Journal of Harbin Engineering University

ISSN: 1006-7043

Vol 47 No. 03

March 2026

Early Alzheimer's Detection Prediction Using Explainable Artificial Intelligence (Xai) Techniques For Convolutional Neural Networks (Cnns)

[1]R.Malarvizhi, [2]Dr. R. Rangaraj

Assistant Professor, Department of Computer Science,

N.M.S.S.Vellaichamy Nadar College, Madurai.

Professor & Head, Department of Computer Science

Hindusthan College of Arts & Science, Coimbatore.

Abstract- This paper focuses on Explainable Artificial Intelligence (XAI) techniques for Convolutional Neural Networks (CNNs) to enhance interpretability and transparency in model predictions. Leveraging methods like Grad-CAM, Saliency Maps, and Layer-wise Relevance Propagation (LRP), the work explains CNN decisions by identifying regions in input images that contribute most to predictions. Grad-CAM generates heatmaps by combining feature maps with class-specific gradients, while Saliency Maps highlight critical pixels using gradient sensitivity. LRP decomposes the output prediction into pixel-level relevance scores through backpropagation. The proposed algorithm integrates these techniques to provide comprehensive visual explanations, aiding in trust and transparency. Key applications include medical diagnosis and autonomous systems, where understanding model decisions is critical for sensitive and high-stakes scenarios. This framework ensures robustness in decision-making and promotes the responsible deployment of CNN-based systems.

Keywords: Explainable AI (XAI), Convolutional Neural Networks (CNNs), Grad-CAM, Saliency Maps, Interpretability, Autonomous Systems.

1. Introduction

Alzheimer's disease (AD) represents a growing global health challenge with significant socio-economic implications. As the most common form of dementia, AD progressively impairs cognitive function, affecting memory, reasoning, and daily activities. Early detection of Alzheimer's is crucial for timely intervention and improved patient outcomes. The integration of data mining techniques offers a promising avenue for enhancing early detection and classification of Alzheimer's disease. In this context, this paper introduces a comprehensive exploration of prediction and classification methodologies using data mining tools to contribute to the ongoing efforts in improving early Alzheimer's detection.

Alzheimer's disease is characterized by the accumulation of abnormal protein deposits, such as beta-amyloid plaques and tau tangles, in the brain. Detecting these biomarkers at an early stage can facilitate the identification of individuals at risk before the manifestation of noticeable cognitive decline. Data mining, a multidisciplinary field that combines techniques from statistics, machine learning, and database management, has emerged as a powerful tool for uncovering patterns, trends, and relationships within large datasets. By leveraging the vast amount of information available in various forms – from genetic data to neuroimaging scans and clinical records – data mining techniques enable the extraction of meaningful insights that can aid in the early diagnosis and prognosis of Alzheimer's disease.

One of the key challenges in Alzheimer's research lies in the complexity and heterogeneity of the data associated with the disease. Data mining techniques provide the means to navigate through this complexity, allowing for the identification of relevant features and patterns that may be indicative of early-stage Alzheimer's. This paper focuses on the prediction and classification aspects of data mining, aiming to develop algorithms that

can accurately discriminate between individuals with early-stage Alzheimer's and those without.

The predictive modeling involves training algorithms on historical data to recognize patterns associated with Alzheimer's risk factors or biomarkers. This enables the development of models capable of predicting the likelihood of an individual developing Alzheimer's in the future. On the other hand, classification algorithms are designed to categorize individuals into predefined groups based on observed features. In the context of Alzheimer's, these groups typically include categories like "cognitively normal," "mild cognitive impairment (MCI)," and "Alzheimer's disease." Accurate classification is vital for ensuring appropriate interventions and treatments are administered to individuals at different stages of cognitive decline.

The introduction of this paper sets the stage for a detailed exploration of various data mining techniques employed in predicting and classifying early Alzheimer's cases. The subsequent sections will delve into specific methodologies such as feature selection algorithms, ensemble learning, and neural networks, highlighting their contributions to the field. Furthermore, the paper will discuss the challenges associated with data mining in Alzheimer's research, ethical considerations, and potential avenues for future research. By elucidating the role of data mining in early Alzheimer's detection, this paper aims to contribute to the ongoing quest for more effective diagnostic and prognostic tools in the battle against this debilitating neurodegenerative disease.

