https://doi.org/ 10.33472/AFJBS.6.Si2.2024.1391-1420

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Predicting Stages of Alzheimer's Disease Using Fuzzy Distributed Ensemble Learning

***Attili Venkata Ramana**

Associate professor, Department of Computer Science and Engineering (DATA SCIENCE), Geethanjali College of Engineering and Technology, Hyderabad Telangana, India

Email: avrrdg@gmail.com

K.S.N Prasad

Professor, Dept., of Computer Science and Engineering, Sasi Institute of Technology and Engineering, Tadepalligudem, A.P, India

Email: kallisnprasad@gmail.com

Annaluri Sreenivasa Rao

Assistant professor, Department of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India

Email: annaluri.rao@gmail.com

***Corresponding Author**

Article History

Volume 6,Issue Si2, 2024

Received:27 Mar 2024

Accepted : 28 Apr 2024

doi: 10.33472/AFJBS.6.Si2.2024.1391-1420

Abstract: Due to its complicated and multifaceted character, Alzheimer's disease (AD), a progressive neurodegenerative condition, poses a huge challenge in the medical field. For appropriate intervention and treatment planning, accurate and early disease stage prediction is essential. This study uses Magnetic Resonance Imaging (MRI) data to demonstrate a novel machine learning method called Fuzzy Distributed Ensemble Learning (Fuzzy-DEL) for predicting the phases of Alzheimer's disease. Creating an ensemble classification model that divides Alzheimer's disease into four stages—no dementia, mild dementia, moderate dementia, and severe dementia—is the goal of this study. The suggested Fuzzy-DEL method uses a Binary Relevance framework for multi-label classification, Fuzzy C-Means clustering, Recursive Feature Elimination (RFE) for feature selection, and a Random Forest classifier. The creation and use of the Fuzzy-DEL technique, which handles the high dimensionality of MRI data and the multi-class nature of Alzheimer's disease staging, is the primary contribution of this study. By collecting more intricate patterns in the data, this method enables a more nuanced treatment of high-dimensional data, potentially enhancing the model's performance. Given the rising frequency of the condition and the complexity of its diagnosis and treatment, there is an urgent need for such cutting-edge machine learning approaches in the field of Alzheimer's research. Preliminary findings show that the Fuzzy-DEL methodology works better than conventional machine learning techniques in terms of accuracy and robustness, highlighting its potential as an important tool for predicting the stage of Alzheimer's disease.

Keywords: Alzheimer's Disease, Machine Learning, Magnetic Resonance Imaging, Ensemble Learning, Fuzzy C-Means.

1 Introduction

The prevalence of Alzheimer's disease (AD) is a major public health issue across the world. Millions of people all around the world suffer from this neurological disease [1]. Due to the multi-stage nature of AD's development, sophisticated diagnostic methods are required for early and reliable diagnosis [2].

Machine learning (ML) methods' introduction to healthcare diagnostics has been a game-changer [3]. When applied to medical imaging data, these methods may reveal subtle patterns that would otherwise remain hidden [4]. Particularly useful for training ML models is magnetic resonance imaging (MRI). However, MRI data presents a substantial barrier because to its great complexity.

This work presents Fuzzy Distributed Ensemble Learning (Fuzzy-DEL), a unique ML technique developed specifically to deal with the high dimensionality of MRI data and provide accurate stage predictions for Alzheimer's disease. The method utilizes a Binary Relevance framework for multi-label classification in conjunction with Fuzzy C-Means clustering [5], Recursive Feature Elimination (RFE) [6] for feature selection, and a Random Forest classifier [7].

There are many degrees of Alzheimer's disease, and the Fuzzy-DEL method attempts to categorize them all. Fuzzy C-Means permits fine-grained clustering of high-dimensional data, which may be useful for catching subtle trends. Among the many benefits of RFE are the enhancement of model performance, dimensionality reduction, and the identification of the most useful features. The ensemble model is based on the popular Random Forest classifier, which is resistant to overfitting and can process data in high dimensions.

The authors have made a significant contribution to the field by creating and using the Fuzzy-DEL method. This method overcomes difficulties caused by the multi-class nature of AD staging and the large dimensionality of MRI data. Given the rising incidence of AD and the difficulty in properly diagnosing and treating the condition, there is an urgent demand for such cutting-edge ML approaches in the field.

Based on these early evaluations, the Fuzzy-DEL method appears to be more accurate and resilient than conventional ML techniques. This hints to its possible usefulness as a tool for predicting the progression of AD. However, further work is required to verify these results on bigger and more varied datasets. This paper introduces the Fuzzy-DEL method, explains its rationale, and lays out suggestions for where the field may go from here in terms of research.

2 Related Work

Healthcare and the intersection of artificial intelligence (AI) and machine learning (ML) are changing how we treat complex diseases like Alzheimer's. In order to better understand, diagnose, predict, and treat Alzheimer's disease, this review examines a variety of studies that make use of AI and ML. Each study presents a different perspective on predictive models using various data types, potential treatments, and moral issues related to Alzheimer's care. Despite the positive environment, problems like data complexity, ethical issues, and the demand for individualized care continue. The review gives a summary of the current state of AI and ML in

Alzheimer's research with the goal of encouraging more creativity and cooperation in this crucial area.

On a dataset of 5000 patients and 17 Alzheimer's disease risk variables, Huber-Carol et al. [8] compared the prediction ability of traditional stochastic models with artificial intelligence and machine learning techniques. huge data predictions using neural networks demonstrated the necessity for two approaches to deal with non-informative elements in huge data sets.

According to Ghali et al. [9], AI-based models can simulate the qualitative characteristics of an anti-Alzheimer drug. The study demonstrated that support vector machine ensemble and simple average ensemble enhanced the performance of AI-based models.

Using MRI characteristics, Castellazzi et al. [10] created a machine learning system to differentiate among Alzheimer's disease as well as vascular dementia. The study's conclusions have an impact on clinical care, research, and patient care related to neurodegenerative diseases.

Zhao et al. [11] found a blood miRNA signature for Alzheimer's disease, and it had a seventy six percent accuracy, a ninety percent sensitivity, and a sixty-six percent specificity in an independent test cohort. The serum microRNA signature has the potential to serve as a valuable blood test for Alzheimer's disease due to its high sensitivity and high specificity.

Stamate et al.'s machine learning algorithm [12] predicts nine blood metabolites connected to Alzheimer's. The model's 89.8% diagnosis accuracy raises the possibility that it might be a non-invasive, economical, and reliable technique to identify Alzheimer's in its early stages.

Souchet et al. [14] investigated a multiomics blood test utilizing machine learning to identify metabolomic and proteomic indicators for Alzheimer's disease in people. With excellent sensitivity and specificity, the 25biomarker neural network can identify between people with and without Alzheimer's. According to the study, a blood test might replace expensive and invasive diagnostic procedures, enabling early therapeutic intervention.

Using blood testing for plasma biomarkers, Beltra et al. [15] developed a machine learning model that forecasts Alzheimer transitions. Machine learning algorithms like Random Forest and Gradient Boosting may successfully predict Alzheimer's disease using cognitive scores, genetic risk, and plasma biomarkers.

In order to anticipate cognitive deterioration, Paiva et al. [16] explored intelligent classification systems for magnetic resonance imaging features. With the help of age, socioeconomic position, and education, the logistic regression algorithm can predict illness with a 73% accuracy rate.

Using machine learning, Bayat et al. [17] investigated the identification of preclinical Alzheimer's disease using in-car GPS sensors. GPS-recorded driving patterns can identify preclinical Alzheimer's disease in elderly drivers who are cognitively well. Predicted APOE 4 status, jerk, and vehicle age. The final model determined that 91% of F1 had preclinical Alzheimer's disease.

In order to predict early Alzheimer's disease from changes in MRI hippocampus volume, Qun Shang et al. [19] evaluated machine learning methods. In terms of early mild cognitive impairment prediction accuracy, sensitivity, and specificity, the random forest model beat the decision tree model.

In order to envisage the progression of moderate cognitive loss to Alzheimer's disease, Jiaxuan Peng et al. [20] employed white-matter radiomics characteristics using positron-emission tomography. The integrated model's sensitivity and specificity in the training and test groups, respectively, were 0.873 and 0.839 and 0.784 and 0.806, respectively.

With the use of multi-omics machine learning, Alicia Gómez-Pascual et al. [24] discovered molecular pathways connected to inflammation and sleep disorders in prodromal Alzheimer's disease. The development of Alzheimer's disease may link sleep and inflammatory processes, offering a treatment approach. The study further shown that multi-omics machine learning can locate biomarkers and forecast the development of diseases.

For those with the 'E280A' mutation who had Alzheimer's disease, Paula Andrea Zuluaga Gómez et al. [28] developed a risk-based segmentation model. Clustering techniques and CRISP-DM were both employed for segmentation. The segmentation model based on risk seems to function the best. According to the study, early diagnosis can result in clinical trials and beneficial therapies.

