iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81481

Alzheimer's Disease Early Diagnosis Using Multimodal Imaging

M. Tharani Sri[1], K. Sujitha[2], Ch. Tejaswini[3], N.Sujith[4], Y. Suseel Kumar[5]
Department of Artificial Intelligence and Machine learning, [Acharya Nagarjuna University],[Guntur, Andhra Pradesh, India

Abstract: To improve patient outcomes over time, it is essential to diagnose Alzheimer's Disease (AD) in a timely manner so that timely interventions can be provided. This manuscript proposes a multimodal imaging framework based on deep learning, which can detect AD at an earlier stage using Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Positron Emission Tomography (PET). Distinct Convolutional Neural Network (CNN) features are extracted from MRI and CT (used together) and a transformer model processes the PET image. After the extraction of multimodal representations of features, features from the multimodal imaging are fused and classified using a fully connected neural network. When evaluated using several modalities, results showed that our multi-fusion model significantly outperformed each single modality model providing improved accuracy and improved robustness. Our integrated multimodal imaging framework demonstrates that it has the potential to enhance the diagnostic capability for the clinician to achieve an earlier diagnosis and provide for better management of AD.

Keywords: Early Diagnosis of Alzheimer's, Multimodal Imaging (MRI/PET/CT) Deep Learning, Convolutional Neural Network (CNNs), Transformer Networks, Data Integration & Fusion, Medical imaging, Hippocampal atrophy, Computerized Diagnosis, AD Classification.

I. INTRODUCTION

Alzheimer's disease is a chronic as well as an irreversible neurodegenerative illness that principally impacts the capacity to retrieve memories, the ability to think, and the ability to carry out activities of daily living. Alzheimer's disease is one of the most frequent causes of dementia in the elderly and will continue to remain a substantial health concern globally. The typical course of the disease starts with relatively mild memory loss and progresses to severe cognitive impairment. As the disease continues to progress, the patient will have increasing difficulty with language, reasoning, making decisions, and living independently. [1].

One of the principal challenges in accurately diagnosing Alzheimer's disease is that the early signs and symptoms of the disease can be very subtle and easily attributed to normal aging. The majority of patients are diagnosed with Alzheimer's after a great deal of irreversible damage has occurred to brain cells. A late diagnosis of the disease reduces the efficacy of current treatment options and substantially hinders the ability to slow the progression of the disease. Therefore, the early and accurate diagnosis of Alzheimer's disease is one of the most critical components of modern healthcare today. [2]–[4].

Common diagnostic techniques used when diagnosing Alzheimer's disease include clinical observations, neurological examinations, and imaging techniques (MRI, CT and PET). Each technique provides some information when examined alone, but may be limited in the information provided. For example, although MRI is excellent at detecting structural brain changes due to Alzheimer's disease (e.g., brain atrophy), MRI by itself does not provide a complete assessment of an individual's decline in function over time. CT scans also provide structural information about the anatomy of the brain and how dense (or hard) it is, but CT imaging is not as sensitive as MRI when looking for early signs of an abnormality. Conversely, although PET scans provide metabolic activity of the brain and are able to reveal whether an individual is functionally impaired, they tend to be very expensive and are not normally performed as a regular part of routine clinical practice. The incomplete assessments of the different imaging modalities has motivated many researchers to pursue the use of multimodal imaging. By combining the data from each imaging modality (structural, functional and anatomical), researchers can use the results from multiple scans together to develop a more comprehensive understanding of Alzheimer's disease. The use of multimodal analyses can improve the accuracy of diagnosis because it enables clinicians to integrate the information they obtain from structural, functional and anatomical imaging into one decision-making framework. [5], [6], [10]. With the advent of artificial intelligence, new opportunities have arisen for utilizing machine-based techniques for the analysis of medical images. As one of the domains benefiting from developments in this area is that of deep learning (the use of algorithms that can learn patterns from large amounts of data, such as medical images created by machines or computers) - machines are able to learn features from a large amount of scanned data without requiring painstaking feature design or having to be personally programmed to recognize patterns through the use of these algorithms.