Aim of the Proposed Approach

This paper proposes an algorithmic framework for prediction and classification aimed at early Alzheimer's detection. By combining data mining techniques, specifically Correlation-based Feature Selection (CFS) and Recursive Feature Elimination (RFE), the goal is to enhance the efficiency of identifying relevant biomarkers from complex datasets associated with Alzheimer's disease. The algorithm seeks to not only improve diagnostic accuracy but also provide a more interpretable set of features that could offer insights into the underlying mechanisms of the disease.

Structure of the Paper

The subsequent sections of this paper will delve into the proposed algorithm, detailing the application of CFS and RFE in feature selection for early Alzheimer's detection. Experimental results and discussions will follow, showcasing the algorithm's efficacy and potential clinical implications. The paper concludes by emphasizing the significance of data mining in revolutionizing early Alzheimer's detection and suggesting avenues for future research in this critical domain.

Objective of the Paper

The primary objective of this research is to develop an advanced algorithm for the prediction and classification of early Alzheimer's using data mining techniques. The paper specifically focuses on the utilization of machine learning algorithms for feature selection and classification to enhance the accuracy of early detection models.

Significance of Prediction and Classification

The prediction aspect involves forecasting the likelihood of an individual developing Alzheimer's based on relevant features, while classification categorizes individuals into different risk groups. Both aspects are crucial for tailoring personalized interventions and optimizing resource allocation in healthcare systems.

Anticipated Contributions

The proposed algorithm is expected to contribute significantly to the field by improving the accuracy and interpretability of early Alzheimer's detection models. The refined feature subset is anticipated to enhance the understanding of relevant biomarkers, facilitating more targeted and effective interventions. The paper's outcomes may also inform the development of scalable and reliable tools for healthcare practitioners, advancing the early diagnosis and management of Alzheimer's disease.

2. Literature Survey

2.1 Stacked Sparse Autoencoder (SSA)

E. Jabason (2018) et.al proposed Missing Structural and Clinical Features Imputation for Semi-supervised Alzheimer's Disease Classification using

Stacked Sparse Autoencoder. This paper addresses the challenge of missing data in Alzheimer's disease (AD) detection, particularly in the Tadpole paper of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Utilizing a novel stacked sparse autoencoder-based imputation method, the proposed algorithm effectively infers missing values and identifies significant structural and clinical features for distinguishing AD, mild cognitive impairment (MCI), and cognitively normal (CN) cases. Experimental results, employing five-fold cross-validation, demonstrate superior performance in accuracy, sensitivity, and specificity compared to existing methods. The approach, leveraging unsupervised learning and MRI data, offers a promising avenue for improved AD diagnosis, with future work focusing on predicting additional clinical features.

2.2 Combined Point Detection Feature Extraction for Alzheimer's Prediction (CPDFE-AP)

Dinu A J (2021) et.al proposed a Novel Modelling Technique for Early Recognition and Classification of Alzheimer's disease.This paper proposes a new algorithm that utilizes a combination of point detection-based feature extraction methods, including SURF, FAST, BRISK, Harris, and Min Eigen, for the early prediction of various stages of Alzheimer's disease. The performance of the proposed method is evaluated by combining it with Random Forest and Tree Bagger classifiers. Experimental results show that the classification accuracy achieved with the Random Forest classifier is 98.42%, while the accuracy with the Tree Bagger classifier is 98.17%. The Random Forest classifier outperforms Tree Bagger, offering higher accuracy, sensitivity, and specificity. The proposed method is versatile, handling both classification and regression tasks effectively, and works well with both categorical and continuous data. Overall, the algorithm demonstrates superior performance compared to methods that rely on single feature extraction and selection techniques for Alzheimer's disease prediction and classification.

2.3 Deep Learning-Based MRI Classification for Alzheimer's Detection (DL-MRI-AD)

A. W. Salehi (2020) et.al proposed a CNN Model: Earlier Diagnosis and Classification of Alzheimer Disease using MRI Alzheimer's Disease (AD) is the most common form of dementia that can lead to a neurological brain disorder that causes progressive memory loss as a result of damaging the brain cells and the ability to perform daily activities. Using MRI (Magnetic Resonance Imaging) scan brain images, we can get the help of Artificial intelligence (AI) technology for detection and prediction of this disease and classify the AD patients whether they have or may not have this deadly disease in future. The main purpose of doing all this is to make the best prediction and detection tools for the help of radiologists, doctors, caregivers to save time, cost, and help the patient suffering from this disease. In recent years, the Deep Learning (DL) algorithms are very useful for the diagnosis of AD as DL algorithms work well with large datasets. In this paper, we have implemented Convolutional Neural Network (CNN) for the earlier diagnosis and classification of AD using MRI images, the ADNI 3 class of images with the total number of 1512 mild, 2633 normal and 2480 AD were used. A significant accuracy of 99% achieved in which the model performed well as we compared with many other related works. Furthermore, we also compared the result with our previous work on which ma-chine learning algorithms were applied using OASIS dataset and it showed that when dealing with large amount of data like medical data the deep learning approaches can be a better option over the traditional machine learning techniques.