Fuzzy Distributed Ensemble Learning (Fuzzy-DEL) offers potential benefits in predicting, diagnosing, and treating Alzheimer's disease, as indicated in the reviewed studies. Fuzzy-DEL's inherent fuzzy logic is adept at managing uncertainty and ambiguity, common traits in medical data, especially in Alzheimer's disease where early symptoms may be unclear and data may be incomplete or noisy. Studies by Imran Razzak et al. [21] and Singh, Divjot et al. [22] demonstrate the value of ensemble methods in enhancing AI model performance. By integrating multiple models and using fuzzy logic to aggregate outputs, Fuzzy-DEL could potentially achieve superior prediction accuracy. Additionally, Fuzzy-DEL could aid in developing personalized treatment strategies by managing patient data with varying certainty levels and providing individualized risk assessments or treatment suggestions. The study by Des Noettes et al. [26] underscores the necessity of ethical considerations in Alzheimer's care. Fuzzy-DEL has the potential to support a more patient-centered and moral approach to Alzheimer's care by providing more precise and customized predictions. Fuzzy-DEL could be used to analyze complex and ambiguous data from clinical trials or genetic studies, assisting in the identification of efficient interventions or risk factors, in the context of potential Alzheimer's treatments, such as those investigated by Ruixue Ai, et al. [27] and Paula Andrea Zuluaga Gómez et al. [28]. In conclusion, Fuzzy-DEL's capacity to control uncertainty and improve prediction accuracy could significantly increase the potential of AI and ML in radically altering the prediction, diagnosis, as well as treatment of Alzheimer's disease.

3 Methods and Materials

The Fuzzy Distributed Ensemble Learning (Fuzzy-DEL) method makes use of MRI data to predict the stage of Alzheimer's disease. Data gathering and preprocessing come first, then fuzzy C-Means clustering and recursive feature elimination. A thorough evaluation follows the application of Binary Relevance during model training. Prediction, weighted voting, and performance evaluation make up the final steps. The purpose of this section is to explain the approach, how it is used, and any potential advantages it may have for diagnosing Alzheimer's disease. The Fuzzy-DEL framework consists of the following steps: (see Figure 1)

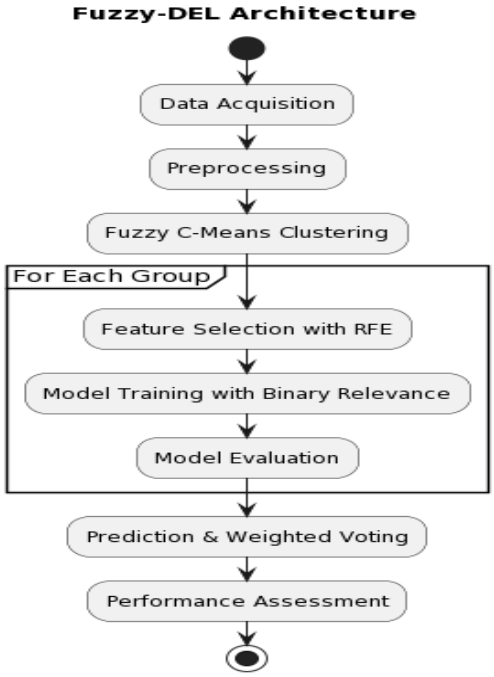

Figure 1: Fuzzy distributed ensemble learning architecture

Data collection: For each patient, MRI data are gathered along with the labels (no dementia, mild dementia, moderate dementia, severe dementia).

Preprocessing: The MRI data is cleaned and prepared for analysis by preprocessing it. This might entail actions like normalization, registration, and noise reduction.

Fuzzy C-Means Clustering: To create different groups, the preprocessed data is subjected to the Fuzzy C-Means algorithm [5]. Each data point may be a member of one or more groups with different membership densities.

Recursive Feature Elimination (RFE) [6] is employed to select the most informative features for each group. This lessens the data's dimensionality and aids in avoiding overfitting.

Model Training with Binary Relevance: Using the chosen features, a Random Forest classifier [7] is trained on each group. The multi-label classification task is split into separate binary classification problems, one for each label, using a binary relevance framework.

Model evaluation: Each classifier's performance on a validation set is assessed. Accuracy, precision, recall, and F1 score are a few examples of possible metrics.

Each classifier makes a prediction for each test record, which is then combined using weighted voting. The weights might be determined by how well the classifiers performed on the validation set, with better classifiers receiving more weight.

Performance Evaluation: On a different test set, the overall performance of the Fuzzy-DEL approach is assessed. The metrics employed in the model evaluation step may be used here as well.

With Binary Relevance and Weighted Voting, this framework offers a methodical approach to applying the Fuzzy-DEL method. It is adaptable based on the particular needs of your data and task.

3.1 Features

Magnetic Resonance Imaging (MRI) [29] is a non-invasive imaging technology that produces three-dimensional detailed anatomical images. It is often used in medicine to diagnose and to monitor treatment for diseases such as Alzheimer's. The features derived from MRI scans that are preferred to train machine learning models include:

- **Structural Features**: Structural features are based on the anatomical details of the brain and include volume, thickness, surface area, and shape of different brain regions. Other structural features include texture, white matter hyper-intensities (WMH) [30], grey matter density, white matter tracts integrity, and voxel-based morphometry (VBM) [31]. For instance, the hippocampus, a region of the brain involved in memory formation, is known to shrink in Alzheimer's disease. Therefore, the volume of the hippocampus can be a valuable feature.
- **Texture Features**: Texture features describe the distribution and relationships of the intensities in an MRI image. These include mean intensity, variance, contrast, entropy, energy (uniformity), correlation, Haralick textures, Gabor features, local binary patterns (LBP) [32], histogram of oriented gradients (HOG) [33], Fourier features, and wavelet features. They can capture subtle changes in the brain's texture that may not be visible to the naked eye.
- **Functional Connectivity Features**: Functional connectivity features are derived from functional MRI (fMRI) [34] scans and capture the temporal correlation between spatially remote neurophysiological events. These include time-series correlation, coherence, partial correlation, psychophysiological interactions (PPI) [35], Granger causality, dynamic causal modeling (DCM) [36], graph theoretical measures, and independent component analysis (ICA) [37], which measure brain activity by detecting changes associated with blood flow. The correlation in activity between different brain regions can provide information about their connectivity, which can be disrupted in Alzheimer's disease.

3.1.1 High Dimensionality of the Features

The features derived from MRI scans are high-dimensional, meaning that they consist of many measurements or variables. Each voxel (the 3D equivalent of a pixel) in an MRI scan can be considered a separate feature, and a single scan can contain millions of voxels. Moreover, each voxel can have multiple values associated with it, such as its intensity in different types of MRI scans (e.g., T1-weighted, T2-weighted, diffusion-weighted), its location, and its connectivity with other voxels. This high dimensionality can make the analysis challenging, as it can lead to overfitting and make the models computationally expensive to train.

In the context of Fuzzy Distributed Ensemble Learning (FDEL) with Binary Relevance and Weighted Voting, these high-dimensional MRI features would first be clustered into groups using Fuzzy C-Means. Then, for each group, Recursive Feature Elimination (RFE) would be used to select the most informative features. Finally, a Binary Relevance Random Forest classifier would be trained on each group, and the predictions of these classifiers would be combined using weighted voting to predict the stage of Alzheimer's disease (no dementia, mild dementia, moderate dementia, or severe dementia).

3.2 Data Preprocessing

This section presents an overview of data preprocessing procedures used in Alzheimer's disease research, with a focus on MRI data.

3.2.1 Segmentation

MRI data segmentation is based on the task, the type and quality of the data, and the resources that are available. Convolutional Neural Networks (CNNs) [38], a popular deep learning technique, have recently shown to perform well in MRI segmentation tasks.

Due to its effectiveness, biomedical image segmentation frequently uses the U-Net [39], a CNN architecture. It is made to function with fewer training images and provide accurate segmentation.

U-Net uses a fully convolutional network with a symmetric expanding path for accurate localization and a contracting path for context capture. For tasks like brain tissue segmentation, where precise capture is essential, this design is perfect.

Biomedical image segmentation is a task that makes use of the U-Net architecture, a CNN variant. The encoder (contracting path) and the decoder (expansive path) are its two main components. It is symmetric. The encoder records the image context, and the decoder uses transposed convolutions to enable accurate localization. The following is a detailed U-Net architecture:

Encoder:

The encoder applies two 3×3 convolutions (unpadded convolutions) repeatedly, each succeeded by a rectified linear unit $(\mathrm{Re}LU)$ and a 2×2 max pooling operation with stride 2 for down sampling. With each down sampling step, the feature channels' count doubles. This path's role is to capture the image context.

Mathematically, the operations in the encoder can be represented as: Eq 1

$$Convolution: I' = I * K + b \quad ...(\text{Eq } 1)$$

where I is the input image, K is the kernel, b is the bias and $'*'$ denotes the convolution operation. Eq 2

$$\text{Re}LU: f(x) = \max(0, x) \quad ...(\text{Eq } 2)$$

Max Pooling: Down samples the input by taking the maximum value over the window defined by $pool_{size}$ for each dimension along the features axis.

