iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: XII Month of publication: December 2024

DOI: https://doi.org/10.22214/ijraset.2024.65729

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Analysing of Alzheimer Diseases Using Data Augmentation and CNN

Dr. B. Narsimha[1], A. Ajay[2], B. Thirupathi[3], B. Satish Kumar Reddy[4]
[1]Professor, Teegala Krishna Reddy Engineering College
[2, 3, 4]Student, Teegala Krishna Reddy Engineering College

Abstract: Alzheimer's disease is the most prevalent form of dementia, characterized by a progressive decline in cognitive function. It typically begins with mild memory impairment and gradually leads to severe neurological deterioration, affecting an individual's ability to converse, interact, and respond to their environment. This irreversible condition involves critical regions of the brain responsible for thought processes, memory retention, and decision-making. The growing prevalence of Alzheimer's disease worldwide underscores the importance of timely diagnosis and effective interventions to enhance patients' quality of life and explore more efficient therapeutic strategies. Advanced diagnostic methods using imaging technologies, such as Magnetic Resonance Imaging (MRI), have revolutionized the early detection of Alzheimer's disease. These scans provide critical insights into the structural changes in the brain, enabling researchers to differentiate between healthy individuals and those with neurodegenerative conditions. However, manual analysis of these imaging data is time-consuming and prone to variability, necessitating automated and robust computational approaches for accurate diagnosis. In this research, we propose a novel framework for classifying MRI scans using Principal Component Analysis (PCA) for feature extraction, data augmentation techniques for expanding the dataset, and a 2D Convolutional Neural Network (2D CNN) for classification.
Keywords: Alzheimer's Disease, Dementia, Magnetic Resonance Imaging (MRI), Principal Component Analysis (PCA), Convolutional Neural Networks (CNN), Data Augmentation

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia worldwide. It leads to a gradual decline in cognitive functions, memory, reasoning, and the ability to perform everyday tasks. The disease predominantly affects older adults, but early-onset cases also exist. AD is characterized by the accumulation of amyloid plaques and tau tangles in the brain, leading to neuronal loss and brain shrinkage.

Early detection of Alzheimer's is crucial for effective intervention and management. Current diagnostic methods rely on clinical assessments, neuropsychological tests, and imaging techniques like MRI and PET scans. However, these methods are often subjective, time-consuming, and limited by data availability. Consequently, patients are frequently diagnosed at advanced stages when cognitive decline has significantly progressed, limiting treatment effectiveness.

Advancements in artificial intelligence (AI) and machine learning (ML) offer promising avenues to address these challenges. Convolutional Neural Networks (CNNs), a subset of deep learning models, have shown remarkable success in analyzing medical imaging data. By automatically extracting features from MRI scans, CNNs can classify brain conditions with high accuracy, even in early stages. Data augmentation techniques further enhance these models by increasing the diversity of training data, enabling better generalization and robustness.

This research explores the integration of data augmentation and CNNs to develop an automated framework for Alzheimer's diagnosis. By leveraging AI, the aim is to improve diagnostic precision, reduce subjectivity, and facilitate early intervention, thereby enhancing the quality of life for patients and their families.

II. CHALLENGES IN CURRENT DIAGNOSTIC TECHNIQUES

The diagnosis of Alzheimer's disease (AD) relies heavily on clinical evaluations, cognitive tests, and neuroimaging techniques such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET). Despite their importance, these methods face several challenges that hinder early and accurate detection:

1) *Subjectivity in Diagnosis:* Current methods are highly dependent on the clinical expertise of neurologists and radiologists. Variations in experience and interpretation can lead to inconsistent and delayed diagnoses, especially in early stages when symptoms overlap with normal aging.