2.4 Brain Network and Clinical Text Integrated Deep Learning for Alzheimer's Diagnosis (BNCT-DLAD)

P. Zhou (2019) et.al proposed Early Diagnosis of Alzheimer's Disease Based on Resting-State Brain Networks and Deep Learning. Computerized healthcare has undergone rapid development thanks to the advances in medical imaging and machine learning technologies. Especially, recent progress on deep learning opens a new era for multimedia based clinical decision support. In this paper, we use deep learning with brain network and clinical relevant text information to make early diagnosis of Alzheimer's disease (AD). The clinical relevant text information includes age, gender and ApoE gene of the subject. The brain

network is constructed by computing the functional connectivity of brain regions using resting-state functional magnetic resonance imaging (R-fMRI) data. A targeted autoencoder network is built to distinguish normal aging from mild cognitive impairment, an early stage of AD. The proposed method reveals discriminative brain network features effectively and provides a reliable classifier for AD detection. Compared to traditional classifiers based on R-fMRI time series data, about 31.21% improvement of the prediction accuracy is achieved by the proposed deep learning method, and the standard deviation reduces by 51.23% in the best case that means our prediction model is more stable and reliable compared to the traditional methods. Our work excavates deep learning's advantages of classifying high-dimensional multimedia data in medical services, and could help predict and prevent AD at an early stage.

2.5 Graph Signal and Convolutional Network for Alzheimer's Detection (GSCN-AD)

H. Padole (2018) et.al proposed Early Detection of Alzheimer's Disease using Graph Signal Processing on Neuroimaging Data. Brain imaging signals obtained using different imaging modalities mostly reside on irregular structures. While most of the classical signal processing methods is designed for signals having a regular structure, a new field of signal processing called Graph Signal Processing (GSP) is growing rapidly which deals with the irregularly structured data. So, GSP has become a natural choice for many brain image analysis applications. In this paper, we consider the problem of detection of Alzheimer's Disease (AD) in the early stages using fMRI data obtained from ADNI dataset. Firstly, we extract efficient discriminating features from resting state fMRI data using our novel hypothesis which is based on the outcomes of two neurological experiments carried out independently. Then we classify these graph signals by designing a classifier based on recently proposed graph convolutional neural network (GCNN). GCNN is the generalization of convolutional neural network (CNN) to the irregular domain using the concepts of GSP. We constructed brain graphs using different connectivity measures and compared the performance obtained using these graphs to find the best suitable connectivity measure for our application. Our proposed model outperforms state-of the-art AD detection methods with a classification accuracy of 92.44%. This improvement can be associated with the fact that we first extracted highly discriminating features using graph frequency analysis performed with suitably constructed graphs and then applied properly designed GCNN classifier to classify the input graph signals.

3. Research Methodology

Explainable AI (XAI) for Convolutional Neural Networks (CNNs) focuses on providing transparency and interpretability for complex deep learning models, which are typically viewed as "black boxes." The goal of XAI is to help users understand how a CNN arrives at its predictions, especially in critical fields like healthcare and autonomous driving. Below are some of the key XAI techniques for CNNs along with their associated formulas and equations:

3.1. Activation Maps and Feature Visualization

Activation maps help visualize which parts of an input (such as an image) are important for the CNN's prediction. CNNs work by convolving input images with filters (kernels), generating feature maps at each convolutional layer. These feature maps show how the model responds to different spatial areas of the input image.

- Let X represent the input image and f_k the k^{th} filter in the convolutional layer.
- The output of a convolution operation between the image and a filter is given by:

$$Y_k = X * f_k$$

Where * denotes the convolution operation. The resulting Y_k is the feature map corresponding to the filter f_k.

- **Activation Map Visualization**: After convolution, we visualize the activations at the output, often using techniques like Grad-CAM to show which parts of the feature map contributed to the final decision.

3.2. Grad-CAM (Gradient-weighted Class Activation Mapping)

Grad-CAM provides a way to visualize which regions of an input image contribute most to the CNN's decision by generating heatmaps. This is useful in understanding which areas in an image influenced the classification decision.

