Iraqi Journal of Science, 2026, Vol. 67, No. 10, pp: 5464-5477
DOI: 10.24996/ijs.2026.67.10.27

ISSN: 0067-2904

Automated Alzheimer's Disease Stage Prediction Using a Deep Learning Combined Approach

Ahmad Adel Habeeb [1]*, Bushra Mahdi Hasan [1], Ammar A. Mousa Alabbassi [2]
[1]*Department of computer science, College of Education, University of Kufa, Najaf, Iraq*
[2]*Department of Electronics and communication, College of Engineering, University of Kufa, Najaf, Iraq*

Received: 22/12/2024 Accepted: 11/2/2026 Published: 30/10/2026

Abstract

Alzheimer's Disease (AD) is a neurological ailment that impacts memory, thinking, and behavior, leading to a decline in daily activities. It affects around 55 million people worldwide; this number is expected to increase as the world's population gets older, making it one of the most significant public health challenges of the present day. This study presents a deep learning approach for diagnosing AD stages using Magnetic Resonance Imaging (MRI). The experiments were performed using two different datasets that comprise a large number of brain MRI scans. Our proposed method combines the Visual Geometry Group 16-layer network (VGG16) for extracting spatial features along with a Long Short-Term Memory (LSTM) network to learn temporal relationships between tomographic slices. The input images were preprocessed and features derived through standard methods before conducting the classification. The proposed architecture was implemented using deep learning frameworks like TensorFlow and Keras, then evaluated using accuracy, precision, recall, and F1-score. The experimental results achieved more than 99% accuracy the training and about 98% on the held-out validation set from 8k MRI images for identifying AD stages. The VGG16-LSTM deep learning architecture demonstrated high performance in classifying AD stages from MRI and achieved high accuracy. This research presents a promising approach for diagnosing AD that has potential for clinical relevance.

Keywords: Image Processing, Deep Learning, VGG16, LSTM, RNN, Alzheimer's Disease, Machine Learning.

تنبؤ تلقائي بمراحل مرض الزهايمر باستخدام نهج مركب من التعلم العميق

احمد عادل حبيب[1], بشرى مهدي حسن[1]، عمار عبد الزهرة موسى العباسي[2]
[1]قسم علوم الحاسوب، كلية التربية، جامعة الكوفة، النجف، العراق
[2] قسم الكترونيات الاتصالات، كلية الهندسة، جامعة الكوفة، النجف، العراق

الخلاصة

*Email: ahmeda.alkheiat@uokufa.edu.iq

Open Access Articles. This articles is licensed under

مرض الزهايمر (AD) هو اضطراب عصبي يؤثر في الذاكرة والتفكير والسلوك، مما يؤدي إلى تراجع القدرة على أداء الوظائف اليومية. ويصيب المرض نحو 55 مليون شخص حول العالم، ومن المتوقع أن يرتفع هذا العدد مع تقدّم أعمار السكان عالميًا، الأمر الذي يجعله أحد أبرز التحديات الصحية العامة في الوقت الراهن. يعرض هذا البحث مقاربة معتمدة على تقنيات التعلم العميق لتشخيص مراحل الزهايمر باستخدام صور الرنين المغناطيسي. وقد أُجريت الدراسة باستخدام مجموعتي بيانات مختلفتين تحتويان على عدد كبير من صور الرنين المغناطيسي للدماغ. يقترح المنهج المقدم دمج شبكة VGG16 (Visual Geometry Group ذات 16 طبقة (لاستخراج الخصائص المكانية مع شبكة الذاكرة طويلة وقصيرة الأمد (LSTM) لنمذجة العلاقات الزمنية بين المقاطع. خضعت الصور لعمليات معالجة أولية، وتم استخراج الخصائص وفق طرق معيارية قبل إجراء عملية التصنيف. وقد جرى تنفيذ البنية المقترحة باستخدام أُطر التعلم العميق مثل TensorFlowوKeras، ثم جرى تقييم أدائها باستخدام معايير الدقة، والاسترجاع، والدقة الإيجابية، ومعامل F1. أظهرت النتائج التجريبية دقةً تجاوزت 99% على بيانات التدريب وحوالي 98% على بيانات التحقق باستخدام 8000 صورة رنين مغناطيسي لتصنيف مراحل الزهايمر. وأثبتت بنية VGG16–LSTM كفاءتها العالية في تصنيف مراحل المرض من صور الرنين المغناطيسي، محققةً نتائج دقيقة وواعدة. تمثل هذه الدراسة مقاربة واعدة لتشخيص الزهايمر ذات إمكانية تطبيقية في السياق السريري.

1. Introduction

Alzheimer's disease (AD) is one of the common types of dementia, which is mostly defined as a multi-symptomatic disorder [1]. AD can be described as severe dementia because it includes symptoms like changes in personality and behavior, memory loss, issues with cognition [2], and executive function deficits. Moreover, in most cases, patients might also exhibit symptoms of mental instability.

Until now, there is no well-known treatment for AD; however, at least temporarily, the symptoms can be reduced to make the patient feel better with close observation and medications. Mostly, AD is common among seniors, especially those who are in their fifties or older, and also a small percentage of those with Down syndrome. [3] The growing Incidence of AD has become a considerable global public health concern, creating economic and healthcare pressures on families and societies. The number of dementia cases increased from 20.3 million in 1990 to 43.8 million in 2016 and is expected to reach 152 million by 2050 [4], largely due to increased life expectancy. A range of medical, lifestyle, and social risk factors further complicates its prevention and management [5]. In 2020, researchers also identified that brain damage, air pollution [6], and heavy alcohol [7] use are considered related causes.

