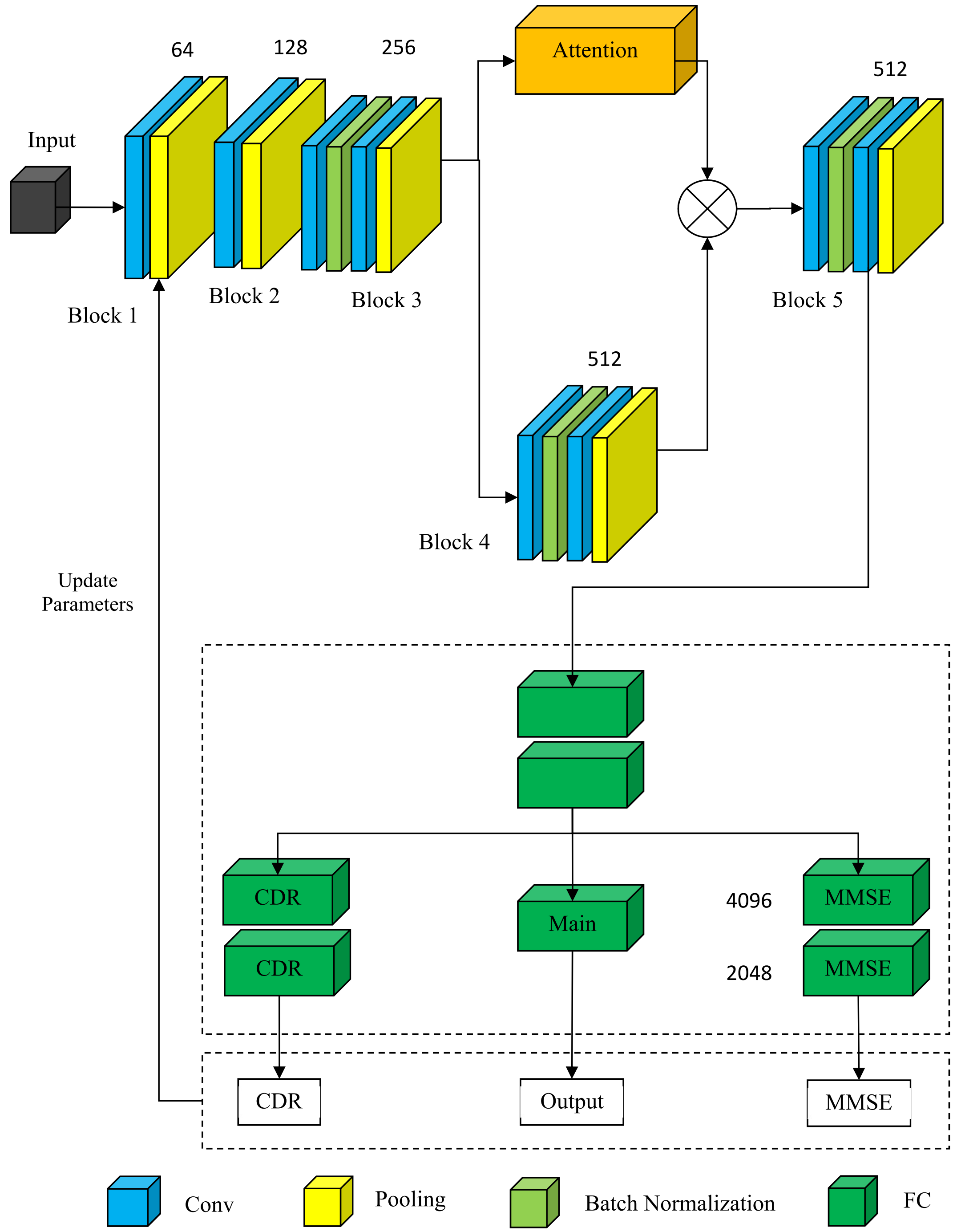

FIGURE 1 3D convolutional network based on the attention mechanism and multi-task learning.

merged layer that combines all learnt features from the sub-tasks is utilised after the dense layer. Before the output layers, all jobs employ two completely linked layers. To enhance performance generalisation, the multi-task learning paradigm employs associated tasks. By using a single architecture to share the same subset of parameters and provide an inductive bias between them during training, the hard parameter-sharing technique is a popular method for learning many tasks. The scientific and business worlds have been interested in it because of its simplicity, capacity to increase generalisation, and ability to lower processing costs. Determining how the gradients of many tasks should be mixed to facilitate simultaneous learning, however, is difficult since tasks frequently clash with one another. We employ the concept of multi-objective optimisation to solve this issue and offer a technique that takes into consideration the temporal behaviour of the gradients to build a dynamic bias that modifies the weighting of each job during backpropagation. To ensure that the simultaneous learning is leading to the performance maximum of all tasks, the method's outcome is to devote more attention to the tasks that are

TABLE 1 Structure of improved C3D network.

Structural units	Improve network
Block 1	3 × 3 × 3, 64, Conv1a
	1 × 2 × 2, Pool
Block 2	3 × 3 × 3, 128, Conv2a
	2 × 2 × 2, Pool
Block 3	3 × 3 × 3, 256, Conv3a
	256, batch normalisation
	3 × 3 × 3, 256, Conv3b
	2 × 2 × 2, Pool
Block 4	3 × 3 × 3,512, Conv4a
	512, batch normalisation
	3 × 3 × 3,512, Conv4b
	2 × 2 × 2, Pool
Block 5	3 × 3 × 3, 512, Conv5a
	512, batch normalisation
	3 × 3 × 3, 512, Conv5b
	2 × 2 × 2, Pool

diverging or that were not benefitted during the previous iterations. Therefore, we empirically demonstrate that the suggested strategy outperforms the cutting-edge methods for learning contradictory tasks.