1. **Grad-CAM Generation**: Grad-CAM uses the gradients of the output with respect to the last convolutional layer's feature maps. Let C be the predicted class, and let A_k represent the feature map at the k^{th} location of the last convolutional layer.

$$\frac{\partial C}{\partial A_k}$$

Where $\frac{\partial C}{\partial A_k}$ is the gradient of the class score with respect to the activations at location k.

2. **Weighted Combination**: The gradients are then globally averaged to produce weights α_k:

$$\alpha_k = \frac{1}{Z}\sum_{i,j}\frac{\partial C}{\partial A_k}$$

Where Z is the normalization factor (typically the number of spatial locations in the feature map).

3. **Heatmap Construction**: A weighted combination of feature maps is then computed to produce the final class activation map:

$$Grad - CAM = ReLU\left(\sum_k \alpha_k A_k\right)$$

Where the $ReLU$ function ensures that only positive contributions are considered. This heatmap shows which regions of the image were most influential in predicting the class.

3.3. Saliency Maps

Saliency maps highlight which parts of the input image are most important for the CNN's decision by computing the gradient of the output with respect to the input pixels. Saliency maps are used to show which areas of the input contribute to the decision.

1. **Gradient Calculation**: Let $F(X)$ be the CNN's output for the input image X. The saliency map $S(X)$ is computed as the gradient of the output with respect to the input image pixels:

$$S(X) = \frac{\partial F(X)}{\partial X}$$

This gives a matrix of the same size as the input image, where each pixel indicates how much it contributes to the final output.

2. **Saliency Map Generation**: To visualize the importance of each pixel, we typically take the absolute value of the gradient:

$$S_{abs}(X) = \left|\frac{\partial F(X)}{\partial X}\right|$$

This map is then visualized as an image where brighter regions indicate higher importance.

3.4. Class Activation Mapping (CAM)

Class Activation Mapping (CAM) is a technique similar to Grad-CAM but is specifically applicable to CNNs with a Global Average Pooling (GAP) layer before the classification layer. It generates class-specific heatmaps by using the final convolutional layer's feature maps and class weights.

1. Feature Maps: Let A_k represent the feature maps at the k^{th} position in the final convolutional layer, and u represent the weights associated with these feature maps from the classification layer.

2. **Class Score**: The class score for a particular class C is calculated as the weighted sum of the feature maps A_k:

$$s(C) = \sum_k w_k A_k$$

3. **Activation Map Generation**: To generate the class activation map for class C. the weighted feature maps are combined:

$$CAM = \sum_k w_k A_k$$

This results in a heatmap showing which regions of the image are most important for predicting lass C.

3.5. Layer-wise Relevance Propagation (LRP)

Layer-wise Relevance Propagation (LRP) explains CNN decisions by backpropagating the relevance score from the output layer to the input layer. It decomposes the output prediction into contributions from individual neurons and

propagates these contributions through the layers of the network.

1. **Relevance Propagation**: Let R_l denote the relevance score for the layer l and x_l denote the activation of the layer's neurons. The relevance for each neuron is propagated backward using the following rule:

$$R_l = \sum_i \frac{x_l^i \cdot W_{l,i}}{\sum_j x_l^i \cdot W_{l,i}} R_{l+1}$$

Where:

- $W_{l,i}$ is the weight between neuron i in layer l and the neurons in the next layer.
- R_{l+1} the relevance score from the next layer. score from
- This propagation continues until the input layer is reached.

2. **Input Relevance**: The final relevance score at the input layer, R_{input} represents the contribution of each pixel to the model's decision.

$$R_{input} = \sum_i \frac{x_{input}^i \cdot W_{input,i}}{\sum_j x_{input}^j \cdot W_{input,j}} R_{output}$$

Where $R_{output}$is the relevance score at the output layer.

Algorithm for Explainable AI (XAI) for Convolutional Neural Networks (CNNs)

This algorithm outlines the steps to apply Explainable AI (XAI) techniques, particularly Grad-CAM**,** Saliency Maps**,** and Layer-wise Relevance Propagation (LRP), to Convolutional Neural Networks (CNNs) to explain their predictions.

Input:

- X: Input image.
- f: Pre-trained CNN model.
- C: Class prediction.
- y: True class label.

Output:

- Heatmap or saliency map showing regions of the input image most important for the model's decision.
- Relevance scores at each layer of the CNN.