Many scientists continue working on AD, with vast resources spent on early diagnosis. Yet, the only reliable method is still post-mortem analysis, focusing on the urgent need for more accurate early detection. Identifying at-risk individuals early could significantly affect disease progression, but this is very difficult due to the aging population and other age-related circumstances. [8] In recent years, several technologies have shown remarkable capabilities in applying artificial intelligence to detect Alzheimer's disease in its early stages. scientists and researchers have employed techniques like transfer learning [9] in pre-trained CNNs on massive datasets before equipping them with the AD classification task. Even though unsupervised learning models have been used for such tasks, models like an autoencoder have been tested for feature selection from high-dimensional datasets. Machine learning models, as the previous research papers suggest, are capable of handling and analyzing complicated, big datasets derived from EHRs. To improve the current state of early and reliable AD diagnosis, AI technologies can provide promising methods to use massive volumes of multidimensional

data, including genetic information, cognitive testing, medical pictures, and more [10]. Early and accurate diagnosis of AD remains a major challenge faced by medical researchers worldwide. As our populations continue to age, identifying those at risk before symptoms fully emerge is critical. Several innovative studies have contributed valuable insights toward this goal by applying machine learning techniques to analyze complex medical data.

2. Related Works

Many previous studies have examined early AD diagnosis; Ramzan et al. (2020) [11] studied rs-fMRI data using deep learning for Alzheimer's identification. Their model used ResNet-18 for defining ADNI scans, also preprocessed their data, and then extracted features with the help of a CNN. Features are then fed to a pre-trained ResNet-18 architecture. The performance of the proposed models was evaluated with accuracy, precision, recall, and F1-score. The pretrained model achieved an average accuracy of 97.92%, precision of 98.13%, recall of 97.92%, and F1-score of 97.96%. The fine-tuned (FT) model obtained an accuracy of 97.88%, a precision of 98.10%, a recall of 97.89%, and an F1-score of 97.93%. The limitation of this study is the small sample size, which may change the accuracy and cause overfitting.
Turkson (2021) [12] suggested a novel approach to classify Alzheimer's Disease (AD) using structural MRI images. The method includes an unsupervised pre-training stage with Spiking Neural Networks (SNNs) to extract features and spatial patterns from MRI data. The resulting features from the SNN were then used to train a supervised CNN, aiming to refine classification performance. The training and testing of the model were performed on the well-known ADNI dataset. Researchers utilized this dataset to evaluate two different methodologies: one utilized a SNN model that had been pre-trained, and the second methodology employed an SNN model to extract features and a CNN to classify AD. The hybrid model produced an accuracy of 90%. However, this paper did not present computational complexity or runtime analysis; therefore, it limits the ability to implement this methodology in real-time clinical settings.
Zena (2022) [13] used the VGG16 architecture for AD identification and analyzed the effectiveness of VGG16 against MobileNet-V2, ResNet-101, and DenseNet-121 based on model accuracy. The VGG16 model yielded 97.625% accuracy when performing AD classification; however, like previous studies, there was no computational complexity analysis performed, which is a critical factor when considering the implementation of healthcare-based solutions.

Hazarika (2023) [14] explored the use of deep neural networks (DNNs) on MRI scans for Alzheimer's disease classification, evaluating architectures such as DenseNet-121, LeNet, and Alex Net to evaluate classification. A hybrid of LeNet and Alex Net was suggested. The new model achieved 93.58% performance, exceeding DenseNet-121's 86.55% accuracy. Also, Kadhim et al. (2024) [15] suggested a tailor-made CNN. This CNN was trained from scratch on MRI scans, developed to learn features directly from the scans for early Alzheimer's diagnosis, and the model achieved an accuracy of 96.67%. However, for a critical model such as disease diagnosis, where misclassification can have serious consequences, it is important to achieve the highest possible accuracy.

In addition to the above works, Bhadrashetty and Sandhya (2024) [16] employed CNNs with Efficient Net variants for dementia severity staging on the OASIS dataset. By applying an ordinal classification strategy, their model better captured disease progression across four stages (none, very mild, mild, and moderate), achieving around 98% accuracy. However, the study relied solely on OASIS data, excluded severe cases, and did not assess computational

efficiency or external validation, which limits its generalizability to real-world clinical deployment.

A growing corpus of literature in 2025 is dedicated to the investigation of Alzheimer's Disease (AD), utilizing a variety of methodological approaches. Aghajanian and colleagues (2025) [17] investigated the detection of AD progression from mild cognitive Deficit to Mild Cognitive Impairment (MCI) using longitudinal structural MRI. They constructed a hybrid deep learning approach that combined a 3D ResNet for feature extraction from single MRIs with a time-aware Long Short-Term Memory (T-LSTM) network to model sequences of images. They also extracted and assessed radiomics features. Their model was trained and tested on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The results demonstrated that the longitudinal deep learning model performed well in predicting instant MCI-to-AD progression (c-index ~ 0.90) with significantly better performance than single-time-point models. A limitation of this work is the small sample size, which limited the use of more complicated architectures and prospect for external validation.

Elmotelb et al. (2025) [18] focused on optimizing the reproducibility of multi-classification and staging in AD through a novel deep learning approach, which included rigorous hyperparameter tuning. By systematically tuning the model's architecture, the study suggests training the model on public, augmented Alzheimer's MRI data, thus presenting classification accuracies of over 98% for the various disease stages. The main contribution is that the authors establish the important role of thorough hyperparameter tuning in obtaining reliable disease staging; however, the use of an augmented, publicly sourced dataset indicates a caveat with the model's generalizability to raw, uncurated data from disparate clinical frameworks.

Lu et al. (2025) [19] designed and tested an interpretable, MRI-based deep learning model to stratify risk for AD and risk of progression. The study utilized a model originally trained on North American cohorts and leveraged it to the independent Chinese SILCODE cohort without finetuning for the purpose of testing cross-ethnic generalizability. The authors tested this model, using structural T1-weighted MRI images to create individual patient AD risk scores and brain risk maps. The model performed well in distinguishing AD from normal cognition (NC), achieving an AUC of 91.3% at a sensitivity of 95.2%. Importantly, the model incorporated prognostic utility, showing detection of 86.7% of individuals who went on to established AD at baseline and identified abnormal risk scores up to 14 years before clinical detection of the disease. The risk numbers developed from structural MRI scans were also found to be associated with key plasma biomarkers, including p-tau 217 and NfL, and cognitive test performance. The authors used interpretability modalities to reveal that there were three different AD brain subtypes, and one subtype was accumulating significantly faster in cognitive decline in patients with Mild Cognitive Impairment (MCI) over the course of the study. A significant limitation is that the authors validated the model in a single site, a Chinese cohort, which may imply limited generality despite very good performance results from the model in the original testing cohort.