3.2 | Attention module

The use of the attention mechanism has increased recently in related DL fields, and it has had positive effects on speech recognition, natural language processing, and picture processing [14, 28]. Speech recognition is a function that enables the programme to translate spoken language into written text. It is sometimes referred to as automatic speech recognition, computer speech recognition, or speech to text. Despite the availability of voice recognition software and hardware, the most sophisticated solutions rely on AI and ML. To comprehend and process spoken language, they incorporate grammar, syntax, signal composition, and audio and voice signal structure. They should ideally learn as they go, developing their replies with each engagement. The finest solutions also enable firms to modify and adjust the technology to their own needs, from brand recognition to language and voice idiosyncrasies.

Specifically, the area of “artificial intelligence” (AI) known as “natural language processing” (NLP) in computer science focuses on giving computers the ability to understand spoken and written language similarly to humans. NLP, or computational linguistics, is the use of rule-based modelling of human language with statistical, ML, and DL models. Together, these technologies enable computers to interpret human language in the form of text or audio data and to “understand” the whole meaning of what is being said or written, including the purpose and emotion of the speaker or writer. NLP is the driving force behind computer algorithms that translate text between languages, heed spoken commands, and swiftly sum up enormous volumes of information—even in real time.

Image processing is transforming an image into a digital format and performing particular operations on it to extract useful information. If particular defined signal processing techniques are applied, the image processing system normally interprets all images as 2D signals. The main component of an image processing system is a general-purpose computer, which can range in size from a PC to a supercomputer. To achieve a particular level of performance, specially constructed computers are occasionally used in specialised applications. The process of processing images begins with picture acquisition. Preprocessing is another name for this phase in the image processing process. It requires getting the picture from a source, often one that is hardware-based. The act of emphasising and bringing to light specific interesting aspects in a hidden image is known as image enhancement.

The attention mechanism's objective is to choose the most pertinent information from a massive amount of input, much to the visual attention mechanism in the human brain. After training, the attention weight in the test phase is fixed, but the input samples are different, and the final attention also varies. The value of the force is also added, that is, the attention of each piece is specific. Depending on the style of expression, the attention mechanism can be classified as either hard or soft [17, 32, 33]. A certain size is either paid attention to or not when it comes to complex attention, which is a 0/1 problem. Hard attention gives greater attention to areas or channels, using varying values between 0 and 1 to show the degree of attention to each region. Soft attention may be differentiated, and the model computes the gradient to determine the attention module's weight using backpropagation. The attention module in this paper is built using a spatial domain attention method based on soft attention. The spatial domain attention mechanism demonstrates that the information in other channel domains is what determines how much attention is paid to various spots on the picture.

A tensor of numbers in the range [0, 255] can be thought of as a picture. When using a set of words for categorisation, translation, or other NLP tasks, the original text Transformer utilises those words as input. To implement the Transformer design for ViT, we make the fewest changes necessary to convert it from operating on words to directly operating on images. We then see how much the model can learn about a visual structure on its own. Because ViT does not natively know the relative positions of patches in the image or even that the image has a 2D structure, it must acquire such essential knowledge from the training data and store structural information in the position embeddings. When used on ImageNet, where we initially train ViT, it earns a top-1 accuracy score of 77.9%. The best CNN trained on ImageNet with no additional data currently achieves 85.8%; so although this is respectable for a first try, it is far from the state of the art. ViT overfits the ImageNet task since it has no built-in understanding of pictures, despite mitigating techniques (such as regularisation). We train ViT on ImageNet-21k (14 million pictures, 21 thousand classes) and JFT (300 million images, 18 thousand classes) to examine the effect of dataset size on model performance. We then compare the findings to those of a state-of-the-art CNN named Big Transfer (BiT) trained on the same datasets. When trained on ImageNet (1M pictures), ViT severely underperforms the CNN counterpart (BiT), as was previously noticed.

As shown in Figure 2, in addition to the existing convolutional layer, pooling layer, and BN layer of the primary network, the Attention module also includes a normalisation layer that limits the output range. Convolutional layers are the fundamental components of CNNs. Applying a filter to an input to create an activation is a simple procedure known as convolution. By continuously using the same filter on an input, such as a picture, a feature map is created that displays the locations and degrees of a recognised feature in the input. Convolutional neural networks are innovative in that they may adhere to the constraints of a specific predictive modelling problem, such as image classification, while automatically learning a large number of parallel filters customised to a training dataset. As a result, input photographs may have incredibly unique characteristics that are not found anywhere else. Convolutional neural networks, or CNNs, are a specific type of the neural network model developed for use with two-dimensional image data, while they may also be utilised with one- and three-dimensional data. This layer performs "convolution", a procedure. A linear process called convolution in the context of a CNN entails multiplying a set of weights with the input, much like in a conventional neural network. Given that the method was designed for two-dimensional input, a two-dimensional array of weights called a filter or kernel is multiplied by an array of input data.

The following takes the Attention module as an example to introduce the details of each network layer. Assuming that the output of Block 3 in Figure 1 is F_3, F_3 is inputted to the Attention module and Block 4 simultaneously, the former first passes through a convolutional layer, and the formula of the convolutional layer is shown in Formula (1).

$$F^i_{Att\_l} = \sum_{j \in M^N} F^j_N * K^{ij}_{Att\_l} + b^i_{Att\_l} \tag{1}$$

Among them, $F^i_{Att\_l}$ refers to the ith feature map with respect to the nth network layer in the Attention module, F^j_N refers to the jth feature map output by Block N, $K^{ij}_{Att\_l}$ refers to the convolution kernel connecting $F^i_{Att\_l}$ and F^j_N in the nth network layer of the Attention module, $*$ represents the convolution operation, $b^i_{Att\_l}$ is the bias item, and M^N represents the number of feature maps output by Block N.