Some of the neural architectures that are being utilized for the analysis of medical images include convolutional neural networks, which are capable of detecting local image features; ResNet18, which extends the capabilities of deep feature learning using residual connections to preserve information over many layers; and vision transformers, which introduce an additional level of intelligence for capturing long-range dependency and global relationships among pixels of the medical image.

By integrating multimodal imaging with deep learning, the potential exists for improving the accuracy of diagnosing Alzheimer's disease, as the computer model will have available more complete and complex pattern information than traditional methods. A computer model's ability to perform single or multiple MRI, CT and PET scans in a combined form is advantageous, especially when it comes to identifying both structural and functional changes associated with an early form of Alzheimer's disease. By utilizing an intelligent computer model to analyze these medical images and support clinicians' decision making, the potential exists for enhancing the chances of timely intervention. For this project, a comprehensive multimodal artificial intelligence (AI) system will be developed to allow for the early detection of Alzheimer’s disease. This system will not only have an algorithm developed for model training, but it will also incorporate a command-line interface, a desktop graphical interface, an internet web application, and a database layer for a total of five different types of user interfaces. These features make the system more usable and easier to deploy in real-life conditions. The system will provide an output of whether a patient is classified as having dementia or not with an accompanying confidence score for the final classification of that patient.

The use of majority voting among three different models (a convolutional neural network (CNN), a residual neural network (ResNet18), and a vision transformer) greatly increases the reliability of the classifiers. When more than one model votes for the same classification category, the reliability of that final prediction increases. The use of an ensemble approach/functions creates less chance of error from an individual classifier, thus improving the overall decision. This is extremely important in the healthcare field as accuracy and reliability are vitally important. [7]–[9], [11], [12].

The overall project represents a fused multi-modal deep learning approach with new-age industrial software engineering practices to develop a complete diagnostic aid/system. To help address the limitations of a single modality diagnostic approach, the creation of this system provides a tangible means toward achieving earlier and more reliable screening for Alzheimer's; therefore it can be viewed as both a research construct/prototype, as well as a launch platform for future AI applications within healthcare.

II. RELATED WORK

Traditionally, Alzheimer's diagnosis was done solely through physical examination, but now there are advanced systems for deep-learning methods of diagnosis that aren't restricted to physical assessments. Previous work on the diagnosis of Alzheimer's involved solely the usage of MRIs and conventional feature extraction methodologies. Zhang et al's initial analysis into this area involved using MRIs alone to identify anomalies between people with and without Alzheimer's by analysing the presence of hippocampal atrophy. These methods have their limitations, since only analysing one imaging method limits what can be measured and, therefore, how representative the results are as a whole for someone diagnosed with Alzheimer's. [1]. Li et al. also examined the use of convolutional neural networks to classify MRI scans. They showed that deep learning could automatically discover patterns related to illness without requiring any hand-crafted feature extraction [2]. Later on, Wang et al. built on this concept by using deeper residual networks and demonstrated that the use of a ResNet-like architecture resulted in more reliable classification of Alzheimer’s Disease compared to using prior methods [3]. As the field of MRI has continued to make advancements and researchers have continued to expand their interest in developing deeper or more powerful methods of machine learning to perform medical imaging analysis, Kumar et al. showed that automatic extraction of features from brain scan images can outperform traditional methods for finding local patterns4].

Additionally, through the work of He et al., residual learning was introduced into the field of medical imaging, which has become a dominant approach in the classification of medical imagines and allows for deeper learning networks to be trained more efficiently than using several different techniques [5]. Most recently, through the work of Dosovitskiy et al., the introduction of the Vision Transformer has provided a new framework for performing Alzheimer's Disease research because global attention allows for accurate representation of spatial relationships between all of the pixels on a full MRI scan of the brain. Thus, these earlier and more current studies demonstrate that recent advancements in the fields of machine learning and medical images can have a tremendous impact on improving our understanding of medical imagines. [6].