Step 1: CNN Model Prediction

1.1. Input the image X into the CNN model f.

1.2. Perform a forward pass through the model to get the predicted class C and class probabilities.

- $C = argmax(f(x))$
- $P(C) = f(X)$

Step 2: Gradient-based Explanations (Grad-CAM)

2.1. Compute Gradient of Class Score:

- Let y be the output class score for the predicted class C.
- Compute the gradient of y with respect to the last convolutional layer's feature map A_k:

$$\frac{\partial C}{\partial A_k}$$

- These gradients indicate how much the activation of each feature map contributes to the class score.

2.2. Global Average Pooling of Gradients:

- Compute the average gradient over all spatial locations in A_k to get the weight α_k:

$$\alpha_k = \frac{1}{Z} \sum_{i,j} \frac{\partial C}{\partial A_k}$$

where Z is the normalization factor, typically the number of spatial locations in the feature map.

2.3. Class Activation Map (CAM):

- Combine the weighted feature maps to generate the Class Activation Map (CAM):

$$CAM = \sum_k \alpha_k A_k$$

- Apply ReLU to highlight only positive contributions:

$$CAM_{Final} = ReLU(CAM)$$

- This will produce a heatmap indicating the most relevant regions for the model's prediction.

Step 3: Saliency Map

3.1. Compute Gradient of Input Image:

- Compute the gradient of the class score y with respect to the input image X:

$$S(X) = \frac{\partial y}{\partial X}$$

- This gradient shows how sensitive the model's output is to changes in the input pixels.

3.2. Generate the Saliency Map:

- Take the absolute value of the gradient to create the saliency map:

$$S_{abs}(X) = \left|\frac{\partial y}{\partial X}\right|$$

- Visualize this saliency map to show which regions of the image are most influential in the prediction.

Step 4: Layer-wise Relevance Propagation (LRP)

4.1. Initialize Relevance at Output:

- The relevance of the output layer R_{output} is set to the class score:

$$R_{output} = y$$

4.2. Backpropagate Relevance Through Layers:

- Starting from the output layer, propagate the relevance backward through the network. For each layer l, the relevance is distributed to the previous layer $l-1$ based on the following rule:

$$R_l = \sum_i \frac{x_l^i \cdot W_{l,i}}{\sum_j x_l^i \cdot W_{l,i}} R_{l+1}$$

Where x_l^i is the activation of the i^{th} neuron at layer l, and $W_{l,i}$ is the weight connecting neuron i to the next layer.

4.3. Reach the Input Layer:

- Continue back propagating the relevance until reaching the input layer. The relevance scores at the input layer, R_{input} correspond to the contribution of each pixel to the final output:

$$R_{input} = \sum_i \frac{x_{input}^i \cdot W_{input,i}}{\sum_j x_{input}^j \cdot W_{input,j}} R_{output}$$

4.4. Visualize Relevance:

- The relevance scores R_{input} can be visualized as a heatmap to show the parts of the input image most responsible for the CNN's decision.

Step 5: Final Visualization

5.1. After applying Grad-CAM, Saliency Maps, and LRP, you will have several visual explanations:

5.1.1. Grad-CAM Heatmap: Shows which regions of the image are most relevant to the class prediction.

5.1.2. Saliency Map: Indicates which pixels in the image contribute most to the final prediction.

5.1.3. LRP Relevance Map: Provides insights into the contribution of each pixel by backpropagating relevance scores.

Step 6: Interpret the Results

6.1. Analyze the generated heatmaps, saliency maps, and relevance scores to understand the CNN's decision-making process. This helps identify the areas of the image that the CNN is focusing on, which can be crucial for understanding the model's predictions, especially in sensitive domains like medical image classification.

This algorithm provides a comprehensive approach to explain CNN predictions using several XAI methods. By combining Grad-CAM, saliency maps, and LRP, this framework helps to demystify how CNNs interpret and make decisions from input data. These methods contribute to model transparency and trust, which is critical in real-world applications like medical diagnosis and autonomous systems.