The convolutional layer updates parameters during training, which will cause changes in the distribution of subsequent input data. The BN layer can solve the above problems. The principle is to normalise the intermediate data of network. The computation formula of the BN layer is as follows:

$$y_i = \gamma \left(\frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \right) + \beta \tag{2}$$

Among them, x_i refers to the input data; y_i represents the output; μ represents the mean value; σ^2 represents the variance; γ and β are learnable reconstruction parameters introduced to ensure that the learnt feature distribution remains unchanged and are related to the channel dimension. Before the network activation function, the forward conduction formula after introducing the BN layer is as follows:

$$\mathrm{F}_{\mathrm{ATT\_2}} = \mathrm{f}\left(\mathrm{BN}(\mathrm{F}_{\mathrm{ATT\text{-}1}})\right) \tag{3}$$

Among them, f(.) represents the linear rectification function (Rectified Linear Unit, ReLU), BN(.) represents the normalisation processing of the BN layer, $\mathrm{F}_{\mathrm{Att\_1}}$ represents the output with respect to the previous convolution, and $\mathrm{F}_{\mathrm{Att\_2}}$ represents the output after the BN layer and the activation function.

From Formula (1), the output $\mathrm{F}_{\mathrm{Att\_3}}$ of the next convolutional layer can be obtained and inputted to the pooling layer. This paper adopts the maximum pooling method, and the maximum value in the area is taken as the output, which is issued to decrease the resolution and increase the receptive field, and its calculation formula can be expressed as follows:

$$F^i_{Att\_4} = w^i_{Att\_4} \cdot \mathrm{down}\left(f\left(F^i_{Att\_3}\right)\right) + b^i_{Att\_4} \tag{4}$$

Among them, $w^i_{Att\_4}$ represents the pooling parameter of the ith feature map, down $(\cdot)$ represents the pooling function and $F^i_{Att\_3}$ refers the ith feature map of the previous output. After pooling, $\mathrm{F}_{\mathrm{Att\_5}}$ is obtained from Formula (1) through the convolutional layer and then passed to the normalisation layer. The function of the normalisation layer is to output the previous layer. Converted to probability, its calculation formula is as follows:

$$F^i_{Att} = \frac{F^i_{Att\_5} - \min\left(F_{latten}\left(F^i_{Att\_5}\right)\right)}{\mathrm{sum}\left(F^i_{Att\_5}\right)} \tag{5}$$

Among them, F^i_{Att} refers to the ith feature map output by the Attention module, and $F^i_{Att\_5}$ refers to the ith feature map output by the previous convolution. $Flatten(\cdot)$ is an expansion function that returns a folded one-dimensional array. $min(\cdot)$ represents the calculation of the minimum value of each row of the current feature map, and $sum(\cdot)$ means calculating each row's sum.

According to the above formula, the output F_4 of Block 4 can be obtained in the same way: Assuming that the input size is (b, c, d, h, and w), then the size of F4 is (b, 2∗c, d/2, h/2, and w/2), the output size of the Attention module is (b, 1, d/2, h/2, and w/2). The input of Block 5 is the product of the two, as shown in Formula (6):

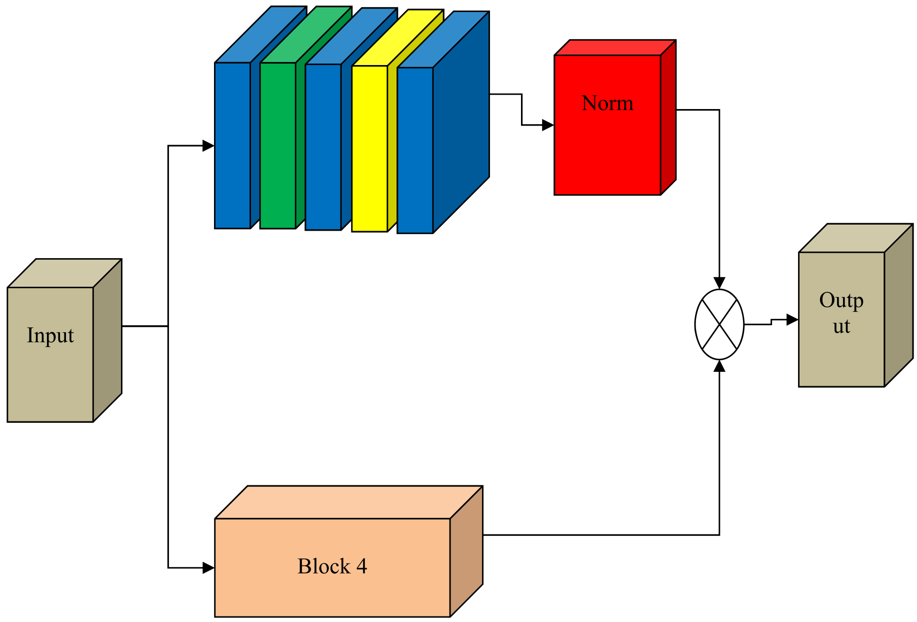

FIGURE 2 Attention module.

$$F = F_{Att} \cdot F_4^i, i = 1, 2, \cdots, 2 * c \tag{6}$$

More specifically, the output of the Attention module is multiplied by each feature map output by Block 4 to realise the attention mechanism's introduction.

3.3 | Multi-task learning module

Multi-task learning transfers the knowledge points of one task to other tasks; the purpose is to use the valuable information contained in multiple studies to learn more accurately; the key is that each job is related, and various sub-tasks are trained in parallel. Furthermore, multi-task learning can share parameters [18, 34, 35], improving the model's generalisation ability to a certain extent. In deep education, there are often two types of multi-tasking: parameter soft-sharing mechanisms and parameter hard-sharing mechanisms. Each job has its own parameters and models; the latter shares the hidden layer's parameters, and the only output layer that preserves distinct parameters is the one for that task. Compared with the two, the parameter hard-sharing mechanism has fewer parameters, which can effectively prevent the model from overfitting, making it more suitable for the small amount of medical image data.

This paper uses multi-task learning on the basis of the parameter hard-sharing mechanism in order to realise AD classification. The main task is AD classification, and the auxiliary sub-tasks are set to CDR score regression and MMSE score regression. Evaluating a patient's cognitive impairment degree is finally integrated into a total score; MMSE is one of the most influential cognitive impairment screening tools. Both are correlated with AD classification and can be used as auxiliary subtasks.