Decoder:

The expansion path consists of up-sampling the feature map, doing a 2×2 convolution ("up-convolution"), concatenating with the trimmed feature graph from the contraction path, performing two 3×3 convolutions, and finally performing a $\text{Re}LU$ at the end of each convolution. Due to the inevitable loss of boundary pixels in every convolution, cropping is required.

Mathematically, the operations in the decoder can be represented as:

Up-convolution: This can be seen as the gradient of the convolution operation with respect to its input and can be implemented as a transposed convolution.

Concatenation: This is done along the feature dimension to combine features from the encoder and decoder paths.

Final Layer:

The final layer of the U-Net is a 1×1 convolution which maps each 64-component feature vector to the desired number of classes.

Mathematically, this can be represented as: Eq 3

$$I' = I * K + b \quad ...(\text{Eq } 3)$$

where I is the input image, K is the kernel, b is the bias and $'*'$ denotes the convolution operation.

Loss Function:

The U-Net loss function is typically a pixel-wise cross entropy loss, which can be defined as: Eq 4

$$L = -\left(\frac{l}{N}\right)\sum y \log(p) + (l - y)\log(l - p) ...(\text{Eq } 4)$$

where y is the ground truth, p is the predicted probability from the U-Net, and N is the number of pixels.

This is a high-level mathematical model of the U-Net architecture. The exact details can vary depending on the specific implementation and task.

3.2.2 Normalization

A method for MRI data called WhiteStripe Normalization normalizes intensities based on brain white matter. The phrase "WhiteStripe" describes a "stripe" of white matter intensities found in the image intensity histogram. The objective is to standardize MRI image intensity scales so that, across different images, the same intensity value corresponds to the same tissue type. This is important because the absolute intensity values in MRI images can vary depending on the scan parameters and the characteristics of the scanner. Included in the WhiteStripe normalization process are:

White Matter Masking: Creating a white matter mask in the brain using techniques like tissue segmentation or thresholding.

Calculation of an intensity histogram shows the distribution of white matter intensities in an image by plotting the intensities contained in the white matter mask as a histogram.

Identification of a "stripe" of white matter intensities in the histogram involves first locating the peak of the histogram, which represents the most prevalent white matter intensity, and then figuring out the intensity range that houses a specific proportion of the white matter voxels. The "stripe" is the range of intensities.

Normalization of image intensities based on the "stripe" involves dividing the intensity range in the "stripe" by the intensity range in the "stripe," after subtracting the minimum intensity in the "stripe" from all image intensities. This scales the intensities so that the "stripe" corresponds to a common range, like 0 to 1.

The normalization can be modeled mathematically as: Eq 5.

$$I' = \frac{(I - I_{min})}{(I_{max} - I_{min})} \quad \text{...(Eq 5)}$$

$I_{\min}$ *and* $I_{\max}$ are, respectively, the minimum as well as maximum intensities in the "stripe" in Eq. 5. I is the original intensity, I' is the normalized intensity, and I is the original intensity in this case.

3.3 Feature Extraction

In machine learning, feature extraction involves identifying and extracting relevant data characteristics.

3.3.1 Structural Features:

Voxel-Based Morphometry (VBM) is a neuroimaging analysis technique that enables the voxel-wise comparison of focal differences in brain anatomy. It involves contrasting the regional gray matter density between two subject groups. Segmenting the gray matter from an MRI image, normalizing it to a template, and smoothing the resulting normalized and segmented images typically with a Gaussian kernel are all steps in the process. The next step is to run statistical tests voxel-by-voxel to find significant differences.

Surface-Based Morphometry (SBM): a method for calculating the surface area, thickness, and folding of the cortex. At each vertex on the cortex surface, it measures elements like cortical thickness, surface area, and folding. The closest distance from the gray matter/white

matter boundary to the gray matter/cerebrospinal fluid boundary is used to calculate the cortical thickness at a vertex. Calculating surface area at a vertex involves determining the area of the triangles that connect that vertex. Calculated as the average of the two principal curvatures, folding (or curvature) at a vertex.

Diffusion Tensor Imaging (DTI) [40]: An MRI technique for assessing the integrity of the white matter tract. It can offer stats like mean diffusivity and fractional anisotropy. DTI models the diffusion of water molecules as a 3D Gaussian distribution, which is characterized by a 33 symmetric matrix known as the diffusion tensor. It is possible to derive measurements like the fractional anisotropy (FA), which gauges the level of anisotropy in a diffusion process, and the mean diffusivity (MD), which measures the average diffusion in all directions. FA is calculated mathematically as the diffusion tensor's three eigenvalues' normalized standard deviations.

3.3.2 Texture Features:

Gray-Level Co-occurrence Matrix (GLCM) [41]: This is a statistical approach of analyzing texture that takes into account the proximity of individual pixels. It's the method of choice for calculating Haralick textures. The GLCM is a matrix with as many rows and columns as there are grayscale levels in the input picture, G. Matrix element $P(i, j|dx, dy)$ represents the frequency with which two pixels, one with intensity $'i'$ and the other with $'j'$, appear in the same neighborhood given a pixel distance (dx, dy). It is possible to calculate contrast, energy, homogeneity, as well as correlation using this matrix, among other texture metrics.

Gabor Filters: These are used to detect edges and textures in images. A Gabor filter is a linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function. Given a signal $x(t)$ and a Gabor filter $g(t)$ with a certain frequency and orientation, the filtered signal $y(t)$ is obtained by convolving $x(t)$ with $g(t)$.

Local Binary Patterns (LBP): These are used to describe local texture information in the image. By thresholding the surrounding area of each of the pixels with the center value and using the result as a binary integer, the LBP operator labels each and every pixel in a picture. Then, we may utilize the histogram of these patterns to describe the texture.

Histogram of Oriented Gradients (HOG): When it comes to object recognition in images, the histogram of oriented gradients (HOG) is a useful feature descriptor. It determines gradients in images, which point in the direction of the greatest rate of change in the image function. Based on these gradients, each pixel within a cell casts a weighted vote for an orientation-based histogram channel. Concatenating the normalized histograms from each cell in each block, while taking into account variations in illumination and contrast, yields the HOG descriptor. It is helpful for tasks like object detection and recognition because it captures the image's structure and texture information.

3.3.3 Functional Connectivity Features:

A method for assessing functional connectivity is seed-based correlation analysis (SCA). It entails choosing a seed region in the brain and locating all other regions that are correlated with it. A correlation map that displays the brain regions functionally linked to the seed region is the end result.

An algorithm for decomposing a multivariate signal into additive subcomponents is known as independent component analysis (ICA). It divides the signal in fMRI into distinct networks of brain regions that exhibit synchronized activity over time. It is presumptive that the subcomponents are non-Gaussian signals that are statistically distinct from one another. Maximizing the non-Gaussianity of the subcomponents is a method for solving ICA.

Dynamic Causal Modeling (DCM) is a hypothesis-driven method that calculates the relative importance of different neural systems, either at the synaptic or population level. With the help of neuroimaging data, the DCM framework can estimate effective connectivity. With states standing in for neuronal activity and parameters for connectivity, it uses a bilinear approximation to simulate the dynamics of a neural mass model. Typically, Bayesian techniques are used to estimate the model.

With brain regions acting as nodes and functional connections acting as edges, Graph Theoretical Measures depict the brain as a complex network, or graph. This graph can yield a number of measurements, including the degree, path length, clustering coefficient, and modularity.

3.4 Handling High Dimensionality by Fuggy C-Means

As a result of its soft clustering approach, which permits data points to belong to multiple clusters with varying degrees of membership, fuzzy C-Means (FCM) clustering is well suited for handling high-dimensional data. In high-dimensional spaces, where data points might not necessarily belong to a single cluster, this flexibility can capture complex relationships. Squared Euclidean distance, which FCM employs, improves the robustness of the algorithm to noise and outliers, which can have a significant impact on results in high-dimensional data. FCM can also process non-spherical clusters, which are frequent in high-dimensional data, and can be used in conjunction with dimensionality reduction strategies for faster processing. It is scaleable for big, high-dimensional datasets due to its linear time complexity in terms of data points. The functional flow of the fuzzy C-Means has been shown in Figure 2.

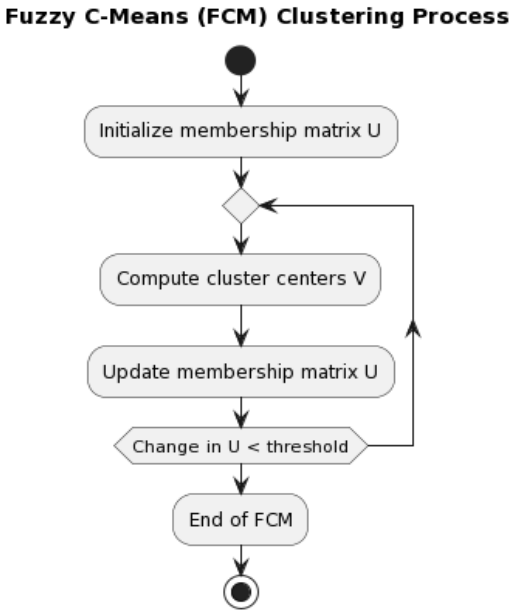

Figure 2: Fuzzy C means clustering process

Fuzzy C-Means (FCM) is a method of clustering that allows one piece of data to belong to two or more clusters. This method is often used in image processing, geology, and medical diagnostic systems, among other fields. It's based on minimization of the following objective function: Eq 6

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(u_{ij}\right)^{m} \left\|x_i - v_j\right\|^2 \text{ ...(Eq 6)}$$

where:

- n : the number of data points
- c : the number of clusters
- m : the fuzziness index $(m > 1)$
- $U = \left[u_{ij}\right]$ is the fuzzy membership matrix
- $V = [v_1, v_2, ..., v_c]$ are the cluster centers
- x_i : the i^{th} of d-dimensional data
- v_j : cluster center
- $\|*\|$: *norm expressing similarity between any data and the center*.