2) *Data Limitations:* The availability of comprehensive and diverse datasets is a significant hurdle. Small sample sizes and a lack of representation across demographics (e.g., age, ethnicity) can result in biased models and limit the generalizability of diagnostic techniques.
3) *Manual Feature Extraction:* Traditional methods often involve manual extraction of imaging features, such as brain volume measurements or atrophy patterns. This process is time-consuming, labor-intensive, and prone to human error.
4) *Poor Sensitivity to Early Stages:* Early-stage Alzheimer's disease often presents with subtle and non-specific symptoms. Existing diagnostic tools struggle to detect these early changes, resulting in diagnoses only at advanced stages when interventions are less effective.
5) *High Computational Complexity:* Advanced imaging techniques like PET scans and volumetric MRI are computationally expensive and require sophisticated infrastructure. This limits their widespread accessibility, particularly in resource-constrained settings.
6) *Ethical and Privacy Concerns:* The use of sensitive patient data in medical imaging poses ethical challenges. Ensuring compliance with data protection regulations like HIPAA while maintaining robust diagnostic systems is a critical challenge.
7) *Over fitting in Machine Learning Models:* Machine learning models trained on small datasets risk overfitting, which reduces their ability to perform well on unseen data. This limitation is further exacerbated by the class imbalance often observed in Alzheimer's datasets.

These challenges highlight the need for innovative diagnostic approaches that leverage artificial intelligence, automation, and robust data handling techniques. By addressing these issues, future systems can provide earlier and more accurate diagnoses, improving patient outcomes and paving the way for timely therapeutic interventions.

III. RESEARCH APPROACH

To address the challenges in diagnosing Alzheimer's disease, this study presents a robust framework integrating Principal Component Analysis (PCA), data augmentation, and Convolutional Neural Networks (CNN). The proposed methodology aims to enhance diagnostic accuracy, improve early detection, and overcome the limitations of traditional approaches.

A. Data Acquisition and Preprocessing

- MRI Scans: High-quality brain MRI images are collected as the primary dataset.
- Preprocessing: Images are resized, normalized, and denoised to standardize input data. This step removes artifacts and ensures consistency, making the dataset suitable for further analysis.

B. Feature Extraction Using PCA

- Dimensionality Reduction: Principal Component Analysis is applied to reduce the high-dimensional MRI data into a smaller set of significant features.
- Improved Efficiency: By retaining only the most informative features, PCA simplifies the data while preserving critical patterns, reducing computational complexity and risk of overfitting.

C. Data Augmentation

- Techniques: Augmentation methods such as rotation, flipping, scaling, and intensity adjustments are employed to artificially expand the dataset.
- Benefits:
  - Increases dataset diversity, improving model robustness and generalization.
  - Addresses class imbalance by creating more samples for underrepresented categories.

D. Classification with 2D Convolutional Neural Networks (CNN)

- Architecture
  - Convolutional Layers: Extract spatial features from MRI images, such as structural anomalies.
  - Pooling Layers: Reduce spatial dimensions, improving computational efficiency and feature abstraction.
  - Fully Connected Layers: Integrate extracted features for final classification into categories like healthy, mild cognitive impairment (MCI), or Alzheimer's disease.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

- Training
  - The CNN is trained on augmented and PCA-processed data, ensuring it learns diverse and meaningful patterns.
  - Loss and accuracy trends are monitored to optimize performance.

E. *Evaluation Metrics*

- Metrics**:** Accuracy, sensitivity, specificity, and F1-score are used to evaluate the model's performance. Comparative analysis against traditional methods highlights the improvements achieved by the proposed approach.

F. *Integration and Deployment*

The entire framework is integrated into an automated pipeline, capable of real-time analysis of MRI scans. This ensures scalability and clinical applicability, making the system suitable for diverse healthcare settings. The proposed methodology leverages advanced AI techniques to overcome the limitations of existing diagnostic methods, enabling earlier detection and more reliable diagnosis of Alzheimer's disease.

IV. FEATURE EXTRACTION

Feature extraction is a critical step in the proposed methodology to enhance the performance of machine learning models, particularly when working with complex datasets like MRI scans. In this study, Principal Component Analysis (PCA) is employed as the primary technique for feature extraction. PCA is a dimensionality reduction method that simplifies large datasets by transforming them into a smaller set of uncorrelated variables, known as principal components, while retaining the most significant patterns in the data.