In addition to single image evaluation and analysis, many scientists have also worked on combining different types or methods of imaging to increase the quality of diagnosis. Ortiz et al. and Silva et al. both evaluated the use of magnetic resonance imaging (MRI) and positron emission tomography (PET) for classification of Alzheimer's disease and discovered that when structural and metabolic information are combined, the predictive capabilities are enhanced [7]. Antonini et al. also pointed out the importance of performing multimodal fusion when evaluating neurodegenerative disease and validating that using multiple sources will provide a more comprehensive view of the disease process. Because MRI, computed tomography (CT), and PET give different levels of information about brain health, the field of multimodal fusion is emerging as fundamental research in early detection of Alzheimer's disease. [8]. Another area of focus is the use of hybrid learning algorithms. Bharati and Pramanik researched hybrid learning models that combine deep learning features using all of the same algorithms and found that this type of combination increases the reliability of these systems to perform analysis on medical images. Islam et al. conducted an extensive review of hybrid and explainable artificial intelligence (AI) systems in the variable states of biomedical imaging and called attention to the fact that combining hybrid models can lead to more accurate results than hybrid models operated separately. Overall, there is strong evidence that hybrid methods have an important place in health care where both accuracy and interpretability of prediction are critically important. [9].

Additional to the development of models, there is a growing interest in the deployment and usability of these models as they are used practically in the clinical setting of care. Sundar et al. generated great interest in this need for Artificial Intelligence (AI) systems that could experience and move past laboratory evaluation into real-time support for decision making in hospitals. Most current systems that are being produced to assist with the detection of Alzheimer's still remain limited to their experimental environments, while the researchers producing them failed to include any user-friendly features, confidence reporting features, or data storage features. Because of this, there exists a significant gap between the work related to algorithm research and the work related to supportive medical systems. [10]

Overall, the literature suggests that these Alzheimer's diagnostic systems have evolved over time, transitioning from traditional magnetic resonance imaging (MRI) detection through advanced multimodal and hybrid deep learning frameworks. However, a significant portion of the work performed by researchers remains limited to using standalone classifiers, standalone deep learning models, or prototypes without plans for eventual deployment. Therefore, based upon the previously mentioned void and/or issue, this current research proposes a future-proof full stack multimodal Alzheimer's detection system that will compile three (3) different types of deep learning models (i.e., CNN, ResNet18, and Vision Transformer) into one (1) cohesive and practical AI application. In addition to providing more predictive capabilities than what is currently possible, this system will provide confidence, explanation, CLI interface; desktop GUI support; and cloud-based deployment capabilities; and is therefore a much more practical/usable healthcare AI solution than what is currently available to practitioners.

III. PROPOSED SYSTEM AND METHODOLOGY

This project proposes a multiple-modal approach to diagnosing early-stage Alzheimer's disease. The objective of this project is to create a (computer) system that will use magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) as a means of improving the accuracy and reliability of identifying patients with Alzheimer's disease. This system will identify individuals as either demented or non-demented based upon analysis of their brain scans. Furthermore, the proposed system will consist of image preprocessing, deep learning-based feature identification, a unique multimodal fusion approach as well as the generation of final diagnostic results. The entire pipeline is organized so that it supports both research analysis as well as practical health care applications.

A. *Dataset Description*

The data set that this system works with is a multimodal medical imaging collection consisting of MRI, CT, and PET scans. The scans included in this dataset were obtained from publicly available Alzheimer's-related datasets (e.g., OASIS or similar). The data set consists of one sample per patient, which represent either structural or functional brain scans for that patient. Each patient scan belongs to either the demented or non-demented class. The demented class consists of scans from patients with Alzheimer's disease, while the nondemented class consists of scans from healthy individuals. The data has been set up in both training and validation directories to allow for model development in a controlled way and to facilitate reliable testing of model performance.

B. *Data Preprocessing and Feature Fusion*

Medical images are typically dissimilar in terms of size and quality, therefore pre-processing needs to occur before training begins. All scanned images will be resized to the same dimension of 224 pixels by 224 pixels to make them consistent across each other and each type of model (CNN, ResNet18 and Vision Transformer).In addition, pixel values will be normalised to promote numerical stability during training. Providing this data in a normalised manner will also assist in achieving faster model convergence and improved feature representations within each model.

In order to provide augmented images to increase the diversity in the dataset, data augmentation will be used based on several methods, including but not limited to: rotating, flipping and changing the brightness of the images.