Lastly, in 2025, Komal et al. [20] put forth a hybrid deep learning architecture based on classifying AD using MRI scans. This proposed framework combined CNNs with self-attention to optimize not only accuracy, but interpretability as well. The authors established two pipelines for analysis based on two different datasets: the OASIS-1 study used a 2D DenseNet-121, which integrated the Transformer encoder, while the OASIS-2 study used a 3D DenseNet along with self-attention blocks. This model achieved a 91.67% accuracy on the OASIS-1 study data and 97.33% on the OASIS-2 study, resulting in strong precision and F1-scores. Attention heatmaps did improve the interpretability of the model, highlighting important brain regions that are clinically relevant, such as the hippocampus and temporal

lobe. A shortcoming of the model was the class imbalance of the datasets, especially for those individuals in advanced AD stages, which may limit generalizability. Additionally, the inclusion of a 3D self-attention module needs more computational resources, which could limit the applicability of the model in clinical settings running in real-time assessments.

Almost all the above studies showed good and promising results in general, using different machine learning and deep learning models to detect and classify AD. Limitations remained in refining the models to the constraints and needs of real-world healthcare deployment. Our study, on the other hand, tried to overcome these limitations by proposing an extensive machine learning approach constructed specifically with practical usability, efficiency, and scalability in mind, trained and then tested on multiple datasets. This makes it more suitable and more accurate compared with real-world cases. This study adopts a powerful feature extraction capability of pre-trained convolutional neural networks, like VGG16, followed by the temporal modelling abilities of recurrent layers such as LSTM. This allows our proposed model to learn powerful visual representations from 3D brain images and temporal correlations between sequences. By combining state-of-the-art computer vision mechanisms along with NLP models within a single architecture, the main aim of this process is to generate a model in a more comprehensive manner that can identify the spatial and temporal signatures characteristic of early AD progression. Moreover, fine-tuning has been performed on the model and the data parameters carefully to ensure inference latency appropriate for implementation on embedded devices that are used more often in healthcare systems.

3. Deep Learning Techniques

3.1 Very Deep Convolutional Networks (VGG)

VGG has become, in recent years, one of the most widely recognized computer vision and deep learning models. Researchers used it for observing image data problems in general, especially large-scale image data recognition tasks. An improved version of it was developed by the Visual Geometry Group at Oxford University, named VGG16. It has undergone extensive evaluation by the scientific society. VGG16 utilizes a thoroughly constructed structure containing sixteen convolutional layers, with localized 3x3 filters applied throughout the network [21]. Relying on the modest filter sizes in its layers, VGG16 is able to reach high results and maintain computational efficiency compared with other approaches [22]. The structural design of VGG16, integrated with increasing the network depth to 16 layers, granted it a better accuracy in object classification tasks, outperforming significantly all prior published results on ImageNet. The '16' signifies the number of layers that learn features from input data, and these are max-pooling, activation functions, and convolutional layers. What makes the VGG-16 model good at a lot of image recognition tasks is its deeper architecture and its ability to pick up on complex patterns from data. The VGG16 model played an essential role in the success of the research team back in 2014 [23], when they came first in the ImageNet Challenge classification and localization. They used VGG16 to learn the best possible representations, which demonstrated state-of-the-art performance in different computer vision tasks and showed huge generalization capability outside the scope of the original ImageNet dataset. This strong design is what led the study group to see its potential and open-source VGG16, allowing for other studies with deep visual representations in alternative contexts [24]. Figure 1 shows a standard VGG16 [25].

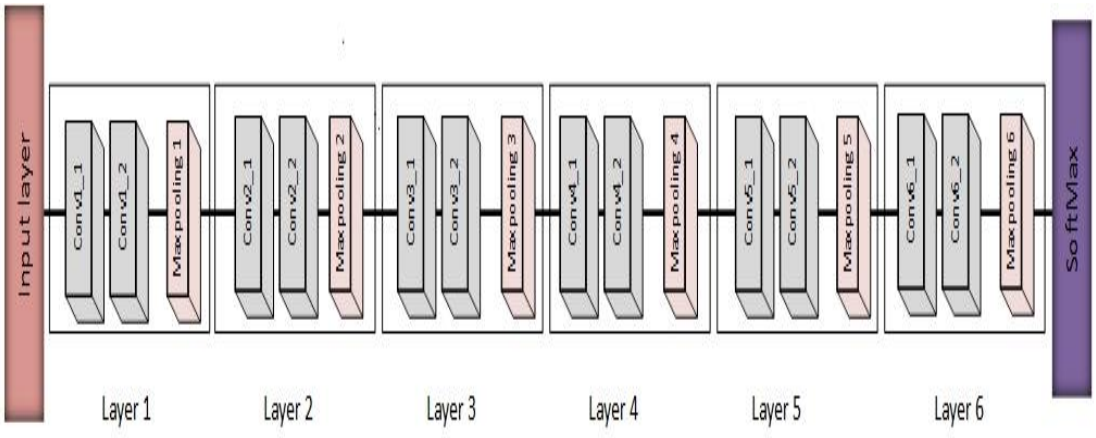


Figure 1: standard VGG-16 architecture [27]