A. *Steps Involved in PCA for Feature Extraction*

1) *Preprocessing of MRI Data:* MRI images are first preprocessed to ensure they are in a consistent format. This includes resizing the images to a standard size, normalizing pixel values to a consistent range, and reducing noise. Preprocessing is crucial to eliminate variations caused by imaging artifacts and to standardize the data for further analysis.
2) *Covariance Matrix Computation:* PCA begins by calculating the covariance matrix of the dataset. The covariance matrix expresses the variance and correlation between different variables (or pixels) in the image. The aim is to capture the linear relationships among the different pixel intensities and identify patterns that explain the most variance in the data.
3) *Eigenvalue and Eigenvector Calculation:* PCA computes the eigenvalues and eigenvectors of the covariance matrix. The eigenvectors represent the directions in which the data varies the most, while the eigenvalues indicate the magnitude of variance along these directions. The top eigenvectors, corresponding to the largest eigenvalues, are selected as the principal components. These components represent the most important features of the dataset that contribute significantly to the variation in the MRI images.
4) *Dimensionality Reduction:* The selected principal components are used to transform the original high-dimensional MRI images into a lower-dimensional space. This reduces the complexity of the data while retaining the most important features for further analysis. By retaining only the top principal components, PCA reduces noise and removes redundant information, making the dataset more manageable for machine learning algorithms.

B. *Reconstruction of the Data*

After reducing the dimensions, the transformed data is passed on to the CNN for classification. The reduced features are used to ensure that only the most relevant information is fed into the deep learning model, improving both training efficiency and accuracy.

1) *Output*

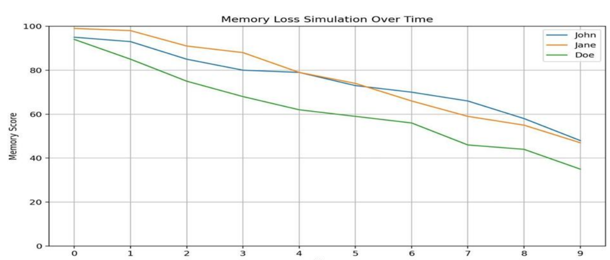
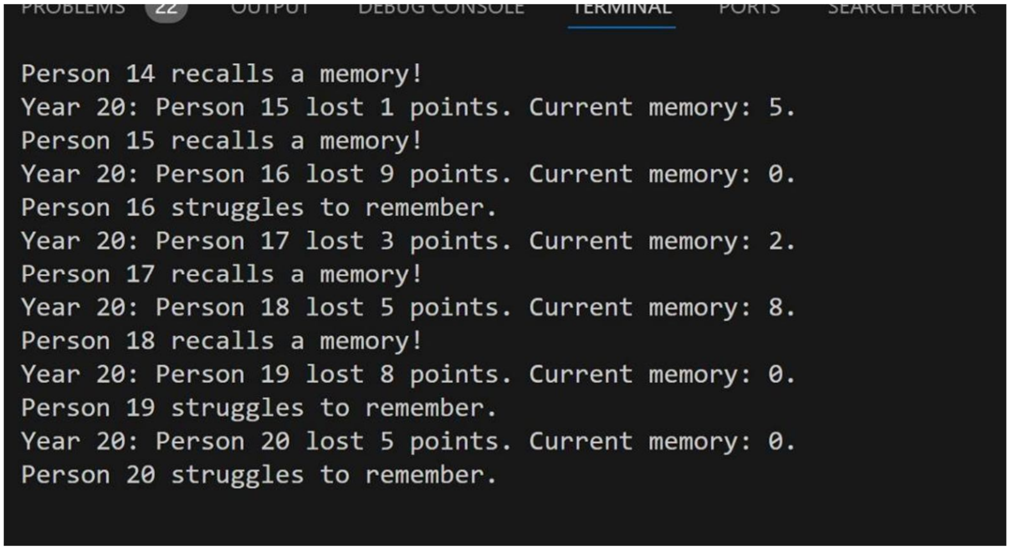