4. Experiment Results

4.1 Accuracy

Accuracy is the degree of closeness between a measurement and its true value. The formula for accuracy is:

$$\boldsymbol{Accuracy} = \frac{(\boldsymbol{truevalue} - \boldsymbol{measuredvalue})}{\boldsymbol{truevalue}} * \mathbf{100}$$

Dataset	SSA	GSCN-AD	Proposed ECNNF
100	62	58	92
200	73	65	96
300	78	71	90
400	82	78	98
500	86	80	94

Table 1.Comparison Table of Accuracy

The Comparison table 1 of Accuracy demonstrates the different values of existing SSA, GSCN-ADand Proposed ECNNF. While comparing the Existing algorithm and Proposed ECNNF, provides the better results. The existing algorithm values start from 62 to 86, 58 to 80 and Proposed ECNNFvalues starts from 90 to 98. The proposed method provides the great results.

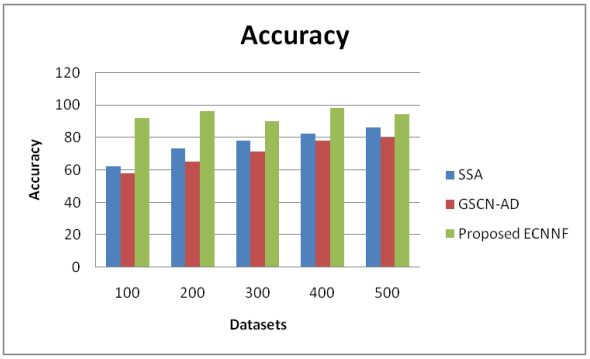

Figure 1 Comparison Chart of Accuracy

The Figure 1 Shows the comparison chart of Accuracy demonstrates the existing SSA, GSCN-AD and Proposed ECNNF. X axis denote the Dataset and y axis denotes the Accuracy. The Proposed ECNNF values are better than the existing algorithm. The existing algorithm values start from 62 to 86, 58 to 80 and Proposed ECNNF values starts from 90 to 98. The proposed method provides the great results.

4.2 Precision

Precision is a measure of how well a model can predict a value based on a given input.

$$\boldsymbol{Precision} = \frac{\boldsymbol{truepositive}}{(\boldsymbol{truepositive} + \boldsymbol{falsepositive})}$$

Dataset	SSA	GSCN-AD	Proposed ECNNF
100	74.12	69.63	95.67
200	78.69	72.82	93.26
300	82.12	76.54	98.21
400	84.41	78.63	97.58
500	86.94	81.72	91.87

Table 2 Comparison Table of Precision

The Comparison table 2 of Precision demonstrates the different values of existing SSA, GSCN-AD and Proposed ECNNF. While comparing the Existing algorithm and Proposed ECNNF, provides the better results. The existing algorithm values start from 74.12to 86.94, 69.63 to 81.72and Proposed ECNNF values starts from 91.87 to 98.21. The proposed method provides the great results.

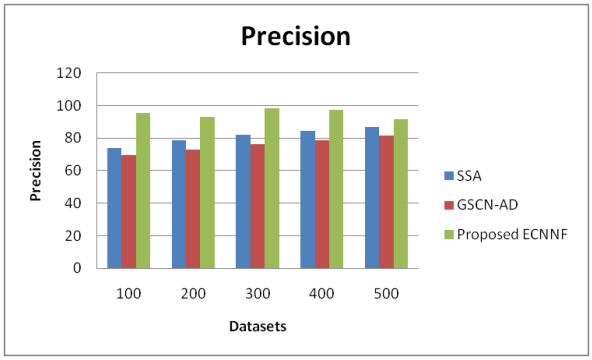

Figure 2 Comparison Chart of Precision

The Figure 2 Shows the comparison chart of Precision demonstrates the existing SSA, GSCN-AD and Proposed ECNNF. X axis denote the Dataset and y axis denotes the Precision ratio. The Proposed ECNNF values are better than the existing algorithm. The existing algorithm values start from 74.12to 86.94, 69.63 to 81.72and Proposed ECNNF values starts from 91.87 to 98.21. The proposed method provides the great results.

4.3 Recall

Recall is a measure of a model's ability to correctly identify positive examples from the test set:

$$\boldsymbol{Recall} = \frac{\boldsymbol{TruePositives}}{(\boldsymbol{TruePositives} + \boldsymbol{FalseNegatives})}$$

Dataset	SSA	GSCN-AD	Proposed ECNNF
100	0.74	0.70	0.89
200	0.76	0.72	0.92
300	0.78	0.67	0.94
400	0.84	0.75	0.96
500	0.86	0.81	0.98

Table 3 Comparison Table of Recall

The Comparison table 3 of Recall demonstrates the different values of existing SSA, GSCN-AD and Proposed ECNNF. While comparing the Existing algorithm and Proposed ECNNF, provides the better results. The existing algorithm values start from 0.74 to 0.86, 0.67 to 0.81 and Proposed ECNNF values starts from 0.89 to 0.98. The proposed method provides the great results.

Figure 3 Comparison Chart of Recall

The Figure 3 Shows the comparison chart of Recall demonstrates the existing SSA, GSCN-AD and Proposed ECNNF. X axis denote the Dataset and y axis denotes the Recall ratio. The Proposed ECNNF values are better than the existing algorithm. The existing algorithm values start from 0.74 to 0.86, 0.67 to 0.81 and Proposed ECNNF values starts from 0.89 to 0.98. The proposed method provides the great results.

4.4 F -Measure

F1-measure is a test's accuracy that combines precision and recall. It is calculated by taking the harmonic mean of precision and recall.

$$F1 - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$$

Dataset	SSA	GSCN-AD	Proposed ECNNF
100	0.88	0.82	0.99
200	0.87	0.81	0.96
300	0.85	0.75	0.94
400	0.83	0.70	0.92
500	0.78	0.67	0.90

Table 4 Comparison Table of F -Measure

The Comparison table 4 of F -Measure Values explains the different values of existing SSA, GSCN-AD and Proposed ECNNF. While comparing the Existing algorithm and Proposed ECNNF, provides the better results. The existing algorithm values start from 0.78to 0.88,0.67 to 0.82and Proposed ECNNF values starts from 0.90to 0.99. The proposed method provides the great results.

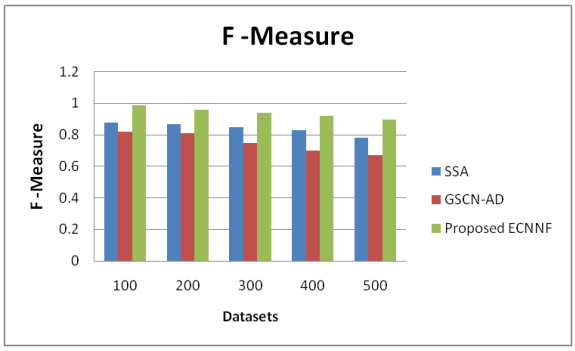

Figure 4 Comparison Chart of F -Measure

The Figure 4 Shows the comparison chart of F -Measure demonstrates the existing SSA, GSCN-AD and Proposed ECNNF. X axis denote the Dataset and y axis denotes the F -Measure ratio. The Proposed ECNNF values are better than the existing algorithm. The existing algorithm values start from 0.78to 0.88,0.67 to 0.82 and Proposed ECNNF values starts from 0.90to 0.99. The proposed method provides the great results.

5. Conclusion

In conclusion, the proposed Explainable Convolutional Neural Network Framework (ECNNF) integrates multiple XAI techniques, including Grad-CAM, Saliency Maps, and Layer-wise Relevance Propagation (LRP), to enhance the interpretability of CNN predictions. By visualizing activation maps, highlighting important features, and back propagating relevance scores, ECNNF provides comprehensive insights into how CNNs make decisions. This transparency is particularly crucial in sensitive applications like medical diagnosis and autonomous systems, where understanding model behavior fosters trust and ensures accountability. The methodology effectively identifies critical image regions influencing predictions, bridging the gap between model complexity and user comprehension. Future research should explore combining ECNNF with

multi-modal data and advanced segmentation techniques to further improve interpretability and robustness, ensuring its applicability across diverse real-world scenarios.

Reference

1. E. Jabason, M. O. Ahmad and M. N. S. Swamy, "Missing Structural and Clinical Features Imputation for Semi-supervised Alzheimer's Disease Classification using Stacked Sparse Autoencoder," 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), Cleveland, OH, USA, 2018, pp. 1-4, doi: 10.1109/BIOCAS.2018.8584844.
2. D. A. J and M. R, "A Novel Modelling Technique for Early Recognition and Classification of Alzheimer's disease," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 21-25, doi: 10.1109/ICSPC51351.2021.9451803.
3. A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta and A. Upadhya, "A CNN Model: Earlier Diagnosis and Classification of Alzheimer Disease using MRI," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 156-161, doi: 10.1109/ICOSEC49089.2020.9215402.
4. R. Ju, C. Hu, p. zhou and Q. Li, "Early Diagnosis of Alzheimer's Disease Based on Resting-State Brain Networks and Deep Learning," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 1, pp. 244-257, 1 Jan.-Feb. 2019, doi: 10.1109/TCBB.2017.2776910.
5. H. Padole, S. D. Joshi and T. K. Gandhi, "Early Detection of Alzheimer's Disease using Graph Signal Processing on Neuroimaging Data," 2018 2nd European Conference on Electrical Engineering and Computer Science (EECS), Bern, Switzerland, 2018, pp. 302-306, doi: 10.1109/EECS.2018.00062.
6. Mahendran, N.; Vincent, P.; Srinivasan, K.; Chang, C. Improving the Classification of Alzheimer's Disease Using Hybrid Gene Selection Pipeline and Deep Learning. Front. Genet. 2021, 12, 784814.
7. Saratxaga, C.L.; Albizuri, A.; Beristain, A.; Remesal, H.; Blesa, J.; Álvarez, I.; Moratal, D. MRI Deep Learning-Based Solution for Alzheimer's Disease Prediction. J. Pers. Med. 2021, 11, 902.
8. Battineni, G.; Chintalapudi, N.; Amenta, F. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). Inform. Med. Unlocked 2019, 16, 100200.
9. Kavitha, C.; Mani, V.; Srividhya, S.R.; Khalaf, O.I.; Romero, C.A.T. Early-stage Alzheimer's disease prediction using machine learning models. Front. Public Health 2022, 10, 3294.
10. Baglat, P.; Salehi, A.W.; Gupta, A.; Gupta, G. Multiple machine learning models for detection of Alzheimer's disease using oasis dataset. In Re-Imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation; Galliers, R.D., Newell, S., Eds.; Springer: Cham, Switzerland, 2020; pp. 614–622.
11. Basheer, S.; Bhatia, S.; Sakri, S.B. Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset. IEEE Access 2021, 9, 42449–42462.
12. Dhakal, S.; Adhikari, S.; Nepal, R.; Rai, P.; Shrestha, S.; Pokharel, S.; Acharya, S. Dementia prediction using machine learning. Procedia Comput. Sci. 2023, 219, 1297–1308.
13. A. B. Rabeh, F. Benzarti and H. Amiri, "CNN-SVM for prediction Alzheimer disease in early step," 2023 International Conference on Control, Automation and Diagnosis (ICCAD), Rome, Italy, 2023, pp. 1-6, doi: 10.1109/ICCAD57653.2023.10152440.
14. A. Kumar, V. Balaji, M. A. Chandrashekar, A. Dukkipati and S. Vadhiyar, "Graph Convolutional Neural Networks for Alzheimer's Classification with Transfer Learning and HPC Methods," 2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Lyon, France, 2022, pp. 186-195, doi: 10.1109/IPDPSW55747.2022.00043.
15. A. Yashodhar and S. Kini, "Comparison of Transfer Learning algorithm in predicting

Alzheimer's disease," 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS), Bangalore, India, 2023, pp. 183-187, doi: 10.1109/ICAECIS58353.2023.10170318

16. A.Raj, S. Bujare, A. Gorthi, J. Malik, A. Das and A. Kumar, "Alzheimer's Disease Recognition using CNN Model with EfficientNetV2," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), 2022, pp. 1-5, doi: 10.1109/ASIANCON55314.2022.9908834.
17. Andersen, E., Casteigne, B., Chapman, W. D., Creed, A., Foster, F., Lapins, A., … Sawyer, R. P. (2021). Diagnostic biomarkers in Alzheimer's disease. Biomarkers in Neuropsychiatry, 5, 100041. doi:10.1016/j.bionps.2021.100041
18. B. Tamm, R. Vandenberghe and H. Van Hamme, "Cross-Lingual Transfer Learning for Alzheimer's Detection from Spontaneous Speech," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-2, doi: 10.1109/ICASSP49357.2023.10096770.
19. B. N. Esi Nyarko, W. Bin, J. Zhou, G. K. Agordzo, J. Odoom and E. Koukoyi, "Comparative Analysis of AlexNet, Resnet-50, and Inception-V3 Models on Masked Face Recognition," 2022 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2022, pp. 337-343, doi: 10.1109/AIIoT54504.2022.9817327.
20. C. A. Ortiz Toro, N. Gutiérrez Sánchez, C. Gonzalo-Martín, R. Garrido García, A. Rodriguez González and E. Menasalvas Ruiz, "Radiomics Textural Features Extracted from Subcortical Structures of Grey Matter Probability for Alzheimer's Disease Detection," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), 2019, pp. 391-397, doi: 10.1109/CBMS.2019.00084.