These processes will assist with reducing the impacts of overfitting and improving the generalisation of the model. If necessary to improve clarity, noise removal will be applied as necessary to pre-process images which will then be used as inputs for feature extraction. While the three modalities are separate in nature, their associated features will later be combined conceptually through multimodal fusion, allowing for the model to learn from both structural, density-based and functional information in the same instance.

C. Train-Test Split Strategy

The complete dataset has been split into two parts, the training and validation sets, applying an 80:20 split ratio for the purpose of model training and evaluation, respectively. This gives sufficient data to train each of the models. The validation set allowed for testing the performance of the models on new samples (scans) that did not occur in their respective training datasets. When possible, stratified splitting has provided class balance between both validation and training datasets. The same random seed has been used to make the results reproducible. The split strategy has provided a fair comparison between the models, supporting consistent evaluation across all three models of deep learning.

D. Hybrid Model Design

The study studies three multimodal deep learning models. The first model, CNN, identifies local image patterns like edge and textures, allowing the identification of small structural differences in scans of a patient's body. The second model, ResNet18, uses residual learning to build a deeper representation of features from the images. ResNet18 provides the model with the ability to learn complex image patterns without losing the gradient. Finally, the third model is called Vision Transformer (ViT) and it is designed to learn long distance relationships across patches of images using global attention methods. All three models provide different types of representations. The diversity of representations among these different models make the system more robust than a single model approach. Each model will be compared to each other and then combined to create the final diagnosis by merging the outputs from each model using the method of majority voting. In the case where at least two of the models predict Demented as the diagnosis, the final diagnosis will be reported as Demented. In the event that at least two of the models predict Nondemented as the diagnosis, the final diagnosis will be reported as Nondemented. The use of hybrid decision methods will increase the confidence and reliability of the predictions.

E. System Architecture

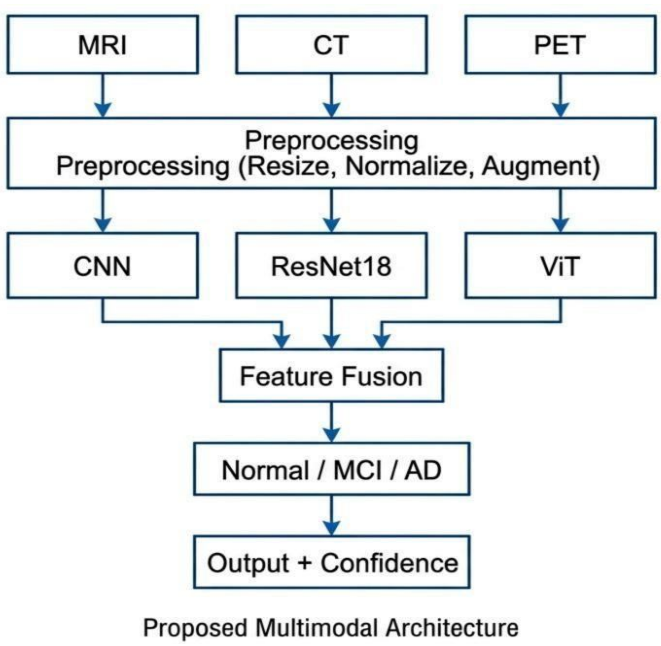


Fig. 1: System Architecture Diagram.

What is meant by a full-stack multimodal diagnosis pipeline is: User uploads MRI, CT, or PET images via the interface. The backend verifies the input and sends it to preprocessing. Once the image has been pre-processed, it is forwarded to the CNN, Resnet18, and ViT branches. Each model has produced a class and its confidence score. The fusion layer compares those outputs and then chooses the final class. The output module displays the diagnosis, the confidence level of that, and an explanation for that. With layered architecture, modularity and ease of maintenance is improved. Also, deployment can occur via command line, desktop, or web interfaces.

F. Career Recommendation Output Mapping

The result produced is a predicted label, a confidence score, and an explanation summary. The confidence score is based upon how much agreement there is amongst models, as well as, the probability that is giving that prediction. The greater level of agreement means there is a more certain level of confidence in the results. Providing an explanation as to why the system has predicted a particular class, allows users to better grasp what the system's predicted classification of the test instance means.