Fig: Memory Loss Simulation Over Time for Alzheimer's Disease

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com



```
PROBLEMS  22   OUTPUT   DEBUG CONSOLE   TERMINAL   PORTS   SEARCH ERROR

Person 14 recalls a memory!
Year 20: Person 15 lost 1 points. Current memory: 5.
Person 15 recalls a memory!
Year 20: Person 16 lost 9 points. Current memory: 0.
Person 16 struggles to remember.
Year 20: Person 17 lost 3 points. Current memory: 2.
Person 17 recalls a memory!
Year 20: Person 18 lost 5 points. Current memory: 8.
Person 18 recalls a memory!
Year 20: Person 19 lost 8 points. Current memory: 0.
Person 19 struggles to remember.
Year 20: Person 20 lost 5 points. Current memory: 0.
Person 20 struggles to remember.
```

Fig : Simulation of Memory Loss in Alzheimer's Disease Over Time

C. Advantages

1) *Improved Accuracy and Efficiency:* By leveraging PCA, the dimensionality of the MRI data is significantly reduced, making it easier and faster for the CNN to process the data. This improves both training efficiency and classification accuracy. PCA helps focus on the most significant features, which enhances the model's ability to accurately detect Alzheimer's-related abnormalities in brain scans.
2) *Data Augmentation for Robustness:* Data augmentation techniques such as rotation, flipping, scaling, and intensity adjustments help expand the dataset artificially. This ensures that the model is exposed to a wider range of data, improving its generalization ability and reducing the risk of overfitting. This is particularly important in the medical field, where datasets are often limited or imbalanced.
3) *Early Detection:* The proposed methodology is designed to detect Alzheimer's disease at an earlier stage compared to traditional methods. By focusing on subtle variations in MRI scans and employing CNNs to classify these patterns, the system can identify early indicators of Alzheimer's, enabling timely intervention and potentially slowing disease progression.
4) *Automation and Speed:* The integration of PCA and CNN allows for an automated diagnostic process, reducing the reliance on human intervention. This not only speeds up the diagnosis but also reduces the potential for human error, leading to more consistent and reliable results. Faster diagnoses enable healthcare professionals to make timely decisions about patient care.
5) *Scalability and Real-Time Deployment:* The methodology is designed to be scalable, meaning it can be deployed in various healthcare settings, from large hospitals to smaller clinics. Additionally, real-time processing capabilities allow the system to be integrated into existing clinical workflows, providing immediate feedback during patient examinations.

D. Disadvantages

1) *Data Dependency:* While data augmentation helps in expanding the dataset, the quality of the model still heavily depends on the quality and diversity of the original dataset. If the dataset is not representative of the general population or lacks sufficient variation, the model might not perform well on unseen data, leading to biased or inaccurate results.
2) *Risk of Overfitting with Inadequate Augmentation:* While data augmentation helps prevent overfitting by increasing dataset diversity, if the transformations applied (e.g., rotations, noise addition) are not realistic or do not reflect real-world data variability, the model could learn patterns that do not exist in genuine clinical data. This could lead to poor generalization and reduced model performance on real MRI scans.
3) *High Computational Requirements:* Although PCA reduces the dimensionality of the data, the CNN model still requires significant computational resources, such as high-performance GPUs, for training and inference. This can be a limitation for smaller healthcare institutions that may not have access to such resources. Additionally, the training time for CNN models can be lengthy, especially with large datasets.

4) *Interpretability and Explainability:* Deep learning models, particularly CNNs, are often considered "black boxes" due to their complex nature. This lack of interpretability can be a barrier to their widespread clinical adoption. Healthcare professionals may find it difficult to trust or understand the rationale behind the model's predictions, particularly when decisions have significant consequences for patient care.
5) *Ethical and Privacy Concerns:* The use of sensitive medical data, such as MRI scans, raises ethical concerns regarding patient privacy and consent. Ensuring that the data is anonymized and complies with regulations like HIPAA is essential to protect patient rights. Furthermore, biased datasets could result in unfair diagnoses for certain demographic groups, such as underrepresented age groups or ethnicities.
6) *Risk of Model Bias:* If the training data is not sufficiently diverse, the CNN may become biased, performing poorly on groups that were underrepresented in the training set. This could lead to disparities in diagnostic accuracy for different population groups, undermining the fairness and effectiveness of the model.