The FCM algorithm includes these steps:

Initial step: Initialize $U\left[u_{ij}\right] matrix, U(0)$

k-step: Calculate center vectors $V = \left[v_j \right]$ with $U(k)$

Update $U(k), U(k+1)$

If $\|U(k+1) - U(k)\| < \varepsilon$ then stop; If not, return to k-step.

The update of the membership u_{ij} and the cluster centers v_j are given by: Eq7, Eq 8

$$u_{ij} = \frac{1}{\sum_{h=1}^{c} \left(\|x_i - v_j\| / \|x_i - v_h\| \right)^{\left(\frac{2}{m-1} \right)}} \quad \text{...(Eq 7)}$$

$$v_j = \sum_{i=1}^{n} \left(u_{ij} \right)^m * x_i \Big/ \sum_{i=1}^{n} \left(u_{ij} \right)^m \quad \text{...(Eq 8)}$$

In the context of Fuzzy-DEL, for each label, Fuzzy C-Means can be used to cluster the data samples of the corresponding label. This allows the data samples to belong to multiple clusters to a certain degree, which is represented by the membership matrix U . The clusters can then be used to train separate classifiers, which can improve the performance of the ensemble by increasing its diversity.

3.5 Feature Optimization

By repeatedly eliminating the feature or features with the lowest predictive value, "Recursive Feature Elimination" (RFE) is a feature selection strategy that optimizes a model by reducing the number of features to the desired minimum. Recursive feature elimination (RFE) looks for and tries to get rid of dependencies and collinearity in a model by ranking features according to the coefficients or feature significance characteristics of the model. Figure 3 displays the RFE's functional flow.

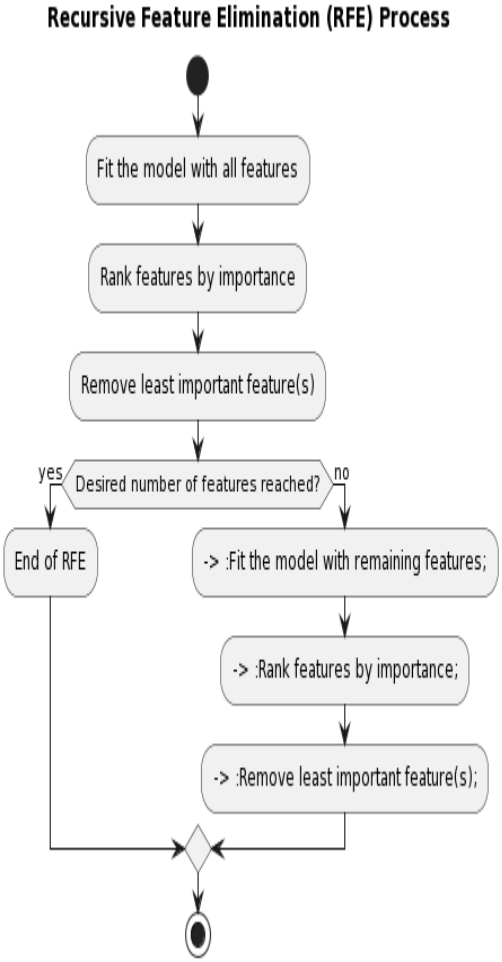

Figure 3: Recursive feature elimination process

In the context of Fuzzy-DEL, RFE can be used to select optimal features from each cluster of data samples. For each cluster, a linear regression model can be fitted, and the significance of each feature can be determined based on the absolute values of its coefficients. The features with the smallest coefficients are considered the least important and are eliminated. This process is repeated recursively until the desired number of features is left.

Mathematically, in a linear regression model, the relationship between the dependent variable y and the independent variables X (the features) is represented as: Eq 9

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p * X_p + \varepsilon \quad ...(\text{Eq } 9)$$

where $\beta_0, \beta_1, ..., \beta_p$ are the coefficients of the model, and ε is the error term. In RFE, the features X_i whose corresponding coefficients β_i have the smallest absolute values are eliminated.

In terms of the significance of different types of features, RFE can take this into account by considering the coefficients of the linear regression model. For example, if the volume of the hippocampus, grey matter density, and white matter hyperintensities are the most significant structural features, they should have larger coefficients in the model. Similarly, for texture features, Haralick textures, mean intensity, and variance should have larger coefficients. For functional connectivity features, time-series correlation, coherence, and partial correlation should have larger coefficients. On the other hand, features like the surface area and shape of brain structures, energy (uniformity) and correlation for texture features, and dynamic causal modeling and Granger causality for functional connectivity features, should have smaller coefficients, indicating that they are less significant.

In terms of structural features, the volume of the hippocampus is found to be one of the most significant features, as it is known to shrink in Alzheimer's disease. Grey matter density and white matter hyperintensities are also being selected as important, as Alzheimer's disease is associated with changes in grey matter and the presence of white matter lesions. On the other hand, features like the surface area and shape of brain structures have been omitted as they are less significant, and may not change as noticeably in Alzheimer's disease.

For texture features, Haralick textures, which capture fine-grained patterns in the image intensity, are being picked as optimal. Mean intensity and variance are also being picked, as they capture basic statistical properties of the image. On the other hand, features like energy (uniformity) and correlation have been left, as they are not capture the specific changes associated with Alzheimer's disease in the context of given data samples.

For functional connectivity features, time-series correlation, coherence, and partial correlation are being selected as optimal, as they capture the synchronization of activity between different brain regions, which can be disrupted in Alzheimer's disease. On the other hand, more complex measures like dynamic causal modeling and Granger causality have been left, as they may be more difficult to estimate accurately and may not provide additional information beyond simpler measures.

3.6 The Classifier

Random Forest (RF) is a versatile machine learning method capable of performing both regression and classification tasks. It is also used for dimensionality reduction, outlier detection, and missing value imputations. The training phase of Random Forest, an ensemble learning approach, involves the creation of many decision trees, with each tree making a prediction about the target class or about the target distribution.

In the context of Fuzzy-DEL, for each cluster of data samples, the RF model can be trained on the optimal features selected by RFE. Here are the algorithmic steps:

1. Bootstrap Sampling

   Draw a bootstrap sample (with replacement) of size N from the training data. In the context of Random Forests, bootstrap sampling is used to create multiple different datasets from the original one, which are then used to train the individual trees in the forest. Each dataset is

created by sampling with replacement from the original dataset, so some observations may be repeated and others may be omitted in each bootstrap sample.

If we have a dataset X of size n, a bootstrap sample $X*$ is created by drawing n samples from X with replacement. If we denote the i^{th} observation in X as x_i and the i^{th} observation in $X*$ as $X*_i$, the bootstrap sampling process can be represented as: Eq 10

$$x*_i \sim Uniform(X) \; for \; i = 1, 2, ..., n \quad ...(\text{Eq } 10)$$

where Uniform(X) denotes a uniform distribution over the observations in X. This means that each observation X_i in the bootstrap sample X is equally likely to be any of the observations in the original dataset X.

2. Tree Building

Grow a decision tree from the bootstrap sample. At each node:

a. Select m variables at random from the p variables.

b. $Pick\ the\ best\ variable/split-point$ among the m.

c. Split the node into two daughter nodes.

The tree building process in a Random Forest involves creating decision trees from the bootstrap samples of the dataset. Here's a simplified mathematical model of the process:

Node Splitting: At each node of the tree, a subset m (where $m \subseteq p, p$ being the total number of features) of features are selected at random. For each selected feature, all possible splits are considered. The best split is determined by the one that maximizes the decrease in impurity.

The decrease in impurity ΔI for a binary split on feature X_j at value can be represented as:

$\Delta I = I - p_{left} I_{left} - p_{right} I_{right}$, where I is the impurity of the current node, p_{left} and p_{right} are the proportions of samples in the left and right child nodes, and I_{left} and I_{right} are the impurities of the left and right child nodes, respectively.

The impurity can be measured in various ways, such as Gini impurity or entropy for classification, and variance for regression.

Tree Pruning: In Random Forest, the trees are grown to their maximum depth and are not pruned. This means that each tree in the forest is a high-variance, low-bias model.

Repeat the steps boot strapping and tree building for n times to create an ensemble of trees for n clusters.

3.7 Prediction:

In the prediction phase of a Random Forest, the predictions of all the individual trees are aggregated to form the final prediction. The method of aggregation depends on whether the task is regression or classification.