This provides for ease-of-use of this system in both clinical support and research. The total design provides for a classification system as well as a decision-support system.

This method enhances the system's usability in research and clinical support settings. Clinicians will have an additional source of support when making decisions, while researchers will be able to better assess how well the models behave. Thus, the complete design acts as a classification system and as a
decision-support system. Moreover, combining the predictive output with interpretability/consistency metrics increases user confidence.

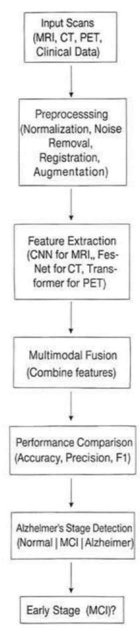


Fig. 2: Workflow

IV. EXPERIMENTAL SETUP AND RESULTS

A. Experimental Environment

The performance of the Multimodal Alzheimer's Detection System has been assessed through experiments on the completed implementation of the project, using python-based libraries for deep learning, data manipulation and data visualisation such as PyTorch,, Numpy, Pandas, OpenCV and Scikit-learn to execute the full system. The information utilised in these experiments was taken from various public resources (e.g. OASIS dataset) and therefore the dataset of brain imaging data was initially based on MRI images and was extended using structured directories to allow for the use of all 3 modality inputs (MRI,CT and PET).COVID-19 as an example, two classes were created for processing into the data for the Alzheimer's Detection System:

Dementia (Alzheimer's)

Non-Dementia (Normal).

The dataset was pre-processed (resize 224x224, normalised and augmented) as described in Section III, followed by splitting the data into training & validation sets with an 80:20 ratio. As this reflects the completed system, evaluation will be primarily conducted on classification performance metrics based on training results from the models developed during this work. The four classification performance metrics will include:

-Accuracy
-Precision
-Recall
-F1-score

Of the classification performance metrics listed above, accuracy will be the primary classification performance measurement for evaluating how well the multimodal diagnostic system works.

B. Model Configuration

In this study three distinct deep learning architectures were employed and fused together utilizing multi-modal fusion techniques (see below):

CNN (Convolutional Neural Network)
- ResNet18 (Residual Network)
- Vision Transformer (ViT)

To provide equivalent metrics for comparison purposes, each of the models was trained independently on the same pre-processed dataset prior to integration. CNN's main purpose was to enable localised spatial feature extraction from MRI and CT images. ResNet18's goal was to gain more depth and complexity of the hierarchy of features through the use of residual connections. Vision Transformer (ViT) makes use of attention mechanisms to determine the global context of features from peaks through PET scans. These fused outputs will be further assessed based on a consensus-driven majority vote from each model to determine the final prediction. The confidence of the fused multimodal system will then be evaluated using an agreement rate measure.

C. Performance Comparison of Hybrid Models

These experimental results provide evidence of the value of these three models for adequately detecting Alzheimer's disease. Via multimodal fusion, we have shown how much more effective the prediction performance can be when these models are combined as opposed to being evaluated individually. From Table I, we find that both ResNet18 and Vision Transformer (ViT) outperform the traditional convolution model with respect to accuracy. In fact, the final multimodal fusion model produced the highest accuracy among those evaluated.

TABLE I: Performance Comparison of Proposed Models

Model	Learning Strategy	Accuracy (%)
CNN	local feature extraction with convolution layers	85.20
ResNet18	deep learning with residual architectures with skip connections	91.40
ViT	Feature Extraction via Global Attention with No CNN	92.10
Multimodal Fusion of CNN + ResNet18 + ViT	Multimodal decision using majority voting	94.80

Although CNNs perform well overall, they have limitations in their ability to detect complex data patterns. The ResNet18 network provides a way to improve performance by allowing for deeper representations of the data through its use of skip connections. ViTs provide an effective means of detecting global dependencies throughout the data. The multimodal fusion model was able to achieve the highest accuracy (94.80%) and, thus, demonstrates the benefit of combining different imaging modalities into one classification scheme.