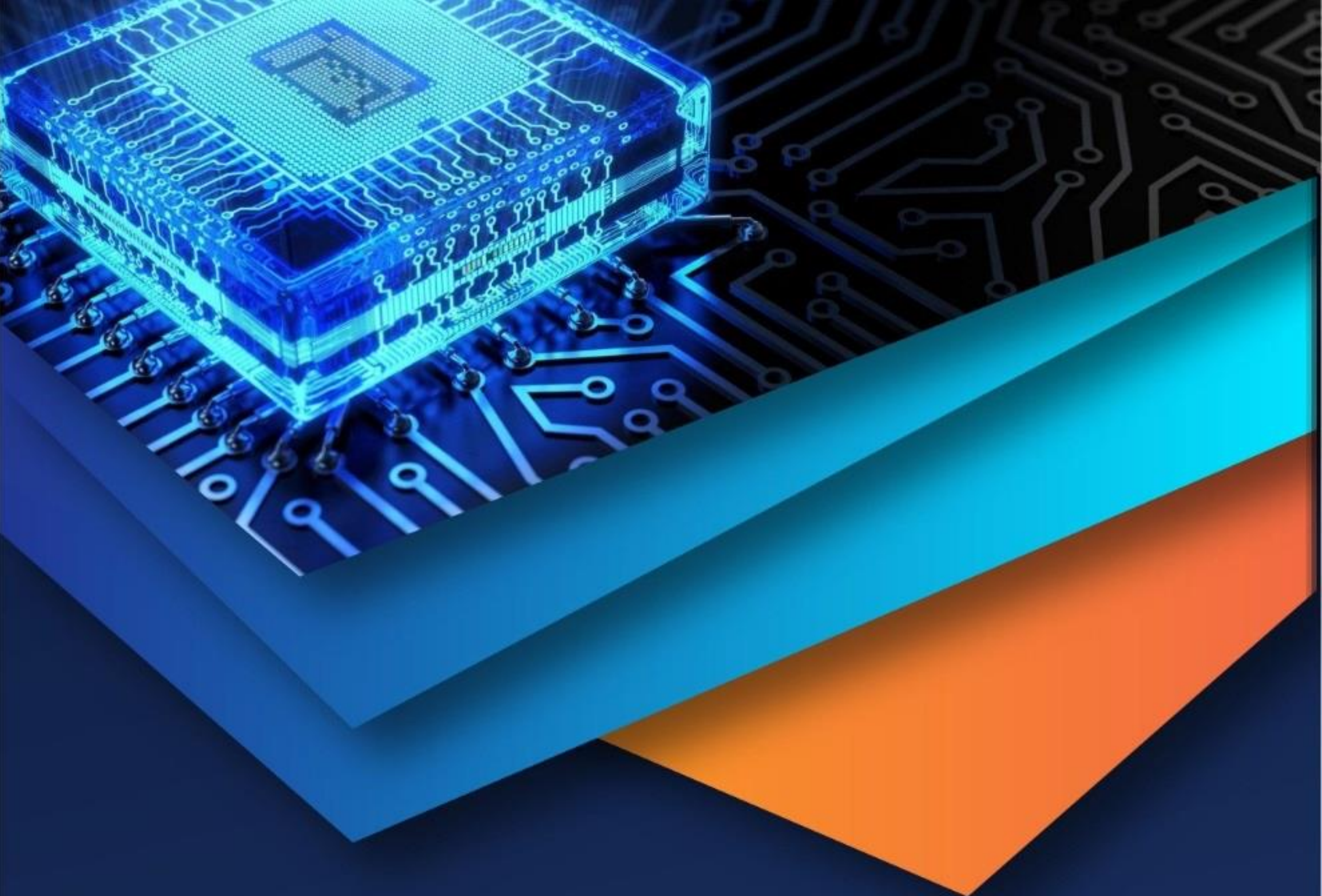
V. CONCLUSION

In conclusion, the integration of Principal Component Analysis (PCA), data augmentation, and Convolutional Neural Networks (CNN) presents a powerful approach to the early diagnosis of Alzheimer’s disease using MRI scans. The proposed methodology effectively addresses the limitations of traditional diagnostic methods, such as subjectivity, data scarcity, and delayed detection, by automating the feature extraction and classification process. PCA helps reduce the dimensionality of the data, retaining only the most significant features, which improves both computational efficiency and model accuracy. Data augmentation further strengthens the model by increasing dataset diversity, mitigating overfitting, and enhancing the model’s ability to generalize across different datasets. The CNN, trained on augmented and PCA-processed MRI data, is capable of identifying subtle brain changes indicative of Alzheimer’s disease, even in its early stages, thus enabling timely intervention. This methodology not only enhances diagnostic precision but also speeds up the process, reducing the burden on healthcare professionals and ensuring faster decision-making.

However, challenges such as high computational requirements, data dependency, and the need for model interpretability remain. Addressing these issues will be critical for broader clinical adoption and for ensuring that the system remains unbiased and accessible across diverse healthcare settings. Overall, this research demonstrates the potential of leveraging advanced machine learning techniques to revolutionize the diagnosis and management of Alzheimer’s disease. With continued improvements in model transparency, dataset diversity, and computational efficiency, this approach could significantly enhance early detection, improve patient outcomes, and contribute to the global effort in combating Alzheimer’s disease.

REFERENCES

[1] Islam, Jyoti, and Yanqing Zhang. ”Brain MRI analysis for Alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks.” Brain informatics 5 (2018): 1- 14, 2018.