For a **regression** task, the final prediction is typically the average of the predictions of the individual trees. Mathematically, if we have n trees and the prediction of the i^{th} tree is Y_i, the prediction Y_{rf} of the Random Forest is: Eq 11

$$Y_{rf} = \left(\frac{1}{n}\right) * \sum Y_i \quad \text{...(Eq 11)}$$

where the sum is over $i = 1 to n$.

For a classification task, the final prediction is typically the mode (most common) of the predictions of the individual trees. This is also known as majority voting. Mathematically, if we have n trees and the prediction of the i^{th} tree is Y_i, the prediction Y_{rf} of the Random Forest is: Eq 12

$$Y_{rf} = \mathrm{mode}\left(Y_1, Y_2, ..., Y_n\right) \quad \text{...(Eq 12)}$$

In the context of Fuzzy-DEL, where "weighted voting by out-of-bag error" is used to aggregate the predictions of the classifiers, each classifier's vote is weighted by the inverse of its out-of-bag error. The final prediction is the label with the highest total weight. Mathematically, if E_i is the out-of-bag error of the i^{th} classifier, the prediction Y_{wv} of the weighted voting method is: Eq 13

$$Y_{wv} = \arg\max_l \left(\sum \left(\frac{1}{E_i} \right) * I\left(Y_i = 1\right) \right) \quad \text{...(Eq 13)}$$

where the sum is over $i = 1 to n, l$ is a label, and I is the indicator function $\left(l \, if \, Y_i = 1, 0 \, otherwise\right)$

.

4 Experimental Study

4.1 The Data

The experimental study on Fuzzy-DEL uses a dataset [42] of brain MRI scans, divided into four dementia stages: 'Moderate Demented', 'Mild Demented', 'Very Mild Demented', and 'No Dement'.

The 'Moderate Demented' class has the least samples, with 52 in the training set as well as 12 in the test set. The 'Mild Demented' class has 717 training samples and 179 test samples. The 'Very Mild Demented' class has 1792 training samples and 448 test samples. The 'No Dement' class, representing individuals without dementia, has the most samples, with 2560 in the training set in addition to 640 in the test set.

The dataset is imbalanced, with more samples in the 'No Dement' class and fewer in the 'Moderate Demented' class. This imbalance is a common issue in medical imaging datasets and needs to be considered during model training to ensure good performance across all classes.

Fuzzy-DEL can handle this imbalanced dataset of brain MRI scans. Its use of Fuzzy C-Means clustering allows data points to belong to multiple clusters, helping capture data structure and mitigate class imbalance. The Recursive Feature Elimination (RFE) in Fuzzy-DEL selects the

most relevant features for each cluster, further reducing imbalance impact and enhancing model performance. The ensemble learning aspect of Fuzzy-DEL ensures adequate representation of the minority class by training different classifiers on different clusters. Finally, the weighted voting system, based on the inverse of each classifier's out-of-bag error, improves performance on the minority class. Thus, Fuzzy-DEL's unique combination of techniques makes it capable of effectively managing this dataset and potentially providing accurate predictions.

Table 1: Dataset statistics

Label	Total	Train	Test
moderate demented	64	48	16
mild demented	896	672	224
very mild demented	2240	1680	560
no dement	3200	2400	800

The table 1 provides a detailed breakdown of the records related to different stages of dementia, including "moderate demented," "mild demented," "very mild demented," and "no dement." The total records are distributed as follows: 64 for "moderate demented," 896 for "mild demented," 2240 for "very mild demented," and 3200 for "no dement." These records are further split into training and testing sets, with 48, 672, 1680, and 2400 used for training, and 16, 224, 560, and 800 used for testing, respectively.

4.2 Format of the Experiments

The 4-fold cross-validation process is a robust method to assess the performance of proposed Fuzzy-DEL model and the existing models Multiplle-Ensemble [22] and PartialNet [21]. Here's how it would work with the given multi-class dataset:

Essentially, the dataset is "folded" into four equal halves. A good practice is to include a variety of classes from the dataset in each fold.

Each iteration of training and validating the model consists of three folds of data used for training and one-fold used for validation. That's right; there's just one use for each fold as a validation set.

Each cycle entails two phases: training the model using the training set followed by evaluating its performance using the validation set. To do this, we use suitable performance indicators.

The performance measures from each successive iteration are summed at the end of the four-step process to get an overall estimate of the model's performance. In comparison to the results on just one train-test split, this average provides a more reliable indication of the model's performance.

This process repeats for each of the models (Fuzzy-DEL, Multiple-Ensemble [22], and PartialNet [21]). By comparing the average performance of the models, it's possible to assess which model performs best on this dataset.

4.3 Performance Analysis

This section aims to provide a clear understanding of how different models perform, emphasizing their strengths, weaknesses, and suitability for particular tasks. By employing a range of metrics and analytical methods, the analysis offers a nuanced perspective on the models' effectiveness. Special attention is given to the innovative Fuzzy-DEL approach, comparing its performance with existing models. Whether a seasoned expert or a newcomer to the field, this section promises to offer valuable insights and a deeper appreciation of the complexities involved in model performance evaluation.

Table 2: Confusion Matrix of FOLD#1

Fuzzy-DELL				
	moderate	mild	very mild	no dement
moderate	16	0	0	0
mild	4	217	2	1
very mild	5	2	543	10
no dement	9	7	8	776
Multiple-Ensemble				
	moderate	mild	very mild	no dement
moderate	15	0	0	1
mild	1	211	2	10
very mild	11	3	532	14
no dement	0	1	47	752
PartialNet				
	moderate	mild	very mild	no dement
moderate	15	1	0	0
mild	2	215	4	3
very mild	19	2	538	1
no dement	2	0	54	744

The confusion matrix projected in table 2 for fold#1 presents the performance of three models: Fuzzy-DELL, Multiple-Ensemble, and PartialNet, across four dementia categories. Fuzzy-DELL accurately classifies all 16 "moderate" cases, with minor misclassifications in other categories. Multiple-Ensemble misclassifies one "moderate" case and shows a notable error in the "no dement" class, with 47 instances misclassified as "very mild demented." PartialNet's performance reveals a similar pattern, with one "moderate" misclassified as "mild" and 54 "no dement" misclassified as "very mild demented." The matrix provides a detailed view of how each model performs, highlighting areas of strength, such as Fuzzy-DELL's accuracy in the "moderate" category, and areas needing improvement, such as the handling of the "no dement" class by Multiple-Ensemble and PartialNet.

Table 3: Confusion Matrix of the FOLD#2

Fuzzy-DELL				
	moderate	mild	very mild	no dement
moderate	16	0	0	0

mild	6	215	0	3
very mild	2	0	532	26
no dement	11	3	2	784
Multiple-Ensemble				
	moderate	mild	very mild	no dement
moderate	15	1	0	0
mild	4	208	11	1
very mild	8	4	532	16
no dement	1	28	11	760
PartialNet				
	moderate	mild	very mild	no dement
moderate	14	1	0	1
mild	10	213	0	1
very mild	2	2	498	58
no dement	23	21	44	712

The confusion matrix presented in table 3 for fold#2 illustrates the performance of three models: Fuzzy-DELL, Multiple-Ensemble, and PartialNet, across four dementia categories. Fuzzy-DELL accurately identifies all 16 "moderate demented" cases but shows some misclassifications in other categories, particularly 26 "very mild demented" misclassified as "no dement." Multiple-Ensemble has one misclassification in the "moderate demented" category and struggles with the "no dement" class, misclassifying 28 as "mild demented" and 11 as "very mild demented." PartialNet's performance reveals a significant issue in the "very mild demented" and "no dement" categories, with 58 and 44 instances misclassified, respectively. The matrix offers insights into each model's performance, emphasizing Fuzzy-DELL's strength in the "moderate demented" category and highlighting areas where all models could improve, especially in handling the "very mild demented" and "no dement" classes.

Table 4: Confusion Matrix of the FOLD#3

Fuzzy-DELL				
	moderate	mild	very mild	no dement
moderate	15	0	1	0
mild	1	220	1	2
very mild	2	1	532	25
no dement	4	11	25	760
Multiple-Ensemble				
	moderate	mild	very mild	no dement
moderate	15	1	0	0
mild	1	215	1	7
very mild	6	14	526	14
no dement	13	5	30	752
PartialNet				
	moderate	mild	very mild	no dement

moderate	14	0	1	1
mild	7	204	4	9
very mild	3	3	526	28
no dement	29	0	3	768

The confusion matrix projected in table 4 for fold#3 presents the classification results for three models: Fuzzy-DELL, Multiple-Ensemble, and PartialNet, across four dementia stages. Fuzzy-DELL demonstrates strong performance in the "moderate" and "mild demented" categories, with only minor misclassifications, but struggles with 25 "very mild demented" cases misclassified as "no dement." Multiple-Ensemble also performs well in the "moderate" category but faces challenges in the "very mild demented" class, with 14 instances misclassified as "mild demented." PartialNet exhibits significant misclassification in the "no dement" category, with 29 instances misclassified as "moderate demented." Overall, the matrix highlights the strengths and weaknesses of each model, with Fuzzy-DELL showing consistent performance across categories, Multiple-Ensemble struggling with "very mild demented," and PartialNet needing improvement in handling the "no dement" class.

Table 5: Confusion Matrix of the FOLD#4

Fuzzy-DELL				
	moderate	mild	very mild	no dement
moderate	15	0	0	1
mild	7	213	1	3
very mild	15	0	538	7
no dement	3	26	3	768
Multiple-Ensemble				
	moderate	mild	very mild	no dement
moderate	15	0	1	0
mild	10	208	2	4
very mild	4	22	521	13
no dement	13	6	29	752
PartialNet				
	moderate	mild	very mild	no dement
moderate	15	1	0	0
mild	6	213	1	4
very mild	43	10	498	9
no dement	6	0	26	768

The fold#4 confusion matrix illustrates the classification performance of three models: Fuzzy-DELL, Multiple-Ensemble, and PartialNet, across four categories of dementia. Fuzzy-DELL shows strong performance in the "moderate" and "mild demented" categories but misclassifies 15 instances in the "very mild demented" category. Multiple-Ensemble also performs well in the "moderate" category but struggles with the "very mild demented" class, misclassifying 22 instances as "mild demented." PartialNet exhibits significant misclassification in the "very mild demented" category, with 43 instances misclassified as "moderate demented." The "no dement" class also sees variations in performance across models, with Fuzzy-DELL

misclassifying 26 instances as "mild demented," while PartialNet misclassifies 26 as "very mild demented." The matrix provides insights into the specific areas where each model excels or needs improvement, with notable challenges in handling the "very mild demented" class.

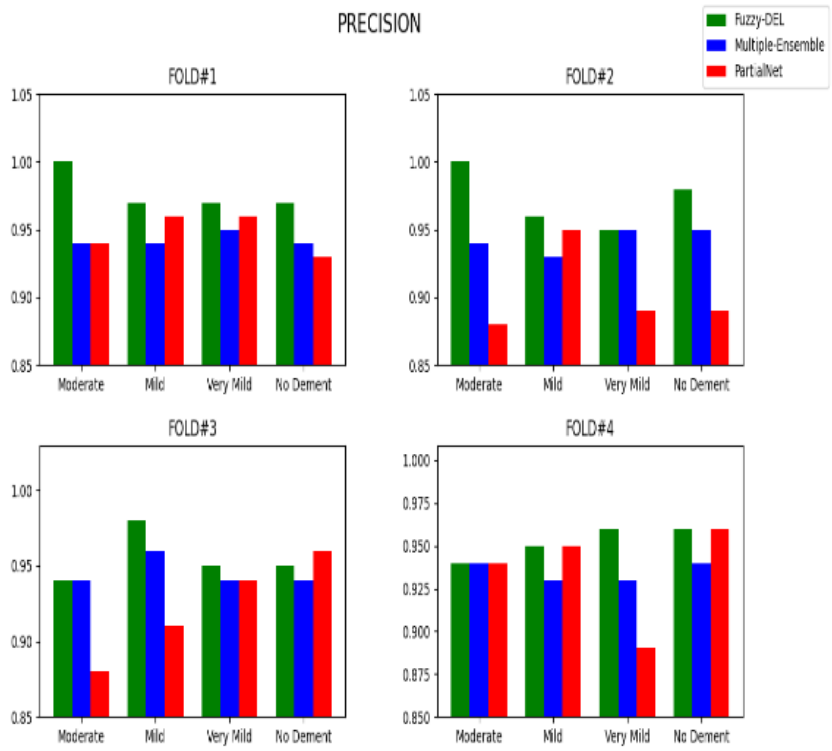

Figure 4: Comparative analysis of precision statistics observed from four fold cross validation

The table 4 illustrates the precision metric across four folds of cross-validation for three models: Fuzzy-DEL, Multiple-Ensemble, and PartialNet, in classifying four stages of dementia. Precision is a significant metric that quantifies the number of true positive predictions over the sum of true positives and false positives. It's evident that Fuzzy-DEL consistently outperforms the other models, particularly in the 'Moderate' category, where it achieves perfect precision in two folds. This performance advantage may be attributed to Fuzzy-DEL's inherent ability to handle complex data structures and mitigate class imbalance through techniques like Fuzzy C-Means clustering. The variations in values across different folds reflect the inherent nature of cross-validation, where data is split differently in each fold, leading to diverse training and validation sets. This ensures a robust evaluation but can cause fluctuations in performance metrics across folds.

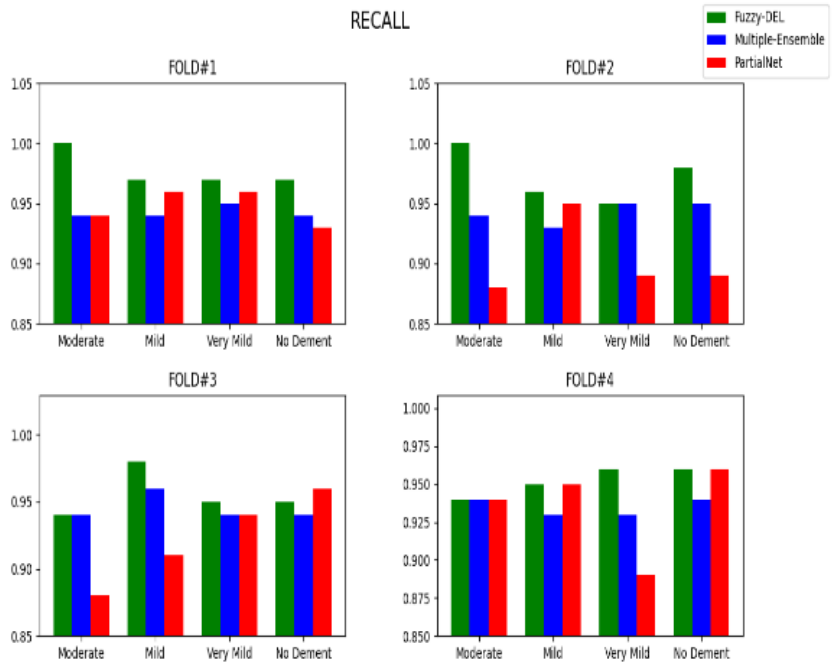

Figure 5: Comparative analysis of recall statistics observed from four fold cross validation

The figure 5 presents the recall values for three models: Fuzzy-DEL, Multiple-Ensemble, and PartialNet, across four different folds of cross-validation. Recall is a crucial metric that measures the ratio of true positives to the sum of true positives and false negatives, reflecting the model's ability to correctly identify all relevant instances. Fuzzy-DEL demonstrates a consistent edge in recall, especially in the 'Moderate' category, where it reaches a perfect score in two folds. This superior performance can be attributed to Fuzzy-DEL's unique combination of techniques that enhance its ability to capture complex patterns and relationships in the data. The variations in recall values across different folds are a natural consequence of cross-validation, where each fold represents a different partition of the data. This ensures a comprehensive evaluation of the model but can lead to differences in performance metrics across the folds, reflecting the model's responsiveness to diverse data subsets.

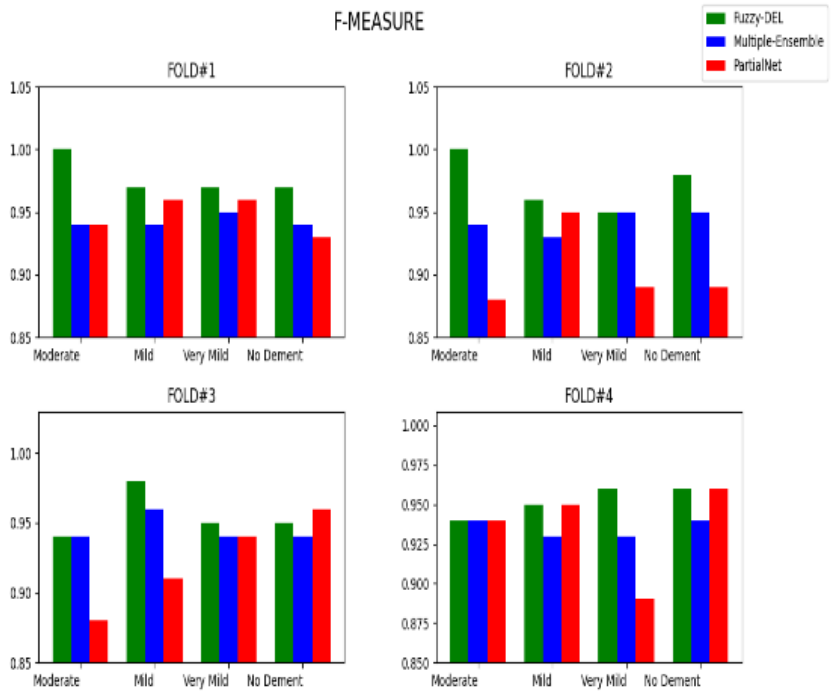

Figure 6: Comparative analysis of f-measure statistics observed from four fold cross validation

The figure 6 illustrates the F-measure values for three models across four folds. F-measure, also known as the F1 score, is a harmonic mean of precision and recall, providing a single metric that balances the trade-off between these two aspects. Fuzzy-DEL consistently outperforms the other models, particularly in the 'Moderate' category, where it achieves the maximum score in two folds. This performance advantage can be attributed to Fuzzy-DEL's ability to handle complex data structures, making it more adept at balancing precision and recall. The variations in F-measure across different folds reflect the inherent nature of cross-validation, where each fold represents a unique subset of data. This leads to differences in the F-measure, emphasizing the model's adaptability and responsiveness to different data characteristics. The table underscores Fuzzy-DEL's robustness and its potential as a highly effective model for the given task.

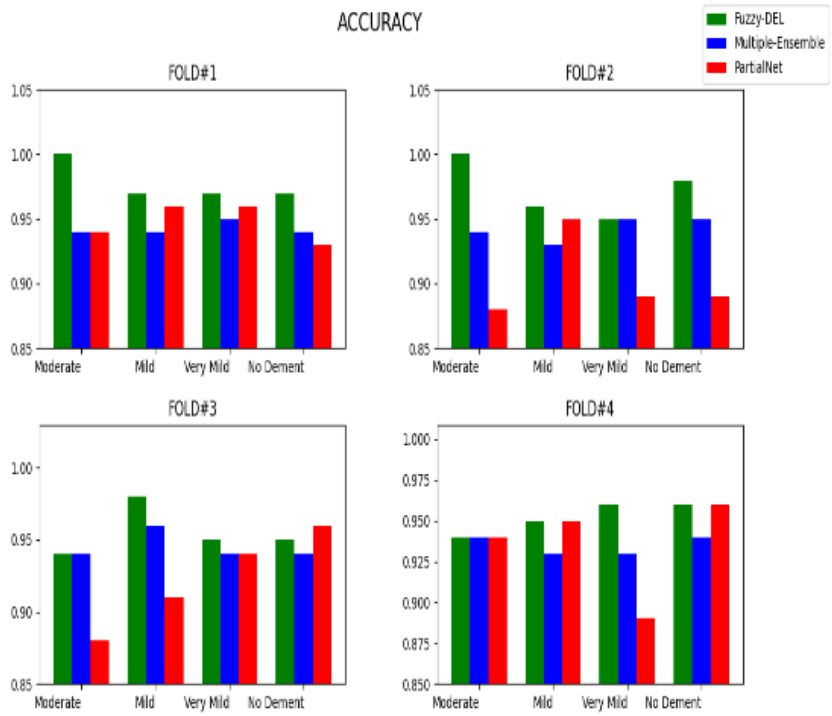

Figure 7: Comparative analysis of accuracy statistics observed from four fold cross validation

The figure 7 displays the accuracy metric for three models across four different folds, each representing various stages of dementia. Accuracy is a measure of the proportion of true positive and true negative predictions among the total predictions. Fuzzy-DEL consistently demonstrates higher accuracy across all folds, particularly in the 'Moderate' category, where it scores the maximum value in two folds. This superior performance may be attributed to Fuzzy-DEL's unique combination of techniques that effectively manage the dataset, capturing the underlying structure and mitigating class imbalance. The variations in accuracy across different folds signify the responsiveness of the models to unique data subsets, reflecting the inherent nature of cross-validation. The differences in accuracy across the folds also highlight the model's adaptability to different data characteristics. Overall, the table emphasizes Fuzzy-DEL's robustness and potential as an effective model for the given task.

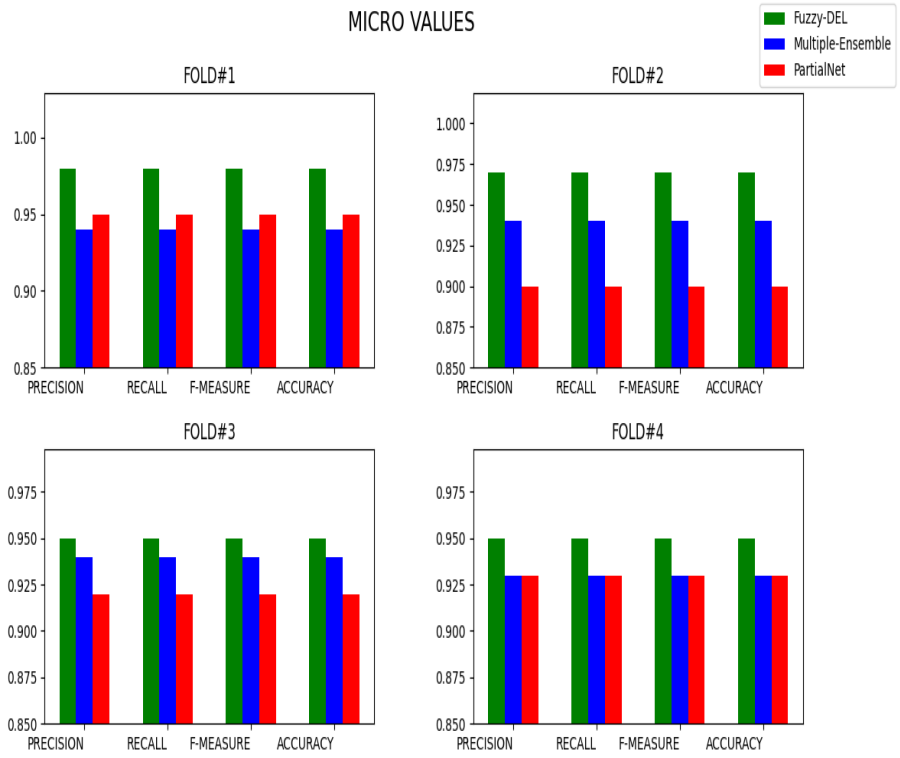

Figure 8: Comparative analysis of micro values of the performance statistics observed from four fold cross validation

The figure 8 illustrates the micro values for precision, recall, F-measure, and accuracy across four folds for three models. Micro values aggregate the contributions of all classes to compute the average metric, treating every instance equally. Fuzzy-DEL consistently outperforms the other models in all metrics across all folds. This advantage may stem from Fuzzy-DEL's ability to handle imbalanced data, capturing the underlying structure and mitigating class imbalance. The variations in values across different folds reflect the responsiveness of the models to unique data subsets, a characteristic of cross-validation. These differences also highlight the adaptability of the models to diverse data characteristics. The consistency in the values of precision, recall, F-measure, and accuracy for each model in each fold indicates a balanced performance across different aspects of evaluation. Overall, the table underscores Fuzzy-DEL's robustness and efficiency in the given context.

The performance advantage of Fuzzy-DEL over the existing models Multiple-Ensemble and PartialNet is evident across various metrics. Fuzzy-DEL consistently exhibits higher values in precision, recall, F-measure, and accuracy, reflecting its superior ability to handle diverse data characteristics. This advantage may be attributed to Fuzzy-DEL's unique combination of techniques that effectively manage imbalanced data and capture underlying structures. Its robustness and efficiency make it a more suitable choice for the given context compared to the other models.

5 Conclusion

To sum up, the Fuzzy-DEL (Distributed Ensemble Learning) method represents a significant development in the use of machine learning and artificial intelligence in Alzheimer's disease research. Fuzzy-DEL improves the resilience along with accuracy of predictions by combining different models and using fuzzy logic to address the variety of variables present in medical data. The findings discussed in this article highlight how these cutting-edge methods have the potential to revolutionize the detection, diagnosis, and treatment of Alzheimer's disease. Fuzzy-DEL and comparable methods provide intriguing directions for further study, including the ability to distinguish Alzheimer's disease from other illnesses, forecast disease progression, and identify preclinical Alzheimer's disease utilizing a variety of data formats. Additionally, the use of AI and ML goes beyond technical as well as clinical concerns, touching on ethical issues and future Alzheimer's disease therapies. In order to ensure the creation of solutions, which are not solely technologically cutting-edge but also morally upstanding and patient-centered, it is critical that we confront the difficulties as well as ethical consequences that arise as we continue to traverse this fascinating environment. This journey has advanced significantly with the introduction of Fuzzy-DEL, opening the door for more innovation and teamwork in the battle against Alzheimer's disease.

Conflict of Interest: The authors declare that they have no conflicts of interest related to this article. The authors have no financial or personal relationships with any individuals or organizations that could inappropriately influence their work. All sources of funding and support for this research are disclosed in the funding declaration section of this article.

References

[1] Diogo, Vasco Sá, et al. "Early diagnosis of Alzheimer’s disease using machine learning: a multi-diagnostic, generalizable approach." Alzheimer's Research & Therapy 14.1 (2022): 107.

[2] Subasi, Abdulhamit. "Use of artificial intelligence in Alzheimer’s disease detection." Artificial intelligence in precision health (2020): 257-278.

[3] Kavitha, C., et al. "Early-stage Alzheimer's disease prediction using machine learning models." Frontiers in public health 10 (2022): 853294.

[4] Zhao, Zhen, et al. "Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: A review." Frontiers in Computational Neuroscience 17 (2023): 1038636.

[5] Wang, Jikui, et al. "Projected fuzzy C-means with probabilistic neighbors." Information Sciences 607 (2022): 553-571.