The way to achieve the proposed multi-task learning is shown in the Multi-task module in Figure 1. The final classification probability and CDR and MMSE scores are obtained using the fully connected layer. The computation formula of the fully connected layer is as follows:

$$y^i = f\left(w^i \cdot \left(\text{ Flatten }\left(x^i\right)\right) + b^i\right) \tag{7}$$

Among them, x^i represents the ith feature map of the input, and w^i defines the fully connected parameters. The channels of the two thoroughly combined layers shared are 4096 and 2048 in turn. An additional dropout layer is added to discard neurons with a certain probability to prevent overfitting and improve the training speed. The main task is followed by adding a fully connected layer with a channel of 2, and the CDR and MMSE tasks are followed by adding two thoroughly combined layers. These three tasks are trained simultaneously, sharing all parameters before the branch and updating according to backpropagation model parameters with the primary task of AD classification as the final output.

The loss function consists of the classification loss of the actual label and the predicted result in the AD classification task, the CDR regression loss and the MMSE regression loss in the auxiliary job. The computation method of the loss function is shown in Formula (8):

$$\text{Loss} = L_{\text{main}} + \alpha L_{\text{CDR}} + \lambda L_{\text{MMSE}} \tag{8}$$

Among them, L_{main} represents AD classification loss, L_{CDR} represents CDR regression loss, L_{MMSE} represents MMSE regression loss, and α and λ are loss balance coefficients.

L_{main} utilises the cross-entropy loss function, L_{CDR} and L_{MMSE} use the Huber loss function, and the computation formula of the latter is as follows:

$$L_\delta(I_i, y_i) = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}\frac{1}{2}(I_i - y_i)^2, |I_i - y_i| \leqslant \delta \\ \delta|I_i - y_i| - \frac{1}{2}\delta^2, |I_i - y_i| > \delta\end{cases} \tag{9}$$

Among them, I_i is the actual label of the ith sample, y_i is predicted value, N is the total number of pieces, and δ is the hyperparameter.

4 | DATASET AND PREPROCESSING

4.1 | Experimental data

The public dataset for the AD Neuroimaging Initiative (ADNI), made available by Michael W. The ADNI dataset consists of four subsets: ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. This study on the categorisation of AD makes use of MRI data from ADNI-1 that was gathered between 2004 and 2009 using the 1.5 T protocol. The experimental data consisted of 638 samples that were split into three groups: the training set, the verification set, and the test set. These groups made up 70%, 10%, and 20% of the total data, respectively.

4.2 | Image preprocessing

An example of the MRI image in the ADNI dataset is depicted in Figure 3. From left to right are the Coronal, Sagittal, and Axial directions of the MRI image, corresponding to the front-to-back, left-to-right, and top-to-bottom sections. Subjects during MRI scan imaging are also images of different angles in the obtained MRI images due to head down, tilted, and backwards. Brain MRI images also contain non-brain parts (such as skulls) that are useless for AD classification non-brain parts are noise signals for classification tasks.

Given the above problems, the preprocessing process of Sagittal direction, as an example, is illustrated in Figure 4, and the precise steps are mentioned below.

Step 1 Origin correction. Use the MATLAB Statistical Parameters Mapping (Statistical Parameters Mapping, SPM) toolbox to correct the origin of the brain MRI image using the AC-PC origin correction method, AC is the anterior commissure (Anterior Commissure), and PC is the posterior commissure (Posterior Commissure). First, adjust the vertical line of the crosshair to roughly pass through the middle of the coronal and axial views so that the intersection point coincides with AC as much as possible, and then adjust the horizontal line of the sagittal crosshair to pass through AC and PC to achieve origin correction.

Step 2 Remove the skull. The process of removing the head is divided into three phases. First, use the SPM tool to separate the MRI image after origin correction into three parts: grey matter, white matter, and other brain tissues, and then add these three parts directly to form a brain binary value. Finally, dot multiplication between the template and the origin-corrected MRI image to obtain the image without the skull.

Step 3 Unify the size. The MRI images in the data set have multiple sizes: 256 × 256 × 166, 192 × 192 × 160 etc. To reduce the calculation burden of the 3D CNN, the MRI image, after the skull was removed, was down-sampled. Finally, an MRI image with a size of 79 × 95 × 79 was obtained.

Step 4 Data augmentation. Due to the lack of experimental data, data augmentation is realised by adding noise, random flipping, and normalisation to the preprocessed MRI images.

5 | EXPERIMENTATION AND ANALYSIS

5.1 | Evaluation criteria

To quantitatively evaluate the proposed method, the most commonly used accuracy (Accuracy, Acc), sensitivity (Sensitivity, Sens), and specificity (Specificity, Spec) is used as the evaluation indicators of the AD classification algorithm on the ADNI dataset. The calculation formulas are as follows:

$$A_{cc} = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$Sens = \frac{TP}{TP + FN} \tag{11}$$

$$Spec = \frac{TN}{FP + TN} \tag{12}$$

FP stands for false positive, referring to the proportion of healthy samples that are incorrectly classified as patients; True positive, abbreviated as TP, refers to the proportion of patient samples that were accurately diagnosed; an assessment of the percentage of correctly diagnosed healthy samples is known as True Negative, or TN; The number of incorrectly positive samples is referred to as the "false negative" (FN) rate. Cut into pieces that fit, a number of patient samples were used.

5.2 | Experimental setup and parameter analysis

In this study, TensorFlow [19] was utilised as the DL framework, and after pre-training on the action recognition dataset UCF101, the convolutional layer parameters of the C3D model were used to initialise the parameters of the matching network layer. The adaptive moment estimation (Adam) optimisation

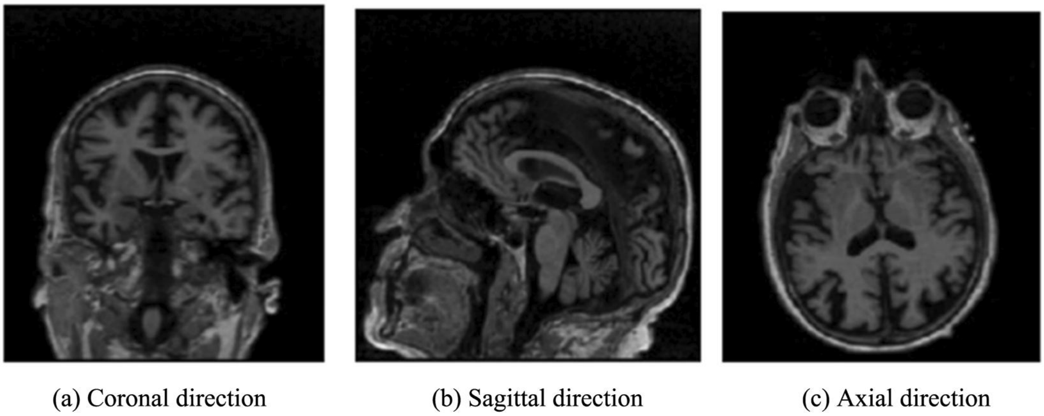

FIGURE 3 Magnetic resonance imaging (MRI) image example.

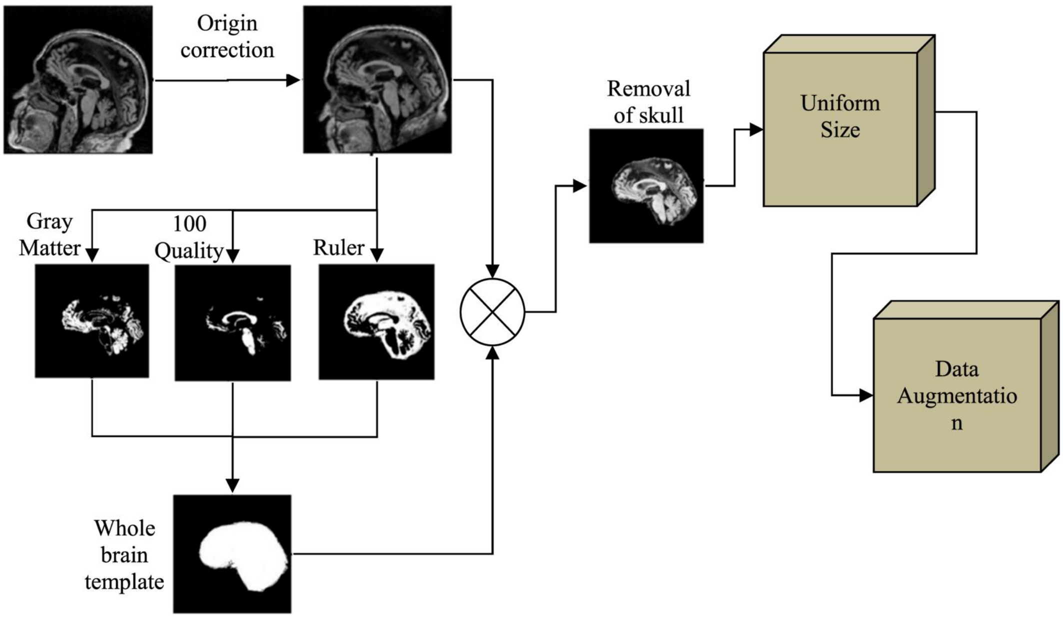

FIGURE 4 Preprocessing flow.

technique was applied, with a total of 10,000 training repetitions, an initial learning rate of 0.0001, and a reduction in learning rate to 1/20 of the original for each iteration of 25 epochs. The number of learning samples used each time was 8, and the hyperparameters δCDR and δMMSE in Formula (9) were set to 0.5 and 7.0, respectively. The loss balance coefficients of auxiliary subtasks CDR regression and MMSE regression were represented by α and λ, respectively, in Formula (8). Since the main task of the model is AD classification, the value range of α and λ typically ranges from 0 to 1. In our HC experiment, we first fixed $\lambda = 0$ and then set α to take values at intervals of 0.1. Figure 5 shows the model's performance on the verification set after training when α takes different values. We found that when α is 1.0, the AD versus HC classification effect is the best; so in the follow-up experiment about multi-task learning, AD versus α was set to 1.0 in HC experiments.

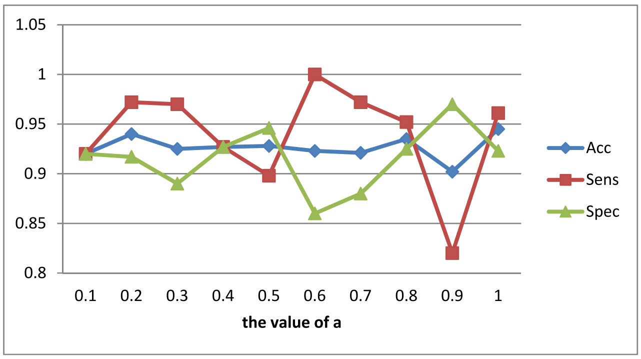

FIGURE 5 Performance of different values of α (AD versus HC).

According to the above experiments, the hyperparameter $\alpha = 1.0$ is fixed, based on ADvs. The parameter λ was adjusted in the HC experiment. The value of λ also takes 0.1 as an interval, and it can be seen from Figure 6 that when $\lambda = 0.3$, the classification effect of the HC experiment is the best. Therefore, for ADvs, in the HC experiment, α was set to 1.0, and λ was set to 0.3.

In the same way, based on MCIvs, the hyperparameters α and λ of the HC experiment can be seen in Figures 7 and 8 that α and λ are 0.2 and 0.6, respectively. In all subsequent experiments based on multi-task learning, α and λ are set according to this conclusion.

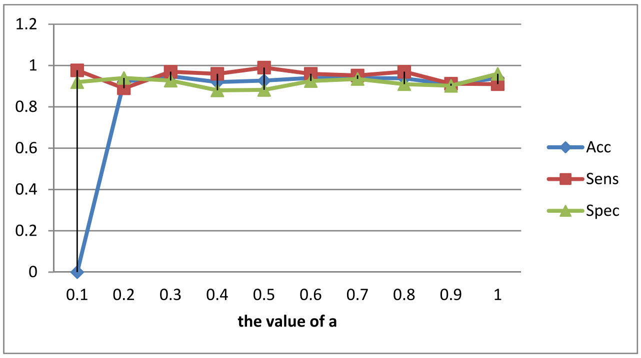

FIGURE 6 Performance of different values of λ (AD versus HC).

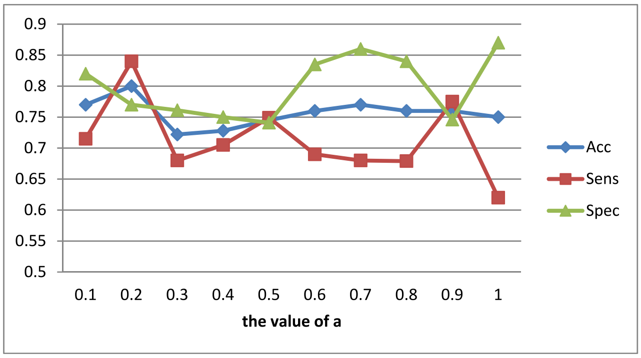

FIGURE 7 Performance of different values of α (MCI versus HC).

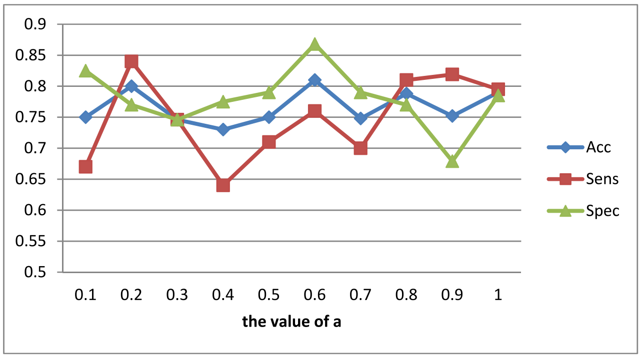

FIGURE 8 Performance of different values of λ (MCI versus HC).

5.3 | Experimental comparison

A model ablation experiment was designed to verify the proposed method's effectiveness, and the results are shown in Table 2.

On the whole MCIvs, the accuracy rate of HC experiments is generally higher than that of AD vs HC experiment which is low because the brain atrophy of the MCI patients is closer to that of HC subjects than that of AD patients, and it is more difficult to distinguish. The C3D + BN model means that the BN layer is introduced based on C3D, and it can be observed from Table 2 that its effect on MCI vs HC is better. The improvement effect of the HC experiment is significant. Before adding the BN layer, the network is difficult to compare the MCI and HC subjects, and the accuracy rate is as low as 0.5, equivalent to random guessing. It is speculated that this is due to the UCF101 data used in pre-training the original C3D network. The set is human action video data, far from the medical classification task. When the C3D network is trained on brain MRI images, the amount of MCI and HC data is too small, which leads to overfitting, and the accuracy rate is extremely high on the test set. The C3D + BN + Attention model means that the attention mechanism is introduced separately based on the former.

The findings in Table 2 demonstrate that it mostly enhances ADvs and the classification accuracy rate of HC trials. The results show that the C3D + BN + Multitask model, which represents the introduction of multi-task learning alone, is more advantageous to the MCIvs HC experiment than the ADvs, for whom the HC experiment has a less favourable impact on the enhancement of the attention mechanism. After both are introduced (the proposed algorithm), compared with the primary C3D network, ADvs, the accuracy of HC experiments increased by 7%, MCIvs. The accuracy of the HC experiment is also increased from 0.5 to 0.8097, which proves the effectiveness of the BN layer, multi-task learning and attention mechanism in improving the accuracy of AD classification.

Table 3 compares the proposed algorithm with the experimental results of some recent studies. For example, literature [20] uses the recurrent gated unit of the RNN to extract slice features for final classification; literature [21] introduces ensemble learning into the convolutional network and proposes a multi-slice ensemble classification model; literature [22] analysed the shape of the hippocampus and used a 3D densely connected convolutional network for AD

TABLE 2 Comparison of ablation experiments.

	Acc	
Model	AD versus HC	MCI versus HC
C3D	0.8721	0.5000
C3D + BN	0.8983	0.7273
C3D + BN + Attention	0.9186	0.7386
C3D + BN + Multitask	0.9157	0.7642
Presented work	0.9448	0.8097

TABLE 3 Comparison of classification performance.

	AD versus HC			MCI versus HC		
Model	**Acc**	**Sens**	**Spec**	**Acc**	**Sens**	**Spec**
Literature [20] algorithm	0.9119	0.9140	0.9100	0.7886	0.7808	0.8000
Literature [21] algorithm	0.8103	-	-	-	-	-
Literature [22] algorithm	0.9229	0.9063	0.9372	0.7464	0.7727	0.6996
Presented work	0.9448	0.9722	0.9250	0.8097	0.8352	0.7670

classification. The results in Table 3 show that, in comparison to the first two methods, the proposed algorithm has AD versus HC/MCIvs, which has a distinct competitive advantage in classification performance. The accuracy rate of the HC experiment increased by 3%–13%/2%, and the sensitivity increased by 6%/5.5%. This is because the first two methods are limited to 2D slice data; however, extracting slices has been dramatically improved compared with the traditional method of removing the middle position of the slice sequence interval; its reserved features are still incomplete, which affects the classification performance.

The third method also uses a three-dimensional convolutional network, but it differs from the algorithm in this paper, its MCIvs. The classification performance of HC is poor, with a difference of 6% in accuracy, 6% in sensitivity, and 7% in specificity in the experiment. This is so that the proposed algorithm can more effectively use the clinical correlation between CDR, MMSE score, and AD classification, as well as the feature information of 3D MRI images based on the attention mechanism, and to improve the performance of the MCIvs HC experimental classification. It is important to note that, in comparison to these three techniques, the suggested algorithm's specificity is subpar. The percentage of HC participants that were accurately classified as disease-free is essentially what defines specificity. Patients diagnosed with the disease usually undergo further follow-up examinations to correct the misdiagnosis, so a certain degree of specificity can be sacrificed to ensure higher sensitivity.

6 | CONCLUSION

Given the aforementioned issues, this paper proposes an end-to-end AD classification method based on a deep 3-D CNN (CNN), which includes an attention mechanism to improve feature extraction capabilities and two auxiliary subtasks, namely CDR Scale (CDR) score regression and MMSE score regression, to optimise AD classification results and realise multi-task learning. The attention mechanism is then implemented, and the 3-dimensional convolutional network is utilised to acquire features with attention weights, which may gain the most advantageous picture position information for classification tasks. In order to properly extract medical picture features and avoid wasting clinical auxiliary information resources for AD classification, this article used deep 3-DCNN and proposed an AD classification algorithm. This study conducts tests on the ADNI data set to demonstrate that the suggested model can effectively classify AD and outperforms competing algorithms. The public dataset for the Michael W. Neal AD Neuroimaging Initiative (ADNI) was made accessible. ADNI-1, ADNI-GO, ADNI-2, and ADNI-3 are the four subsets that make up the ADNI dataset. The MRI data from ADNI-1 that was collected between 2004 and 2009 using the 1.5 T procedure is used in this study on the classification of AD. The 638 samples that made up the experimental data were divided into three groups: the training set, the verification set, and the test set. 70%, 10%, and 20%, respectively, of the total data came from these groups. However, in the proposed algorithm, when comparing MCIvs with ADvs, the performance of HC experiments is poor. In the future, we will further explore the structural features of MCI patient images by combining multi-scale information, introducing adversarial samples, and optimising the classification loss function to adapt to unbalanced sample classification during the training task.

ACKNOWLEDGEMENTS

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through General Research Project under grant number (GRP/75/44).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author.

ORCID

Renato R. Maaliw III https://orcid.org/0000-0002-7310-2708
Mohammad Shabaz https://orcid.org/0000-0001-5106-7609

REFERENCES

1. Shojaei, S., Abadeh, M.S., Momeni, Z.: An evolutionary explainable deep learning approach for Alzheimer's MRI classification. Expert Systems with Applications 220, 119709 (2023). ISSN 0957-4174. https://doi.org/10.1016/j.eswa.2023.119709
2. SubramanyamRallabandi, V.P., Seetharaman, K.: Deep learning-based classification of healthy aging controls, mild cognitive impairment and Alzheimer's disease using fusion of MRI-PET imaging. Biomedical Signal Processing and Control 80(Part 1), 104312 (2023). ISSN 1746-8094. https://doi.org/10.1016/j.bspc.2022.104312
3. Alorf, A., Khan, M.U.G.: Multi-label classification of Alzheimer's disease stages from resting-state fMRI-based correlation connectivity data and deep learning. Computers in Biology and Medicine 151(Part A), 106240 (2022). ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2022.106240
4. EL-Geneedy, M., et al.: An MRI-based deep learning approach for accurate detection of Alzheimer's disease. Alexandria Engineering Journal 63, 211–221 (2023). ISSN 1110-0168. https://doi.org/10.1016/j.aej.2022.07.062
5. Han, R., Liu, Z., Philip Chen, C.L.: Multi-scale 3D convolution feature-based broad learning system for Alzheimer's disease diagnosis via MRI