[2] Liu, Manhua, et al. ”Multi-modality cascaded convolutional neural networks for Alzheimer’s disease diagnosis.” Neuroinformatic 16 (2018): 295-308, March 2018.
[3] Basaia, Silvia, et al. ”Automated classification of Alzheimer’s disease and mild cognitive im pairment using a single MRI and deep neural networks.” Neuroimage: Clinical 21 (2019): 101645, December 2018.
[4] Fang, Xusheng, Zhenbing Liu, and Mingchang Xu. ”Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer’s disease diagnosis.” IET Image Processing 14.2 (2020): 318- 326, 2019.
[5] Jo, Taeho, Kwangsik Nho, and Andrew J. Saykin. ”Deep learning in Alzheimer’s disease: diag nostic classification and prognostic prediction using neuroimaging data.” Frontiers in aging neuro science 11 (2019): 220, August 2019.
[6] Raju, Manu, et al. ”Classification of mild cognitive impairment and Alzheimer’s disease from magnetic resonance images using deep learning.” 2020 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT). IEEE, November 2020.
[7] Chitradevi, D., S. Prabha, and Alex Daniel Prabhu.”Diagnosis of Alzheimer disease in MR brain images using optimization techniques.” Neural Computing and Applications 33 (2021): 223-237, May 2020.
[8] Suh, C. H., et al. ”Development and validation of a deep learning–based automatic brain segmen tation and classification algorithm for Alzheimer disease using 3D T1-weighted volumetric images.” American Journal of Neuroradiology 41.12 (2020): 2227- 2234, November 2020.
[9] Buvaneswari, P. R., and R. Gayathri. ”Deep learning-based segmentation in classification of Alzheimer’s disease.” Arabian Journal for Science and Engineering 46 (2021): 5373-5383, January 2021.
[10] Raees, PC Muhammed, and Vinu Thomas. ”Automated detection of Alzheimer’s Disease using Deep Learning in MRI.” Journal of Physics: Conference Series. Vol. 1921. No. 1. IOP Publishing,
[11] Ramzan, Farheen, 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 44 (2020): 1-16.
[12] Yamanakkanavar, Nagaraj, Jae Young Choi, and Bumshik Lee. ”MRI segmentation and classi f ication of human brain using deep learning for diagnosis of Alzheimer’s disease: a survey.” Sensors 20.11 (2020): 3243



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