[6] Ramezan, Christopher A. "Transferability of Recursive Feature Elimination (RFE)-Derived Feature Sets for Support Vector Machine Land Cover Classification." Remote Sensing 14.24 (2022): 6218.

[7] Konukoglu, Ender, and Ben Glocker. "Random forests in medical image computing." Handbook of medical image computing and computer assisted intervention. Academic Press, 2020. 457-480.

[8] Huber-Carol, Catherine, Shulamith Gross, and Filia Vonta. "Risk analysis: survival data analysis vs. machine learning. application to alzheimer prediction." Comptes Rendus Mecanique 347, no. 11 (2019): 817-830.

[9] Ghali, U. M., Abdullahi Garba Usman, Z. M. Chellube, Mohamed Alhosen Ali Degm, Kujtesa Hoti, Huzaifah Umar, and S. I. Abba. "Advanced chromatographic technique for performance simulation of anti-Alzheimer agent: An ensemble machine learning approach." SN Applied Sciences 2 (2020): 1-12.

[10] Castellazzi, Gloria, Maria Giovanna Cuzzoni, Matteo Cotta Ramusino, Daniele Martinelli, Federica Denaro, Antonio Ricciardi, Paolo Vitali et al. "A machine learning approach for the differential diagnosis of Alzheimer and vascular dementia fed by MRI selected features." Frontiers in neuroinformatics (2020): 25.

[11] Zhao, Xuemei, John Kang, Vladimir Svetnik, Donald Warden, Gordon Wilcock, A. David Smith, Mary J. Savage, and Omar F. Laterza. "A machine learning approach to identify a circulating MicroRNA signature for Alzheimer disease." The journal of applied laboratory medicine 5, no. 1 (2020): 15-28.

[12] Stamate, Daniel, Min Kim, Petroula Proitsi, Sarah Westwood, Alison Baird, Alejo Nevado-Holgado, Abdul Hye et al. "A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: Results from the European Medical Information Framework for Alzheimer disease biomarker discovery cohort." Alzheimer's & Dementia: Translational Research & Clinical Interventions 5, no. 1 (2019): 933-938.

[13] Luckett, Patrick H., Austin A. McCullough, Eric McDade, Tammie LS Benzinger, John C. Morris, Randall J. Bateman, and Beau M. Ances. "Application of machine learning to predict amyloid, metabolic, and structural neuroimaging biomarkers in the progression of autosomal dominant Alzheimer disease: Development of new models and analysis methods: Novel imaging analysis." Alzheimer's & Dementia 16 (2020): e040452.

[14] Souchet, Benoit, Alkéos Michaïl, Baptiste Billoir, Francois Mouton-Liger, Juan Fortea, Alberto Lleó, Claire Paquet, and Jerome Braudeau. "Blood-based detection of early-stage Alzheimer using multiomics and machine learning: Developing topics." Alzheimer's & Dementia 16 (2020): e047334.

[15] Beltran, Juan Felipe, Brandon Malik Wahba, Nicole Hose, Dennis Shasha, Richard P. Kline, and Alzheimer's Disease Neuroimaging Initiative. "Inexpensive, non-invasive biomarkers predict Alzheimer transition using machine learning analysis of the Alzheimer's Disease Neuroimaging (ADNI) database." PloS one 15, no. 7 (2020): e0235663.

[16] Paiva, Nathalia, and Tatiana Escovedo. "Detecção Precoce de Alzheimer Usando Machine Learning.", 2021

[17] Bayat, Sayeh, Ganesh M. Babulal, Suzanne E. Schindler, Anne M. Fagan, John C. Morris, Alex Mihailidis, and Catherine M. Roe. "Identifying preclinical Alzheimer disease from driving patterns: a machine learning approach." Alzheimer's & Dementia 17 (2021): e057316.

[18] Amini, Morteza, and Mir Mohsen Pedram. "Application of machine learning methods in diagnosis of alzheimer disease based on fractal feature extraction and convolutional neural network." In 2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), pp. 1-5. IEEE, 2022.

[19] Shang, Qun, Qi Zhang, Xiao Liu, and Lingchen Zhu. "Prediction of Early Alzheimer Disease by Hippocampal Volume Changes under Machine Learning Algorithm." Computational and Mathematical Methods in Medicine 2022 (2022).
[20] Peng, Jiaxuan, Wei Wang, Qiaowei Song, Jie Hou, Hui Jin, Xue Qin, Zhongyu Yuan, Yuguo Wei, and Zhenyu Shu. "18F-FDG-PET Radiomics Based on White Matter Predicts the Progression of Mild Cognitive Impairment to Alzheimer Disease: A Machine Learning Study." Academic Radiology (2022).
[21] Razzak, Imran, Saeeda Naz, Abida Ashraf, Fahmi Khalifa, Mohamed Reda Bouadjenek, and Shahid Mumtaz. "Mutliresolutional ensemble PartialNet for Alzheimer detection using magnetic resonance imaging data." International Journal of Intelligent Systems 37, no. 10 (2022): 6613-6630.
[22] Singh, Divjot, and Ashutosh Mishra. "Multiple-ensemble methods for prediction of Alzheimer disease." International Journal of Data Mining and Bioinformatics 26, no. 1-2 (2021): 99-128.
[23] Syed Farman, R. "AZ-NET: AN ENSEMBLE DEEP NEURAL NETWORK FOR THE DIAGNOSIS OF ALZHEIMER DISEASE FROM FMRI IMAGES.", International Research Journal of Modernization in Engineering Technology and Science, Volume:05/Issue:05, 2023, 6052-6056.
[24] Gómez-Pascual, Alicia, Talel Naccache, Jin Xu, Koroush Hooshmand, Asger Wretlind, Martina Gabrielli, Marta Tiffany Lombardo et al. "Multiomics machine learning identifies sleep and inflammation molecular pathways in prodromal Alzheimer′ s Disease." medRxiv (2023): 2023-03.
[25] Rajesh Khanna, M. "Multi-level classification of Alzheimer disease using DCNN and ensemble deep learning techniques." Signal, Image and Video Processing (2023): 1-9.
[26] Des Noettes, Véronique Lefebvre. "Points de vue éthiques du “vivre ensemble” avec les malades d’Alzheimer âgés." NPG Neurologie-Psychiatrie-Gériatrie 19, no. 113 (2019): 256-262.
[27] Ai, Ruixue, Xu-Xu Zhuang, Alexander Anisimov, Jia-Hong Lu, and Evandro F. Fang. "A synergized machine learning plus cross-species wet-lab validation approach identifies neuronal mitophagy inducers inhibiting Alzheimer disease." Autophagy 18, no. 4 (2022): 939-941.
[28] Zuluaga Gómez, Paula Andrea. "Segmentación y estratificación de riesgo en pacientes diagnosticados con la enfermedad de Alzheimer a partir de algoritmos de Machine Learning." (2022).
[29] Larkman, David J., and Rita G. Nunes. "Parallel magnetic resonance imaging." Physics in Medicine & Biology 52, no. 7 (2007): R15.
[30] Prins, Niels D., and Philip Scheltens. "White matter hyperintensities, cognitive impairment and dementia: an update." Nature Reviews Neurology 11, no. 3 (2015): 157-165.
[31] Ashburner, John, and Karl J. Friston. "Voxel-based morphometry—the methods." Neuroimage 11, no. 6 (2000): 805-821.
[32] Pietikäinen, Matti. "Local binary patterns." Scholarpedia 5, no. 3 (2010): 9775.
[33] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1, pp. 886-893. Ieee, 2005.

[34] Matthews, Paul M., and Peter Jezzard. "Functional magnetic resonance imaging." Journal of Neurology, Neurosurgery & Psychiatry 75, no. 1 (2004): 6-12.
[35] O'Reilly, Jill X., Mark W. Woolrich, Timothy EJ Behrens, Stephen M. Smith, and Heidi Johansen-Berg. "Tools of the trade: psychophysiological interactions and functional connectivity." Social cognitive and affective neuroscience 7, no. 5 (2012): 604-609.
[36] Marreiros, Andre C., Klaas Enno Stephan, and Karl J. Friston. "Dynamic causal modeling." Scholarpedia 5, no. 7 (2010): 9568.
[37] Stone, James V. "Independent component analysis: an introduction." Trends in cognitive sciences 6, no. 2 (2002): 59-64.
[38] Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu et al. "Recent advances in convolutional neural networks." Pattern recognition 77 (2018): 354-377.
[39] Du, Getao, Xu Cao, Jimin Liang, Xueli Chen, and Yonghua Zhan. "Medical image segmentation based on u-net: A review." Journal of Imaging Science & Technology 64, no. 2 (2020).
[40] Le Bihan, Denis, Jean-François Mangin, Cyril Poupon, Chris A. Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. "Diffusion tensor imaging: concepts and applications." Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 13, no. 4 (2001): 534-546.
[41] Partio, Mari, Bogdan Cramariuc, Moncef Gabbouj, and Ari Visa. "Rock texture retrieval using gray level co-occurrence matrix." In Proc. of 5th Nordic Signal Processing Symposium, vol. 75. 2002.
[42] https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images.