3.2 Recurrent Neural Network (RNN)

RNNs are another form of neural networks developed specifically for sequential data. Recurrent Neural Networks (RNNs) excel at dealing with sequential data types, such as audio, text, and video.[26] The fundamental advantage of RNNs lies in how they utilize looping connections among neuronal units to remember (i.e., they maintain memory of) past inputs, allowing them to process sequentially within the context of the entire input. Therefore, RNNs enable an uninterrupted flow of information from one moment to the next. In the early years of using RNNs, propagation of learning was impeded since back-propagation was limited to a single time step (i.e., back-propagation could not happen across multiple steps). In fact, once researchers established methods to use multiple steps for gradient propagation (as error flows through real networks), this greatly expanded the ability to teach RNNs. Also, until very recently, the ability for RNNs to learn long sequences or long-term relationships was extremely limited. Recently, during the development of new types of RNNs, such as LSTM, researchers solved this issue by providing gate mechanisms for selecting which past state to save or remove. RNNs have also been augmented with new methods to focus on specific input items (e.g., attention and pointer mechanisms), enabling RNNs to produce more intelligent output in terms of sequence order. Overall, RNNs possess a great deal of power due to their ability to remember and build upon past input patterns. The continual progress of RNN technology indicates that RNNs will continue to improve at understanding the complexity of time-dependent input patterns, similar to how we understand our own experience of time [27]

3.3 Long Short-Term Memory (LSTM)

Long short-term memory, also known as LSTM for short, has become the go-to approach for modelling patterns over time in medical data [28]. Early recurrent networks struggled with vanishing signals when exploring back very far, yet LSTMs changed this through clever adaptive "gates" inside the memory cells. These cells can effectively preserve information for as long as they are needed, constantly updating their understanding following earlier results that were already known. Specialized input, forget, and output gates provide LSTMs with pinpoint control over saving critical cues versus discarding unwanted noise. That is the reason why it can connect dots between all symptoms from months ago to get a current diagnosis. This gated architecture is what lets LSTMs excel at tasks requiring cross-temporal context, like how a patient's history or disease has developed, rather than reacting to the present only. Such contextual awareness proves vital in healthcare, where the past so strongly forms the present. In reality, LSTMs have revolutionized how we analyze medical sequences, from

longitudinal biomarker tracking to lesion development over images [29]. By preserving key prior outcomes, LSTM models learn far more expressive representations of underlying biology. This leads to much better predictive results for complex conditions with both short- and long-term factors. Finally, LSTMs' storage and reasoning capabilities across a wide period of time are what make them essential for clinical predictive issues [30].

Data used in this study was obtained from two separate image datasets on Kaggle, the first one for training and the other for validating our model. The training dataset, which is authorized by the Open Data Commons Open Database License, contains 6400 MRI images gathered manually from various online sources, with each label verified and divided into 4 classes [31]. The other dataset, which was used for validation, contains 137,000 MRI images [32]; only 8,000 were used to test the model's performance.

4. Proposed Model

This study presents a combined deep learning model that integrates VGG16 for spatial feature extraction with LSTM to capture temporal dependencies across MRI slices, optimized for early-stage AD diagnosis. While based on existing models, the proposed model is tailored to enhance accuracy and temporal feature learning. A multi-case study design was used to design and test the model.

4.1 Data preparation

The MRI scans sourced from the Augmented Alzheimer's Dataset and the Alzheimer's MRI 4-class Dataset were used for training and evaluation. Images were thoroughly preprocessed using advanced preprocessing methods. This process included data augmentation techniques like rotation, scaling, flipping, noise reduction, as well as color changes on datasets from Kaggle, then added normalization and scaling to increase the model performance and eventually gain highly informative data diversity. Every image was subjected to normalization using the pixel value 1/255, rescaling values of each pixel to the range of [0,1] to standardize the data input distribution. Preprocessing and augmentation techniques were especially favorable to the underrepresented class 'Moderate Demented' by helping with imbalanced classes and having augmentation factors of 3x to 101x among the various classes. Each train, validation, and test split contained 70% training, 20% validation, and 10% testing from each respective dataset through structural sampling to control the number of classes. The preprocessing steps increased the data variation through the images and the scope of the dataset for more successful model training. Table 1 provides an overview of the data before and after preprocessing, while the examples of the sample images are depicted in each respective class after the preprocessing.

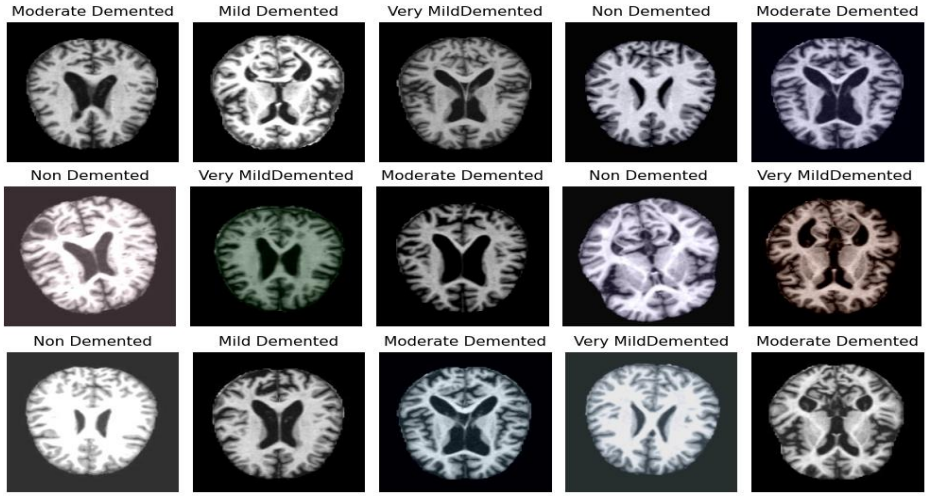


Figure 2: Sample of data after preprocessing procedure.

Table 1: Data size before and after the preprocessing procedure

Class	images before preprocessing	images after preprocessing	Augmentation Factor
Mild Demented	896	8,960	10x
Moderate Demented	64	6,464	101x
Non-Demented	3,200	9,600	3x
Very Mild Demented	2,240	8,960	4x
Total	6,400	33,984	5.3x

4.2 VGG16 Feature Extraction and Spatial Processing

The model proposed in this study was implemented with the Keras API from the TensorFlow deep learning framework. It uses pretrained VGG16 for extracting features from image data. The VGG16 backbone was started with ImageNet pretrained weights, including 13 convolutional layers with 3 by 3 kernels and ReLU activation function, succeeded by 5 max pooling layers with 2 by 2 windows. The model processes 3D color images as input with 224*224*3 pixels as dimensions for each image. These input images were then fed to a pretrained VGG16 with the top layers removed and convolutional blocks retained.

Feature extraction was performed by removing the original fully connected layers (FC6, FC7, FC8) and attaching a Flatten layer, transforming the 7×7×512 feature maps into 25,088 dimensional vectors. To gradually decrease spatial dimensions while extracting the features, the VGG16 model integrates max pooling layers between convolutional layers. In order to guarantee the features, start at the optimal point; these convolutional blocks are initialized with pretrained weights. This allows it to use the VGG16's visual representations and update gradients with specific features of the domain. A progressive fine-tuning approach was implemented: initially freezing all convolutional blocks, then unfreezing the final 6 layers (conv5_1 to conv5_3) to allow domain-specific adaptation while preventing overfitting. All layers are frozen at the beginning, except for the last few convolutional blocks of the model. The output of VGG16 is then flattened into a 25088-dimensional dense feature vector by a Keras Flatten layer. These vectors will help to protect the spatial information extracted by the convolutional network. The VGG16 model generates feature matrices and label arrays.

4.3 Temporal Processing, Classification, and Evaluation

An LSTM layer from Keras was used after the feature extraction was complete and the array of labels was generated to predict the temporal dependency in the sequence of image features. LSTM is highly compatible with our model since it can learn long-term dependencies through the internal cell state. Additionally, processing of the outputs from LSTM includes dense and dropout layers before classification with a SoftMax output layer, whose units are equal to the number of classes. Validation loss and accuracy metrics for model performance observation were also included, along with some callback functions like Early Stopping and Reduce LR On Plateau, which will help in preventing overfitting problems. Figure 3 shows our proposed model. The training and validation accuracy have been recorded within the final evaluation on those splits, with results over 99.5% for training and 98.58% for validation on the second dataset, showing that the model learned and performed well.

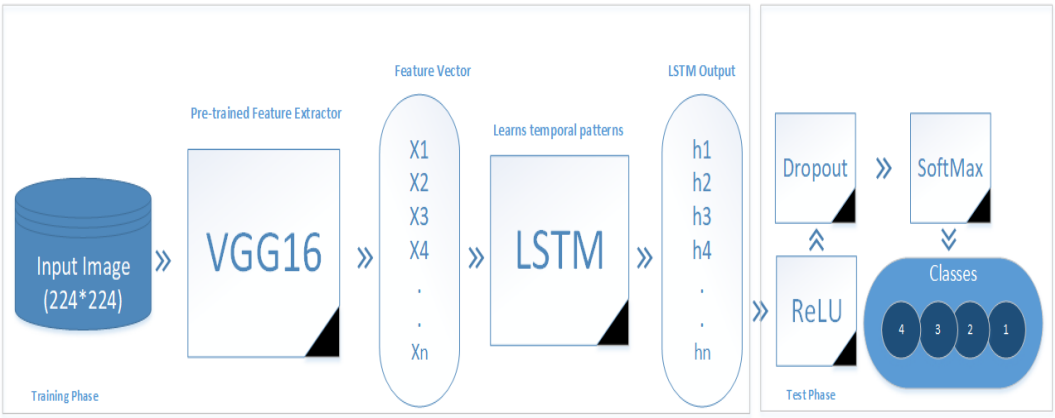


Figure 3: Proposed VGG16-LSTM Model.

5. Results and Discussion

As mentioned before, our proposed VGG16-LSTM model achieved a test accuracy of 98.58%, with strong precision, recall, and F1-scores, demonstrating its effectiveness in classifying Alzheimer's stages from 2D MRI images. A feature extraction function was defined in the feature extraction model, which uses the data generated from input images and loops it through the batches, and then passes it to the VGG16 model up to the flatten layers, then gets the labels and stacks them into a list of features. The output needs to fit the LSTM model, so it has to be reshaped.

The LSTM model's architecture contains layers with 128 units, a dropout layer with a rate of 0.6, and two dense layers with 64 units, as well as the number of unique labels, correspondingly. The model has been constructed using the Adam optimizer and sparse categorical cross-entropy as a loss function. It was trained through 25 epochs, which was enough to make the model stable and give the desired results with a 32 32-batch size. Also, callbacks are formulated to motivate early halting and decrease the learning rate. VGG16 by itself gave a rate of 96.5% accuracy for the classification of AD, as depicted in Figure 4.
Moreover, our enhanced proposed combined model (VGG16 + LSTM) outperformed recent state-of-the-art models in AD stage diagnosis. The proposed model scored an overall accuracy of 98.58%, surpassing other published methods. Unlike models that rely on multiple data modalities or heavy preprocessing, our approach utilizes a single imaging modality with minimal preprocessing yet delivers good results—highlighting the strength of combining spatial and temporal feature extraction.

To evaluate the classification performance in more detail, Figure 4 presents the training curves (loss and accuracy) for the standalone VGG16 model, while Figure 5 illustrates the performance of the proposed VGG16-LSTM model. To further evaluate classification performance, confusion matrices were created for both models (Figure 6 for our suggested model and Figure 7 for the VGG16 by itself). Both models showed a strong diagonal pattern, reflecting accurate predictions through all classes. However, significant differences emerge when comparing specific misclassifications. VGG16 (Figure 7): This model misclassified 207 "non-demented" cases as "very mild demented," showing difficulty in distinguishing early-stage patients. Our suggested model, VGG16-LSTM (Figure 6): The misclassification proportion for the same category dropped pronouncedly to 45 cases only, showing the model's enhanced sensitivity for early detection. This reduction in early-stage misclassification is clinically important, as accurate differentiation of "very mild" cases can enable earlier therapeutic action. Overall, the confusion matrix confirms that the proposed model provides more accurate diagnostic precision and reliability compared to the baseline

VGG16. The proposed model matrix thereby demonstrates greater clinical utility and diagnostic precision, especially for nuanced and critical early-stage differentiation.
As for evaluation metrics, the proposed model achieves 98% on F1-score, recall, and precision, indicating balanced and consistent performance across all classes. Moreover, the model took only 175 seconds to train for 25 epochs, consuming about 3298 MB of GPU memory. This highlights not only its high accuracy and robustness but also its efficiency and scalability in practical healthcare systems. Accuracy results from some of the prior research compared to the proposed framework are listed in Table 2.