images. Applied Soft Computing 120, 108660 (2022). ISSN 1568-4946. https://doi.org/10.1016/j.asoc.2022.108660
6. Sharma, S., et al.: A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. Measurement: Sensors 24, 100506 (2022). ISSN 2665-9174. https://doi.org/10.1016/j.measen.2022.100506
7. Liu, F., et al.: Patch-based deep multi-modal learning framework for Alzheimer's disease diagnosis using multi-view neuroimaging. Biomedical Signal Processing and Control 80(Part 2), 104400 (2023). ISSN 1746-8094. https://doi.org/10.1016/j.bspc.2022.104400
8. Liang, W., et al.: Rethinking modeling Alzheimer's disease progression from a multi-task learning perspective with deep recurrent neural network. Computers in Biology and Medicine 138, 104935 (2021). ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2021.104935
9. Liu, Y., Mazumdar, S., Bath, P.A.: An unsupervised learning approach to diagnosing Alzheimer's disease using brain magnetic resonance imaging scans. International Journal of Medical Informatics 173, 105027 (2023). ISSN 1386-5056. https://doi.org/10.1016/j.ijmedinf.2023.105027
10. Sharma, R., et al.: FDN-ADNet: fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans. Applied Soft Computing 115, 108099 (2022). ISSN 1568-4946. https://doi.org/10.1016/j.asoc.2021.108099
11. Loddo, A., Buttau, S., Di Ruberto, C.: Deep learning based pipelines for Alzheimer's disease diagnosis: a comparative study and a novel deep-ensemble method. Computers in Biology and Medicine 141, 105032 (2022). ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2021.105032
12. Zhao, Y., et al.: Multi-task deep learning for medical image computing and analysis: a review. Computers in Biology and Medicine 153, 106496 (2023)
13. Zhao, Y., et al.: Multi-task deep learning for medical image computing and analysis: a review. Computers in Biology and Medicine 153, 106496 (2023). ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2022.106496
14. Hu, Z., et al.: VGG-TSwinformer: Transformer-Based Deep Learning Model for Early Alzheimer's Disease Prediction
15. Hu, Z., et al.: VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer's disease prediction. Computer Methods and Programs in Biomedicine 229, 107291 (2023). ISSN 0169-2607. https://doi.org/10.1016/j.cmpb.2022.107291
16. Demir, F., et al.: Improving brain tumor classification performance with an effective approach based on new deep learning model named 3ACL from 3D MRI data. Biomedical Signal Processing and Control 81, 104424 (2023). ISSN 1746-8094. https://doi.org/10.1016/j.bspc.2022.104424
17. Muhammed Niyas, K.P., Thiyagarajan, P.: A systematic review on early prediction of Mild cognitive impairment to alzheimers using machine learning algorithms. International Journal of Intelligent Networks 4, 74–88 (2023). ISSN 2666-6030. https://doi.org/10.1016/j.ijin.2023.03.004
18. Dong, A., et al.: Latent feature representation learning for Alzheimer's disease classification. Computers in Biology and Medicine 150, 106116 (2022). ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2022.106116
19. Cobbinah, B.M., et al.: Wei Han, LiweiChe, JunmingShao, Reducing variations in multi-center Alzheimer's disease classification with convolutional adversarial autoencoder. Medical Image Analysis 82, 102585 (2022). ISSN 1361-8415. https://doi.org/10.1016/j.media.2022.102585
20. Wang, Y., et al.: Self-supervised learning and semi-supervised learning for multi-sequence medical image classification. Neurocomputing 513, 383–394 (2022). ISSN 0925-2312. https://doi.org/10.1016/j.neucom.2022.09.097
21. Martinez-Murcia, F.J., et al.: Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders. IEEE Journal of Biomedical and Health Informatics 24(1), 17–26 (2020). https://doi.org/10.1109/JBHI.2019.2914970
22. Gao, X., et al.: Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimer's disease. IEEE Journal of Biomedical and Health Informatics 26(1), 36–43 (2022). https://doi.org/10.1109/JBHI.2021.3097721
23. Choi, J.Y., Lee, B.: Combining of multiple deep networks via ensemble generalization loss, based on MRI images, for Alzheimer's disease classification. IEEE Signal Processing Letters 27, 206–210 (2020). https://doi.org/10.1109/LSP.2020.2964161
24. Tanveer, M., et al.: Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. IEEE Journal of Biomedical and Health Informatics 26(4), 1453–1463 (2022). https://doi.org/10.1109/JBHI.2021.3083274
25. Zhang, X., et al.: An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Journal of Biomedical and Health Informatics 26(11), 5289–5297 (2022). https://doi.org/10.1109/JBHI.2021.3066832
26. Zhu, W., et al.: Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI. IEEE Transactions on Medical Imaging 40(9), 2354–2366 (2021). https://doi.org/10.1109/TMI.2021.3077079
27. Lao, H., Zhang, X.: Regression and classification of Alzheimer's disease diagnosis using NMF-TDNet features from 3D brain MR image. IEEE Journal of Biomedical and Health Informatics 26(3), 1103–1115 (2022). https://doi.org/10.1109/JBHI.2021.3113668
28. Li, J., et al.: 3-D CNN-based multichannel contrastive learning for Alzheimer's disease automatic diagnosis. IEEE Transactions on Instrumentation and Measurement 71, 1–11 (2022). Art no. 5008411. https://doi.org/10.1109/TIM.2022.3162265
29. Rashid, A.H., et al.: Biceph-net: a robust and lightweight framework for the diagnosis of Alzheimer's disease using 2D-MRI scans and deep similarity learning. IEEE Journal of Biomedical and Health Informatics 27(3), 1205–1213 (2023). https://doi.org/10.1109/JBHI.2022.3174033
30. Shamrat, F.M.J.M., et al.: AlzheimerNet: an effective deep learning based proposition for Alzheimer's disease stages classification from functional brain changes in magnetic resonance images. IEEE Access 11, 16376–-16395 (2023). https://doi.org/10.1109/ACCESS.2023.3244952
31. Murugan, S., et al.: DEMNET: a deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. IEEE Access 9, 90319–90329 (2021). https://doi.org/10.1109/ACCESS.2021.3090474
32. Ashtari-Majlan, M., Seifi, A., Dehshibi, M.M.: A multi-stream convolutional neural network for classification of progressive MCI in Alzheimer's disease using structural MRI images. IEEE Journal of Biomedical and Health Informatics 26(8), 3918–3926 (2022). https://doi.org/10.1109/JBHI.2022.3155705
33. Liu, M., et al.: Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores. IEEE Transactions on Cybernetics 50(7), 3381–3392 (2020). https://doi.org/10.1109/TCYB.2019.2904186
34. Shi, J., et al.: Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE Journal of Biomedical and Health Informatics 22(1), 173–183 (2018). https://doi.org/10.1109/JBHI.2017.2655720
35. Liu, M., et al.: Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. IEEE Transactions on Biomedical Engineering 66(5), 1195–1206 (2019). https://doi.org/10.1109/TBME.2018.2869989
36. Li, F., et al.: A robust deep model for improved classification of AD/MCI patients. IEEE Journal of Biomedical and Health Informatics 19(5), 1610–1616 (2015). https://doi.org/10.1109/JBHI.2015.2429556
37. Mehta, R., et al.: Propagating uncertainty across cascaded medical imaging tasks for improved deep learning inference. IEEE Transactions on

Medical Imaging 41(2), 360–373 (2022). https://doi.org/10.1109/TMI.2021.3114097
38. Faisal, F.U.R., Kwon, G.-R.: Automated detection of Alzheimer's disease and mild cognitive impairment using whole brain MRI. IEEE Access 10, 65055–65066 (2022). https://doi.org/10.1109/ACCESS.2022.3180073
39. Gamal, A., Elattar, M., Selim, S.: Automatic early diagnosis of Alzheimer's disease using 3D deep ensemble approach. IEEE Access 10, 115974–115987 (2022). https://doi.org/10.1109/ACCESS.2022.3218621
40. Liu, S., et al.: Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Transactions on Biomedical Engineering 62(4), 1132–1140 (2015). https://doi.org/10.1109/TBME.2014.2372011

How to cite this article: Dhaygude, A.D., et al.: Knowledge-based deep learning system for classifying Alzheimer's disease for multi-task learning. CAAI Trans. Intell. Technol. 1–16 (2024). https://doi.org/10.1049/cit2.12291