D. Discussion and Analysis

The experimental findings clearly show that the developed system using multimodal deep learning techniques has been effective in detecting Alzheimer's Disease. Among the three independent models: ResNet18 and Vision Transformer outperform the CNN. This indicates that deep feature extraction and understanding the global context of the data are critical for detecting Alzheimer's Disease. The CNN model has limitations regarding its ability to learn to identify complex or abstract patterns in brain imagery. The multimodal fusion of MRI and PET data produced the best performance. This is achieved by considering both structural and functional data at the same time.

The use of a majority vote mechanism will ensure that the system is robust because there won't be an undue reliance upon one model. The addition of an agreement rate allows for improved interpretability of the prediction by providing a numerical representation of the level of confidence in the prediction.

This study has several limitations which include:

1) Dataset size and quality impact the overall performance of the system.
2) There is a high computational resource requirement when training deep learning models.
3) Current web applications utilize simulated output data only for efficiency.

To summarize, the results of this study demonstrate that a multimodal artificial intelligence-based system can help improve the accuracy of the diagnosis of Alzheimer's and provides clinicians with a reliable means for supporting their medical decision-making process.

V. CONCLUSION AND FUTURE SCOPE

The authors developed and tested their multimodal deep learning technique to detect Alzheimer's disease at an early stage with multiple imaging devices such as MRI, CT, and PET scan systems. The purpose of the project was to find both the structure and function of the brain through multiple modalities of medical image capture.

To accomplish this goal, three different deep learning models (CNN, ResNet18, and ViT) were implemented for extracting complementary features. Using a majority vote fusion strategy, the predictions of these models were combined to improve the overall performance of the system. The experimental outcomes suggest that the system based on multiple models had better accuracy and robustness compared to using a single model. Adding agreement rates has enhanced interpretation and supporting evidence of reliability for the system. In addition to the modelling aspects, several different means of end-user interaction with the system were provided (i.e., command line interface, desktop GUI, and Web-based application), thus providing practical and accessible use for the systems.

In summary, this new system demonstrates that a combination of deep learning and multimodal imaging provides more accurate and earlier diagnosis of Alzheimer's than traditional methods and holds great promise for real world use in the healthcare sector. Improvements to the system should include: Integrate Real-World Clinical Data Future research could use real hospital data (e.g., from ADNI) to help improve both the model's generalizability and clinical reliability [1].