Table 2: A comparison between the previous results and our model

Model	Proposed Model	Model [5]	Model [6]	Model [7]	Model [8]	*Model [9]*
Accuracy	98.58	97.88%	90%	97.62%	93.58%	96.67

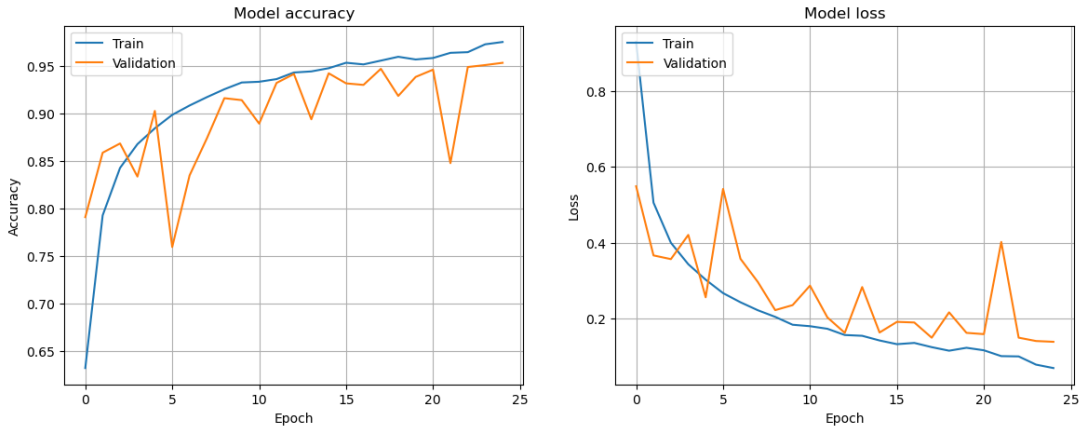


Figure 4: VGG16 Accuracy and Loss for AD Diagnosis.

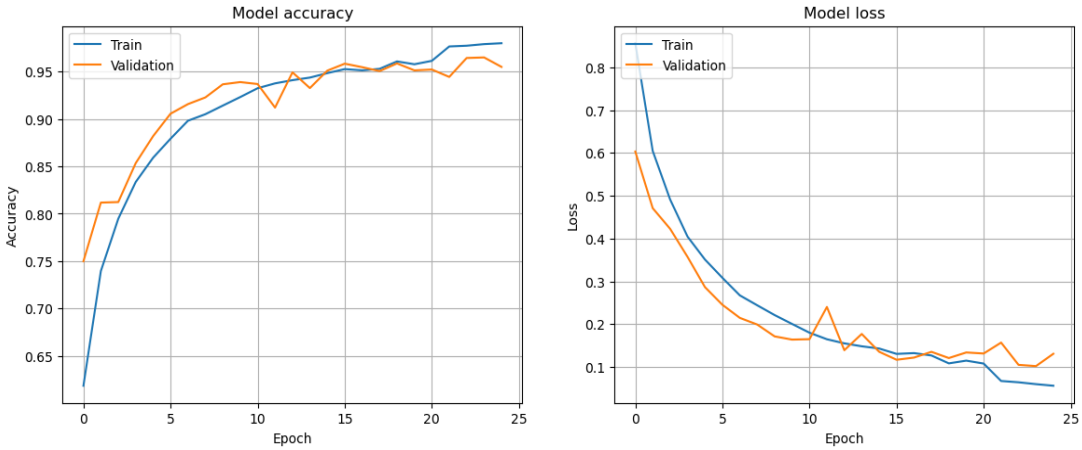


Figure 5 :Proposed Model Accuracy and Loss.

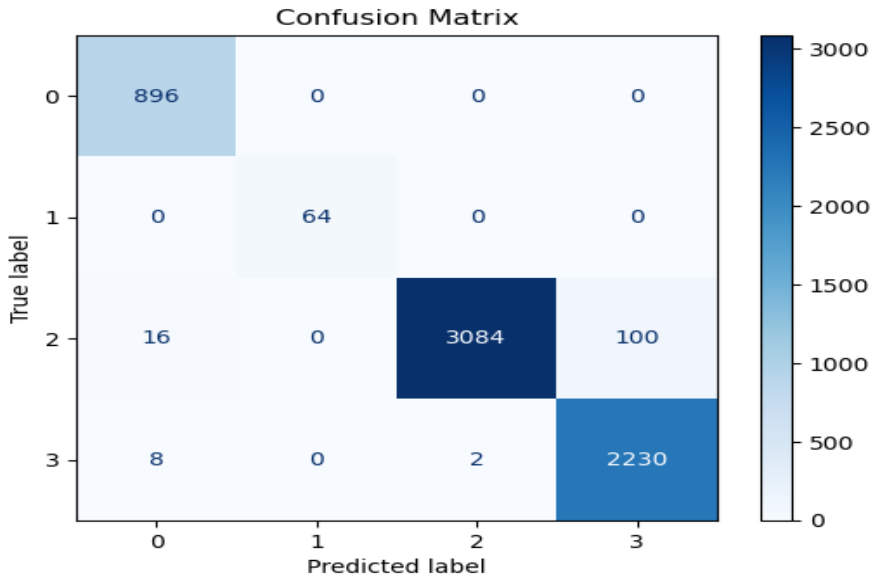


Figure 6: Proposed Model Confusion Matrix.

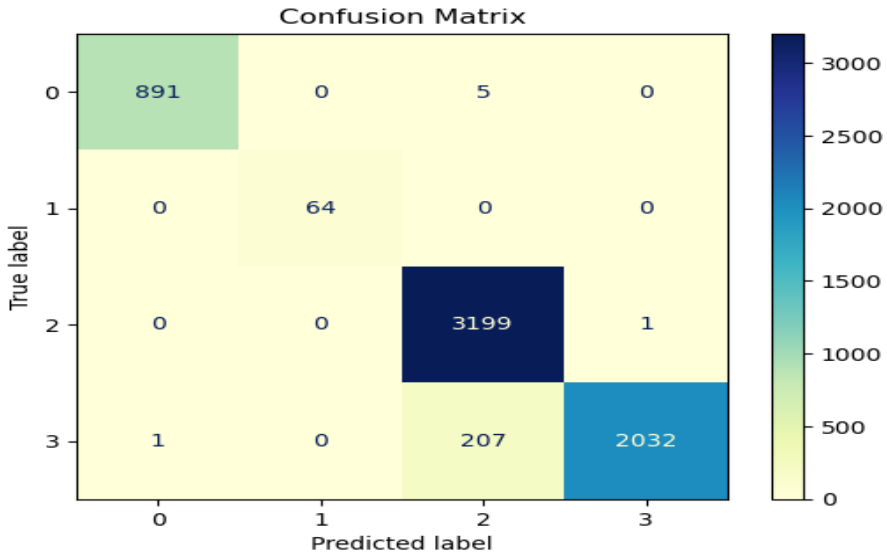


Figure 7: VGG16 Confusion Matrix.

6. Conclusions

In conclusion, our study introduced a combined deep learning model integrating the VGG16 architecture for spatial feature extraction with an LSTM layer to capture sequential dependencies from 2D MRI scans. The model was trained and tested on a publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, using a multi-case design. Images were pre-processed through augmentation, normalization, and scaling, then split into 80% for training and 20% for testing. Experimental results showed that the suggested model achieved a good classification performance for AD stages, scoring 98.58% accuracy, outperforming other existing models. Evaluation metrics such as precision, recall, and F1-score verified the model's sturdiness, and visualization tools like confusion matrices and ROC curves additionally demonstrated consistent performance across the classes.

7. Future Works

As the results show, our suggested model achieved high accuracy; nevertheless, several points can be improved in the upcoming future works, such as:

- Testing the model on larger and more diverse datasets to improve generalization.
- Applying other types of network architectures to increase feature learning.
- Integrating heuristic or ensemble methods to enhance classification accuracy.
- Evaluating the model on more imaging modalities, such as CT scans or PET scans, to reinforce multimodal diagnosis.
- For external validation, the model can be tested on independent external datasets, particularly on data collected directly from hospitals and private clinics, to ensure reliability and generalizability across populations.
- Apply explainable AI techniques like Grad-CAM, saliency maps, or SHAP to recognize critical brain regions affecting the classification process, thereby improving clinician trust and usability.

These directions aim to further strengthen early AD diagnosis through automated, efficient, and reliable deep learning techniques.

References

[1] S. Safiri, A. Ghaffari Jolfayi, A. Fazlollahi, S. Morsali, A. Sarkesh, A. Daei Sorkhabi, B. Golabi, R. Aletaha, K. Motlagh Asghari, R. Hamidi, S. E. Mousavi, S. Jamalkhani, N. Karamzad, A. Shamekh, M. Mohammadinasab, M. J. M. Sullman, F. Sahin, and A.-A. Kolahi, "Alzheimer's disease: a comprehensive review of epidemiology, risk factors, symptoms diagnosis, management, caregiving, advanced treatments and associated challenges," Frontiers in Medicine, vol. 11, 2024 , doi: 10.3389/fmed.2024.1474043

[2] X.-L. Li, N. Hu, M.-S. Tan, J.-T. Yu, and L. Tan, "Behavioral and Psychological Symptoms in Alzheimer's Disease," BioMed Research International, vol. 2014, Article ID 927804, pp. 1–9, Jul. 2014, doi: 10.1155/2014/927804

[3] X. Li, X. Feng, X. Sun, N. Hou, F. Han, and Y. Liu, "Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2019," Frontiers in Aging Neuroscience, vol. 14, Oct. 2022, doi: 10.3389/fnagi.2022.937486.

[4] Wisam H. Hoidy, Shaimaa M. Essa, and Mohammed H. Al-Saadi, "Pro-inflammatory/Anti-Inflammatory Interleukin Imbalance in Iraqi Alzheimer's Patients: Implications for Post-Conflict Environmental Exposures," European Journal of Neuroscience, vol. 62, e70214, Aug. 2025, doi: 10.1111/ejn.70214

[5] J. Hardy, "Alzheimer's disease: The amyloid cascade hypothesis: An update and reappraisal," Journal of Alzheimer S Disease, vol. 9, no. s3, pp. 151–153, Jul. 2006, doi: 10.3233/jad-2006-9s317.

[6] A. Peters, "Ambient air pollution and Alzheimer's disease: the role of the composition of fine particles," Proceedings of the National Academy of Sciences (PNAS), vol. 120, no. 3, e2220028120, 2023, doi: 10.1073/pnas.2220028120

[7] D. Rodríguez-Arroyo, C. Romano, J. Pérez-González, C. Valdés, and A. Martínez-Cruz, "Alcohol and stress exposure across the lifespan are key risk factors for Alzheimer's disease and cognitive decline," Cell Reports Medicine, vol. 5, no. 3, Mar. 2024, doi: 10.1016/j.xcrm.2024.101329.

[8] I. Aberathne, D. Kulasiri, and S. Samarasinghe, "Detection of Alzheimer's disease onset using MRI and PET neuroimaging: longitudinal data analysis and machine learning," Neural Regen. Res., vol. 18, no. 10, pp. 2134–2140, Oct. 2023, doi: 10.4103/1673-5374.367840.

[9] G. I. Stoleru and A. Iftene, "Transfer Learning for Alzheimer's Disease Diagnosis from MRI Slices: A Comparative Study of Deep Learning Models," Procedia Computer Science, vol. 225, pp. 2614–2623, 2023, doi: 10.1016/j.procs.2023.10.253

[10] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 7, pp. 8459–8486, Jan. 2022, doi: 10.1007/s12652-021-03612-z.

[11] F. Ramzan et al., "A deep learning approach for automated diagnosis and Multi-Class classification

of Alzheimer's disease stages using Resting-State FMRI and residual neural networks," Journal of Medical Systems, vol. 44, no. 2, Dec. 2019, doi: 10.1007/s10916-019-1475-2.

[12] R. E. Turkson, H. Qu, C. B. Mawuli, and M. J. Eghan, "Classification of Alzheimer's disease using deep convolutional spiking neural network," Neural Processing Letters, vol. 53, no. 4, pp. 2649–2663, Apr. 2021, doi: 10.1007/s11063-021-10514-w.

[13] J. I. Zena, E. Lucky, C. G. Ellaine, I. S. Edbert, and D. Suhartono, "Deep Learning Approach based Classification of Alzheimer's Disease Using Brain MRI," 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 397–402, Dec. 2022, doi: 10.1109/isriti56927.2022.10053030.

[14] R. A. Hazarika et al., "An approach for classification of Alzheimer's disease using deep neural network and brain magnetic resonance imaging (MRI)," Electronics, vol. 12, no. 3, p. 676, Jan. 2023, doi: 10.3390/electronics12030676.

[15] K. A. Kadhim, F. Mohamed, F. H. Najjar, and G. A. Salman, "Early diagnose Alzheimer's disease by convolution neural network-based histogram features extracting and canny edge," Baghdad Science Journal, vol. 21, no. 2(SI), p. 0643, Feb. 2024, doi: 10.21123/bsj.2024.9740.

[16] M. Z. Hussain, et al., "A fine-tuned convolutional neural network model for Alzheimer's diagnosis using MRI," Scientific Reports, vol. 15, Article No. 1234, Apr. 2025, doi:10.1038/s41598-025-86635-2.

[17] S. Aghajanian, F. Mohammadifard, I. Mohammadi, S. R. Firouzabadi, A. B. Bagheri, E. M. Ghaffary, and O. Mirmosayyeb, "Longitudinal structural MRI-based deep learning and radiomics features for predicting Alzheimer's disease progression," Alzheimer's Research & Therapy, vol. 17, no. 182, Jul. 2025, doi: 10.1186/s13195-025-01827-2.

[18] A. S. Elmotelb, F. F. Sherif, A. S. Abohamama, M. Fakhr, and A. M. Abdelatif, "A novel deep learning technique for multi classify Alzheimer disease: hyperparameter optimization technique," Frontiers in Artificial Intelligence, vol. 8, Art. ID 1558725, Apr. 2025, doi: 10.3389/frai.2025.1558725.

[19] B. Lu, Y. Chen, R. Li, M. Zhang, S. Yan, G. Chen, F. X. Castellanos, P. M. Thompson, J. Lu, Y. Han, and C. Yan, "Interpretable MRI-Based Deep Learning for Alzheimer's Risk and Progression," medRxiv preprint, May 2025, doi: 10.1101/2025.05.06.25326606.

[20] R. Komal, P. Dhavakumar, K. Rahul, B. Jaswantha, and R. Preeth, "Hybrid deep learning framework for magnetic resonance imaging-based classification of Alzheimer's disease," Brain Network Disorders, May 2025, doi: 10.1016/j.bnd.2025.05.002.

[21] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," Expert Systems with Applications, vol. 165, 2020, doi:10.1016/j.eswa.2020.113909.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for Large-Scale image recognition," Computer Vision and Pattern Recognition, Sep. 2014, [Online]. Available: http://export.arxiv.org/pdf/1409.1556

[23] Chetan Sharma, Shamneesh Sharma, Sakshi, and Hsin-Yuan Chen, "Advancements in handwritten Devanagari character recognition: A study on transfer learning and VGG16 algorithm," Discover Applied Sciences, vol. 6, no. 623, pp. 1–20, 2024, doi: 10.1007/s42452-024-06217-1

[24] T. A. T. K. Azaharan, A. K. Mahamad, S. Saon, N. Muladi, and S. W. Mudjanarko, "Investigation of VGG-16, ReSNet-50 and AlexNet performance for brain tumor detection," International Journal of Online and Biomedical Engineering (iJOE), vol. 19, no. 08, pp. 97–109, Jun. 2023, doi: 10.3991/ijoe.v19i08.38619.

[25] L. Yang, S. Xu, X. Yu, H. Long, H. Zhang, and Y. Zhu, "A new model based on improved VGG16 for corn weed identification," Frontiers in Plant Science, vol. 14, Jul. 2023, doi: 10.3389/fpls.2023.1205151.

[26] A. Murad, "Deep Recurrent Neural Networks for Human Activity Recognition," Sensors, vol. 17, no. 11, Nov. 2017, pp. 2556, doi:10.3390/s17112556.

[27] R. Schmidt, "Recurrent Neural Networks (RNNs): A gentle Introduction and Overview," arXiv, 23-

Nov-2019. [Online]. Available: https://doi.org/10.48550/arXiv.1912.05911.

[28] F. M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed, "A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU," Journal on Artificial Intelligence, vol. 6, no. 1, pp. 301–360, 2024, doi:10.32604/jai.2024.054314.

[29] R. DiPietro and G. D. Hager, "Deep learning: RNNs and LSTM," in Elsevier eBooks, 2019, pp. 503–519. doi: 10.1016/b978-0-12-816176-0.00026-0.

[30] L. Boongasame, J. Boonpluk, S. Soponmanee, J. Muangprathub, and K. Thammarak, "Design and implement deepfake video detection using VGG-16 and Long Short-Term Memory," Applied Computational Intelligence and Soft Computing, vol. 2024, no. 1, Jan. 2024, doi: 10.1155/2024/8729440.

[31] "Alzheimer's Dataset (4 Class of Images)," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images.

[32] "Medical Scan Classification Dataset | Kaggle," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/arjunbasandrai/medical-scan-classification-dataset.