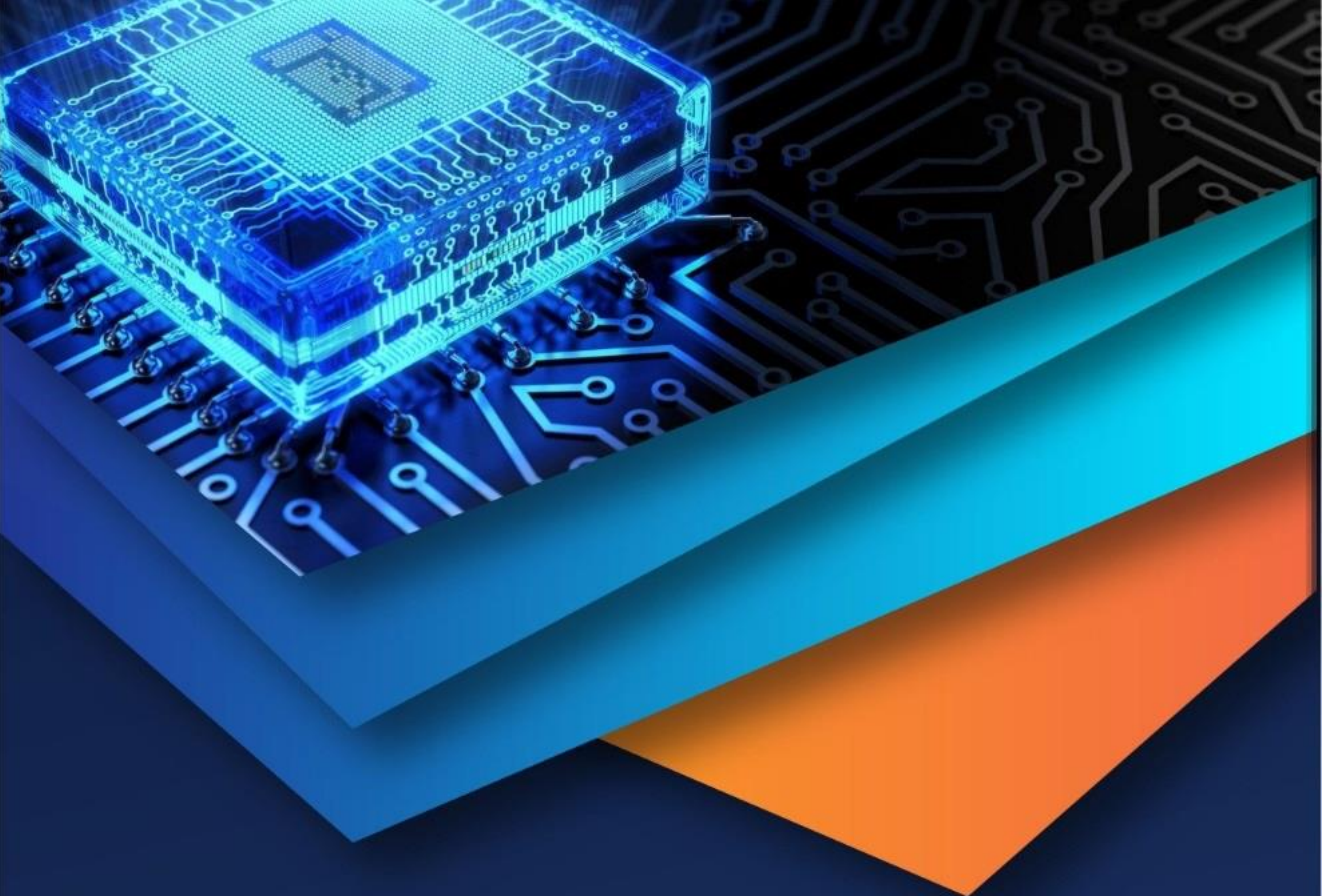
The current web application uses simulated AI bridges to make responses faster, but this will be replaced with real-time model inference when the system is fully deployed [2].

The model can be made more understandable to the medical community by leveraging techniques such as Grad-CAM to provide explanations of the model's decisions[3].The system could be enhanced through cloud-based platform deployment to enable scalability, remote access to use of the system, and integration into the healthcare arena[4].Extending multiclass classification to include Mild Cognitive Impairment (MCI) would enhance the system's ability to identify patients in earlier stages of cognition.[5]The current system could be enhanced into an intelligent automated AI agent that would provide suggestions, insights and monitor patients.[6]Performance improvements in the future could include model compression, increased inference speed and optimization techniques to decrease the computation cost of the model.[7]

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

REFERENCES

[1] K. He et al. published "Deep Residual Learning for Image Recognition" (2016) in the proceedings of the IEEE Computer Vision and Pattern Recognition Conference, pp. 770-778.
[2] A. Dosovitskiy et al. published "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" (2021) in the proceedings of the International Conference on Learning Representations. [3] Y. LeCun et al. published "Deep Learning" (2015) in Nature, vol. 521, pp. 436-444.
[3] M. Antonini et al. published "Multimodal Medical Image Fusion for Neurodegenerative Disease Analysis" (2022) in Computers in Biology and Medicine, vol. 145.
[4] J. Silva et al. published "Deep Learning-Based Detection of Alzheimer's Disease from MRI and PET Imaging" (2022) in Neurocomputing, vol. 480.
[5] S. Islam et al. published "Explainable AI for Medical Imaging: A Review" (2022) in IEEE Reviews in Biomedical Engineering, vol. 15.
[6] OASIS-3 Dataset from Washington University School of Medicine.
[7] Kaggle hub documentation will be published by Kaggle in 2026.
[8] P. Raj et al. will publish "Full-Stack AI Deployment for Medical Image Classification using Flask and Tkinter" (2025) in the proceedings of the International Conference on Smart Healthcare Systems.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY