Jurnal Ilmu dan Teknologi Kesehatan
https://ejurnal.poltekkesjakarta3.ac.id/index.php/jitek/
Volume 13, Issue 2, Page 145 – 162 DOI: 10.32668/jitek.v13i2.2381
e-ISSN: 2338-9109

A Hybrid Approach using Fuzzy Inference System and Deep Learning for Alzheimer's Disease Classification on MRI Images

1st Bagas Wibowo[1]

2nd Andy Maulana Yusuf[2*]

3rd Bintang Vieshe Mone[3]

4th Sabrina Adinda Sari[4]

[1,2,3,4]School of Computing, Telkom University, City, Bandung, West Java, Indonesia

[*]email: *andymaulanayusuf@telkomuniversity.ac.id*

Keywords:
Alzheimer's Disease Detection
Fuzzy Inference System
Deep Learning
Hybrid FCIE
MRI Classification

Abstract

Early detection of Alzheimer's disease using brain MRI image data can substantially improve clinical intervention and patient management. Our study evaluates the performance of an Alzheimer's classification system based on Fuzzy Inference Systems (FIS), specifically for the Mamdani and Sugeno models, in identifying four patient categories: (1) Non-Dementia, (2) Very Mild Dementia, (3) Mild Dementia, and (4) Moderate Dementia. In addition, this study compares the classification performance and computational efficiency of several deep learning architectures, including a traditional CNN (VGG16), a modern model (EfficientNet-B0), and a hybrid Fuzzy Convolutional Inference Engine (FCIE) that integrates CNN-based feature extraction with fuzzy logic reasoning. The dataset used consists of normalized and augmented Alzheimer's MRI images, and each model was trained and validated using a 70%:15%:15% split for training, validation, and testing. Experimental results show that the Mamdani and Sugeno FIS models achieve validation accuracies of about 32% and 35%, respectively, which highlights the limitations of pure texture-based features in capturing complex classification patterns. In contrast, VGG16 and EfficientNet-B0 produced validation accuracies of 82.81% and 85.22%, respectively, with AUC values of 0.95 and 0.96, respectively. However, when both schemes were combined into a hybrid model FCIE achieved the highest validation accuracy of 98.03% and AUC of 0.99. Comparative analysis of metrics, including precision, recall, F1-score, AUC, and training duration, shows a clear trade-off between accuracy and computational efficiency. This study recommends the FCIE model for clinical applications requiring high diagnostic accuracy, while EfficientNet-B0 is suggested for medical environments with moderate GPU resource constraints.

Received: 5 Feb 2026
Accepted: 18 Jun 2026
Published: 18 Jun 2026

INTRODUCTION

Alzheimer's disease is a progressive neurodegenerative disorder that affects millions of individuals worldwide. This disease is primarily characterized by a decline in cognitive function, memory, and behavior, which ultimately reduces the patient's quality of life. According to data from Alzheimer's Disease International (2023), the number of individuals living with Alzheimer's and other dementias has surpassed 50 million globally, with numbers expected to increase as the aging population grows (1). Early detection then becomes crucial, as it allows for timely medical intervention, cognitive therapy, and family support to slow disease progression and improve patient outcomes (2).

Brain Magnetic Resonance Imaging (MRI) has become one of the primary modalities for Alzheimer's diagnosis (3), due to its ability to produce detailed anatomical images without exposure to ionizing radiation. Many studies have shown that texture and morphological patterns in brain structures, such as changes in hippocampal volume, cortical atrophy, and changes in signal intensity in white matter, can serve as early biomarkers of Alzheimer's disease (4,5). However, traditional texture feature extraction is often insufficient to capture complex spatial patterns, prompting the development of automated artificial intelligence (AI)-based approaches to classify MRI scans into Alzheimer's severity categories (6).

Fuzzy logic comes with the ability to handle uncertainty and provide rule-based explanations that are inherently interpretable (7). However, traditional fuzzy systems typically rely on manually engineered features, such as statistical or texture descriptors, which limit their ability to capture the richer spatial information present in raw MRI data (8). In contrast, deep learning approaches, particularly Convolutional Neural Networks (CNNs) such as VGG and Efficient Net (9,10), can automatically learn spatial features and generally achieve higher classification accuracy. Despite these advantages, CNN models are often considered “black boxes,” exhibit low interpretability, and require large datasets and substantial GPU resources (11). To bridge the gap between these two paradigms, the Fuzzy Convolutional Inference Engine (FCIE) has been proposed as a hybrid architecture that integrates fuzzy reasoning into convolutional layers, enabling effective spatial feature extraction while maintaining a level of fuzzy interpretability.

In this study, we analyze five models in depth, namely (1) Mamdani Fuzzy Inference System (FIS), (2) Sugeno FIS, (3) FCIE hybrid model, and (4-5) two CNN architectures (VGG and EfficientNet). The main goal is to compare accuracy, computational efficiency, model size, and interpretability, to identify the most suitable models for specific application scenarios, including clinics with limited hardware, referral hospitals, and hybrid diagnostic settings. Furthermore, this research seeks to identify the most suitable deployment scenarios for each model, ranging from resource-constrained clinical settings to high-performance referral hospitals requiring maximum diagnostic precision.

RESULTS AND DISCUSSION

Research in automated detection of Alzheimer's disease has progressed rapidly from manual feature-based methods to Deep Learning approaches and Hybrid systems. This section reviews recent literature regarding the evolution of the method.

In recent years, Convolutional Neural Networks (CNN) have become the gold standard in medical image analysis due to their ability to automatically extract hierarchical features. Islam and Zhang's research (12) developed a deep CNN model based on the Inception-V4 architecture for multiclass Alzheimer's classification on the OASIS dataset, which showed that the deep network was able to capture more complex features than traditional methods.

Further studies conducted by Helaly et al. (13) by comparing a simple CNN architecture built from scratch with a Transfer Learning model such as VGG19. Their research on the ADNI dataset showed that the fine-tuned VGG19 was able to achieve an accuracy of up to 97%, outperforming the scratch model (95.17%), even though the scratch model had the advantage of better computational efficiency.

El-Geneedy et al. (14) extended this evaluation by comparing their shallow CNN against various SOTA architectures such as DenseNet121, ResNet50, and EfficientNet-B7. Their findings are interesting, while their shallow CNN achieved very high accuracy (99.68%) on a given dataset, some complex models such as EfficientNet-B7 actually showed suboptimal performance (48.68%) in their experimental setup, highlighting that model complexity does not always correlate with accuracy on limited medical datasets.

Although Deep Learning offers high accuracy, this method is often criticized as a "black box" and vulnerable to data uncertainty. Tanveer et al. (15) in their comprehensive survey emphasized that neuroimaging data often contain noise, artifacts, and imprecise boundaries between brain networks. They argue that purely deterministic models like standard CNNs struggle to handle this uncertainty. Therefore, the integration of Fuzzy Logic is crucial due to its ability to model ambiguity and provide better interpretation in this case Explainable AI in clinical diagnosis.

To bridge the gap between CNN's feature extraction capabilities and Fuzzy reasoning capabilities, researchers started developing Hybrid architectures. Korablyov et al. (16) proposed a hybrid Neuro-Fuzzy model where the CNN module acts as a feature extractor which is then processed by the Neuro-Fuzzy Classifier (NFC). This approach has been shown to improve the quality of object recognition while reducing the computational cost of training.

This trend continues in the latest medical applications. Abdkhaleq (17) demonstrated the effectiveness of hybrid architecture on medical diagnosis (CT-Scan case study) by combining CNN (ResNet50) for features, PCA for dimensionality reduction, and ANFIS (Adaptive Neuro-Fuzzy Inference System) for classification. This hybrid model achieves 98.1% accuracy, outperforming the conventional CNN+SVM method, and offers better decision transparency.

This study positions itself among these studies by conducting a head-to-head comparison between pure FIS with Mamdani and Sugeno methods, pure Deep Learning such as VGG16 and EfficientNet, and Hybrid FCIE approaches specifically for Alzheimer's cases, to identify the best trade-off between diagnostic accuracy and computational efficiency.

METHODS

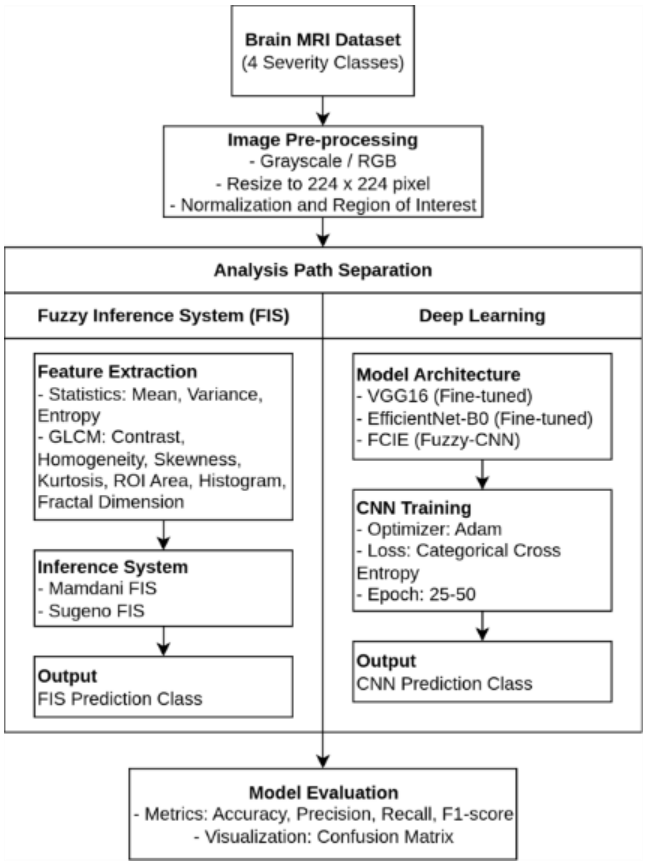

Figure 1. Piplene Research Methodology

As illustrated in Figure 1, the research methodology is structured into four distinct stages to ensure a comprehensive evaluation of the proposed models. First, the Data Preprocessing stage involves categorizing the dataset

into four distinct categories based on disease severity, followed by image resizing, normalization, and data augmentation to enhance model generalization. Second, the Fuzzy Inference System (FIS) development is implemented, beginning with feature extraction and followed by the implementation of two reasoning models, Mamdani and Sugeno. Third, simultaneously with the fuzzy modeling, we train and optimize Deep Learning architectures, specifically VGG16 and EfficientNet-B0, alongside our proposed Hybrid Fuzzy Convolutional Inference Engine (FCIE), incorporating rigorous hyperparameter tuning. Finally, a comprehensive Performance Evaluation is conducted using standard metrics, such as Accuracy, Precision, Recall, and F1-Score, to facilitate a comparative analysis of diagnostic effectiveness and computational efficiency across all models.

A. Dataset

This study utilized the Augmented Alzheimer's MRI Dataset sourced from Kaggle (18) . The dataset includes grayscale T1-weighted brain MRI images with a native resolution of 256×256 pixels, categorized into four classes based on the Clinical Dementia Rating (CDR), (1) Non-Demented (ND), (2) Very Mild Demented (VMD), (3) Mild Demented (MD), and (4) Moderate Demented (MoD). A representative sample of the dataset is represented in Figure 2.

The dataset is termed "Augmented" as it includes images that have experienced prior geometric transformations, resulting in a diverse repository suitable for training deep neural networks. For this study, the dataset was divided into three subsets, 70% for training, 15% for validation, and 15% for testing. This partition was executed using a stratified sampling strategy to maintain the class distribution across all subsets. Next, data shuffling was applied prior to each training epoch to mitigate bias and enhance model generalization.

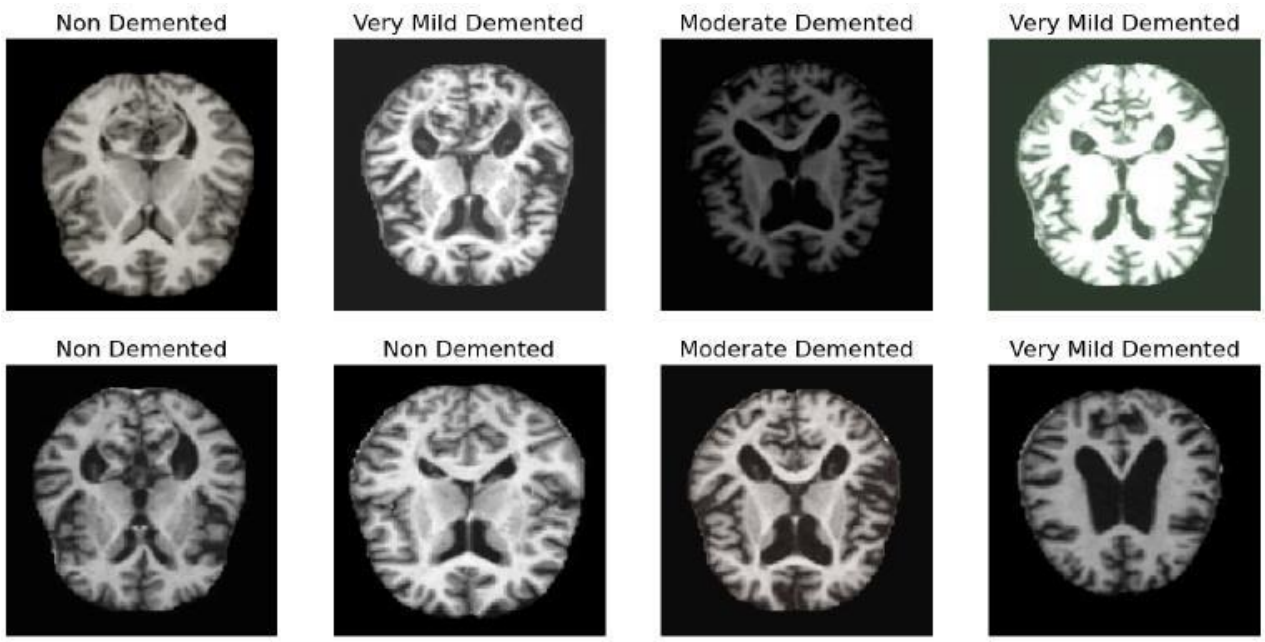

Figure 2. Example data from the dataset we used in this study.

B. Data Preprocessing

To ensure compatibility with the proposed architectures and optimize computational efficiency, a two-stage preprocessing pipeline was implemented, (1) Resizing and Normalization, followed by (2) Model-Specific Data Augmentation.

B.1. Resize and Normalization

Initially, all input images were resized from their native resolution to 224×224 pixels. This dimension was selected to satisfy the input layer specifications of the VGG16, EfficientNet-B0, and the proposed FCIE architectures. Additionally, downscaling the resolution serves to reduce computational load and accelerate convergence during training. Following resizing, intensity normalization was applied by scaling pixel values to the range (0-1). This step is critical for maintaining numerical stability, preventing vanishing or exploding gradients, and ensuring consistency with the pre-trained weights from ImageNet utilized in the deep learning models (19).

B.2. Data Augmentation

To mitigate overfitting and improve the robustness of the deep learning models, on-the-fly data augmentation was implemented using the Keras ImageDataGenerator. It is important to note that augmentation was explicitly excluded for the FIS models (Mamdani and Sugeno). Since fuzzy logic classifiers rely on static texture statistics (e.g., GLCM features), introducing geometric distortions could alter the underlying feature distribution and degrade performance (20). In other words, we do this stage specifically for Deep Learning and Hybrid models. For FCIE, VGG16, and EfficientNet-B0 models, we applied several augmentation techniques, including random rotation (±10° for FCIE, ±15° for VGG16 and EfficientNet), random flip (50% horizontal flip for all models, 30% vertical flip applied only to VGG16 and EfficientNet), random zoom (0.9–1.1 for FCIE and EfficientNet, 0.8–1.2 for VGG16), random shear (0–10°, applied only to VGG16 and EfficientNet), and random brightness adjustment (±10% for FCIE, ±20% for VGG16 and EfficientNet). We apply different values of augmentation for each model to obtain optimal results in the training process, and it is important to note that all augmentation techniques are applied exclusively to the training set, while the validation and test sets remain unaugmented to ensure unbiased model evaluation.

C. Fuzzy Inference System (FIS) Implementation

To evaluate the efficacy of explicit rule-based reasoning in medical diagnosis, we developed two distinct Fuzzy Inference Systems based on the Mamdani and Sugeno models. Unlike deep learning architectures that perform automated end-to-end feature learning, traditional FIS relies on pre-defined numerical descriptors to construct membership functions and inference rules. Consequently, this phase focuses on the manual extraction of representative features from the MRI data to serve as crisp inputs for the fuzzy controllers.

C.1. Statistical and Texture Feature Extraction

Since Fuzzy Inference Systems (FIS) require numerical inputs, the MRI images were processed to extract statistical and textural descriptors. The following statistical and texture-based features were computed,

Table 1. Image feature extraction

Feature	Description	Computation Method
Mean Intensity (μ)	Measures the average pixel intensity within the ROI (Region of Interest).	N/A
Variance (σ^2)	Quantifies the spread of pixel intensities, useful for detecting tissue abnormalities.	N/A
GLCM Contrast (C)	Measures the complexity of the image based on intensity histogram distribution.	N/A
GLCM Homogeneity (Gh)	Measures the closeness of GLCM elements to its diagonal, indicating texture uniformity.	N/A
Skewness (S)	Measures the asymmetry of the intensity distribution relative to the mean.	N/A

Feature	Description	Computation Method
Kurtosis (K)	Quantifies the sharpness or peakedness of the intensity distribution.	N/A
Area of ROI (A)	Represents the area of the segmented region, typically corresponding to hippocampus or cortical tissue.	Computed as the number of non-zero pixels in the binary segmentation mask.
Histogram Features	Represents the frequency distribution of intensity levels within the ROI.	Computed using intensity histograms (e.g., np.histogram).
Fractal Dimension (D)	Measures the spatial complexity of brain structures,higher values typically indicate normal tissue patterns.	Estimated using the Box-Counting or Differential Box-Counting (DBC) method.

Prior to fuzzy inference, each MRI image was processed sequentially through several steps, including Region of Interest (ROI) segmentation, extraction of statistical intensity features, extraction of texture features based on the Gray-Level Co-occurrence Matrix (GLCM), computation of fractal dimension, and calculation of the ROI area. The final output consisted of nine numerical features, all normalized to the range (0.0-1.0).

C.1.1 Region of Interest (ROI) Segmentation

The segmentation process aims to isolate the brain tissue from the background. First, images were resized to 224×224 pixels using bilinear interpolation. To mitigate noise, a 3×3 Gaussian filter was applied, mathematically defined as,

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

where $\sigma = 0.8$. Subsequently, Otsu's Thresholding was employed to determine the optimal intensity threshold T that minimizes intra-class variance. This threshold generates a binary mask $B(x,y)$,

$$B(x,y) = \{1 \; if \; I(x,y) > T; \; 0 \; otherwise\} \quad (2)$$

Post-processing involved morphological opening to remove small noise artifacts, followed by extracting the Largest Connected Component (LCC) to isolate the primary brain region. The final ROI image I_{ROI} is obtained by applying this mask to the original image.

C.1.2 Intensity Feature Extraction

From the segmented ROI, four fundamental statistical moments were computed based on the pixel intensity values $I(x,y)$, where N denotes the total number of pixels in the ROI,

Mean (μ), Represents average brightness,

$$\mu = \frac{1}{N} \sum_{x,y \in ROI}^{I} (x,y) \quad (3)$$

Variance (σ^2), Measures contrast/spread,

$$\sigma^2 = \frac{1}{N} \sum_{x,y \in ROI}^{I} (I(x,y) - \mu)^2 \quad (4)$$

Skewness (S), Quantifies histogram asymmetry,

$$S = \frac{1}{N} \sum_{x,y \in ROI}^{I} \left(\frac{I(x,y) - \mu}{\sigma} \right)^3 \quad (5)$$

Kurtosis (K), Measures the "tailedness" of the distribution,

$$K = \frac{1}{N} \sum_{x,y \in ROI}^{I} \left(\frac{I(x,y) - \mu}{\sigma} \right)^4 - 3 \quad (6)$$

Since Skewness and Kurtosis are unbounded, they were subsequently normalized to the (0, 1) interval using Min-Max scaling before the fuzzification stage.

C.1.3 GLCM Texture Feature Extraction

Texture analysis was performed using the Gray-Level Co-occurrence Matrix (GLCM). The ROI intensity range was quantized into 8 levels to reduce sparsity. The GLCM $P(i,j)$ represents the probability of a pixel with intensity i being adjacent to a pixel with intensity j at a specific distance of $d = 1$ and angle θ. To ensure rotation invariance, we computed the features for four directions ($0°, 45°, 90°, 135°$) and used the average value.

Contrast,

$$Contrast = \sum_{i,j}^{P} |i-j|^2 P(i,j) \quad (7)$$

Homogeneity,

$$Homogeneity = \sum_{i,j}^{P} \frac{P(i,j)}{1 + |i-j|} \quad (8)$$

Entropy,

$$Entropy = -\sum_{i,j}^{P} P(i,j) \log\big(P(i,j)\big) \quad (9)$$

C.1.4 Fractal Dimension (FD)

To measure the geometric complexity of the brain structure, we utilized the Box-Counting Method. The binary ROI mask is covered by a grid of squares with side length ϵ. The number of boxes $N(\epsilon)$ containing at least one foreground pixel is counted. This is repeated for decreasing values of ϵ. The Fractal Dimension (FD) is estimated as the slope of the log-log plot,

$$FD = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)} \quad (10)$$

Higher FD values indicate greater structural complexity, which correlates with healthy brain tissue preservation.

C.1.5 ROI Area Feature

The relative ROI Area (A_{rel}) serves as a proxy for brain atrophy. It is calculated as the ratio of brain pixels (N_{ROI}) to the total image pixels ($W \times H$),

$$A_{rel} = \frac{N_{ROI}}{224 \times 224} \quad (11)$$

This feature requires no further normalization as it naturally lies within the (0, 1) range.

C.2. Fuzzification and Membership Functions

Each of the nine input features was partitioned into three fuzzy sets to represent different linguistic levels. An exception was applied to the *Skewness* feature, which inherently describes the asymmetry of the intensity distribution. Therefore, it was divided into three specific categories, Negative, Zero, and Positive.

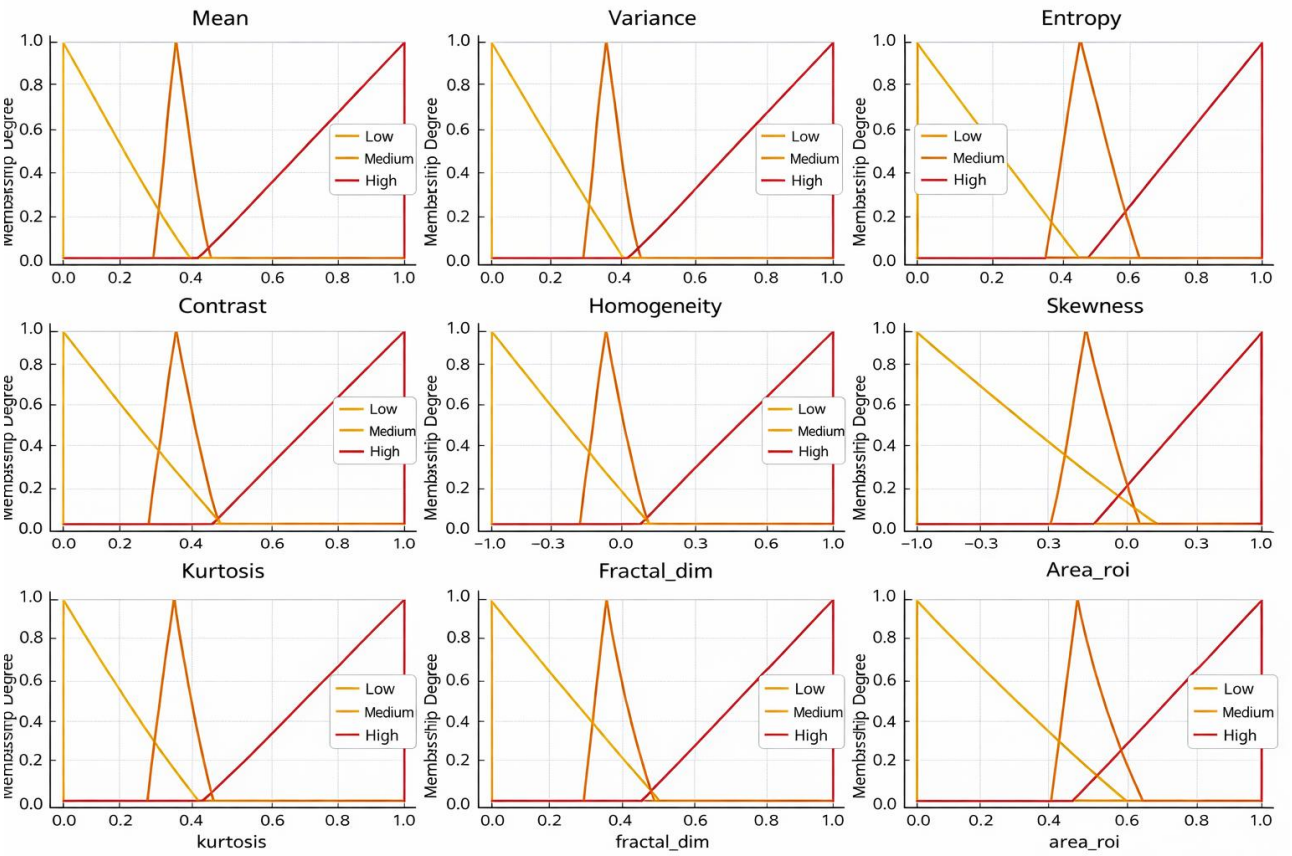

Figure *3*

All triangular membership functions are defined over the normalized domain. The Mamdani Fuzzy Inference System (FIS) employs nine input features, namely mean, variance, contrast, homogeneity, entropy, skewness, kurtosis, fractal_dim, and area_roi. Each feature, except skewness, is mapped into three fuzzy sets Low, Medium, and High which are represented by triangular membership functions. The skewness feature is instead modelled using the fuzzy sets Negative, Zero, and Positive. Each triangular membership function is defined by three parameters within the interval, where represents the peak of the function with a membership degree of 1, while and denote the points at which the membership degree becomes zero. For example, for the Mean feature, the Low, Medium, and High fuzzy sets are defined by the parameters, and, respectively, with similar parameterizations applied to the remaining eight features based on their observed normalized distributions.

The output of the Mamdani FIS, which represents the Alzheimer's disease classification, is also modeled using triangular membership functions for four diagnostic categories. The Non-Demented (ND) category is defined by the parameters with a peak at and is expressed as. The Very Mild Demented (VMD) category uses the parameters with a peak at, defined as. The Mild Demented (MD) category is modelled with and a peak at, expressed as. Finally, the Moderate Demented (ModD) category is defined by with a peak at, expressed as. These triangular output membership functions enable the Mamdani FIS to model gradual transitions between Alzheimer's disease stages while maintaining smooth and interpretable decision boundaries over the normalized output domain.

In the implementation of the FIS-Sugeno model, only three input variables were used in the rule base after being normalized to the range, Mean (μ), Variance (σ^2), and Contrast (C).

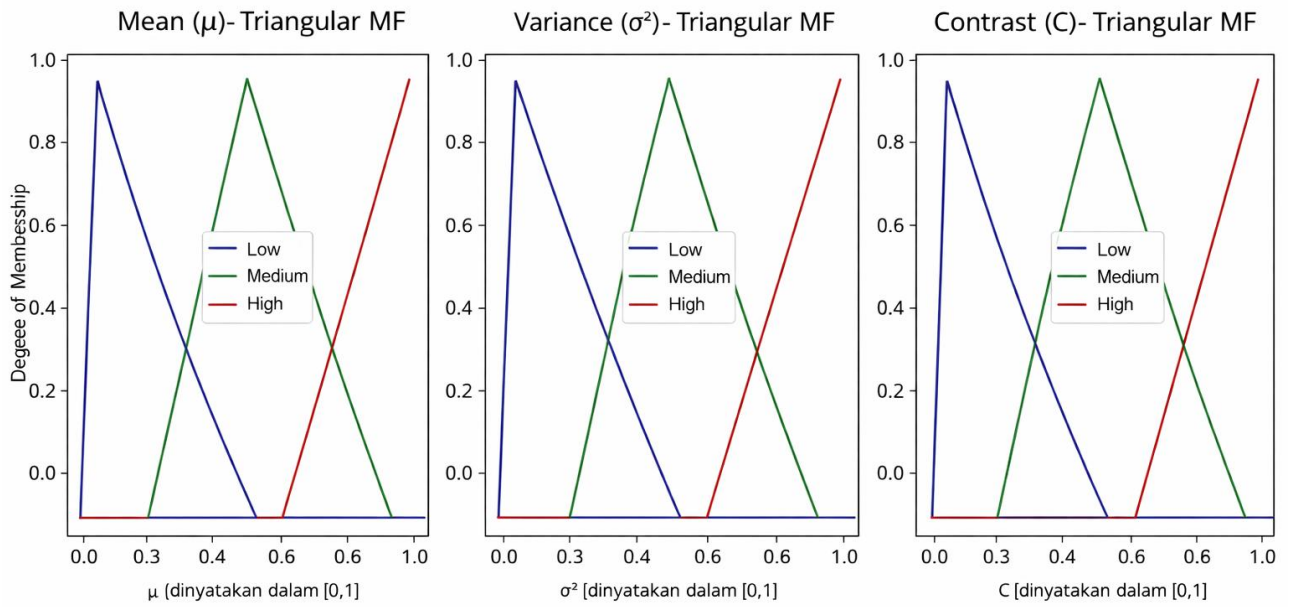

Figure 4. Sugeno Membership Function

Each of these variables is divided into three fuzzy sets (Low, Medium, and High), all of which use triangular membership functions defined by the parameters, where represents the peak of the membership function (membership value of 1), and denote the points at which the membership value becomes 0. For example, for the Mean feature, the Low, Medium, and High fuzzy sets are parameterized as, and, respectively, similar parameter settings are adapted for the Variance and Contrast features.

Unlike the Mamdani model, the consequents in the FIS-Sugeno system do not use output membership functions. Instead, each rule produces a consequent value defined by a linear algebraic function, where the coefficients are heuristically initialized and subsequently optimized using a Genetic Algorithm (GA). The final crisp output is computed as the weighted average of all rule outputs and is then mapped to one of the four Alzheimer's classes (NonDemented, VeryMildDemented, MildDemented, or ModerateDemented) based on predefined numerical thresholds (e.g., NonDemented).

C.3. Fuzzy Rule Base and Inference Mechanisms

The core of the reasoning process lies in the Fuzzy Rule Base, which maps the input feature space to the output diagnostic categories. Based on expert knowledge and preliminary data analysis, we constructed a heuristic rule base consisting of 50 IF-THEN rules. A representative rule structure is expressed as,

$$Rule\ i:\ IF\ x_1\ is\ A_{i1}\ AND\ \ldots\ AND\ x_{50}\ is\ A_{i50}\ THEN\ y\ is\ B_i \quad (12)$$

where x_j represents the j -th input feature, A_{ij} is the corresponding linguistic term (e.g., Low, Medium), and B_i is the output consequent.While both models share the same antecedent (IF part), they differ in their consequent (THEN part) formulation,

Mamdani Model, the consequent B_i is a linguistic fuzzy set (e.g., "Mild Demented"). The final crisp output is obtained using the Centroid Defuzzification method. Then for sugeno Model (First Order), the consequent is a linear function of the input variables, defined as,

$$y_i = p_{i0} + p_{i1}x_1 + p_{i2}x_2 + \cdots + p_{i50}x_{50} \quad (13)$$

The final output is computed as the weighted average of all rule outputs. This linear structure offers computational efficiency but introduces a complex parameter of space that requires optimization.

C.4. Model Parameter Optimization

To optimize the performance of the Mamdani FIS following the initial heuristic definition, a structured Grid Search was conducted on the membership function parameters. This process involved adjusting the midpoint *b* of each input membership function by ±10% from its initial value, shifting the boundary points *a* and *c* within a ±0.05 range to maintain proportional overlap, and modifying the peak values of the output membership functions with a small offset (±0.02). Grid Search was selected due to its ability to systematically explore the parameter space, effectively handling the nonlinear and multimodal characteristics of the Mamdani FIS, where small parameter changes may yield substantial effects, while preserving interpretability, as the resulting functions remain in the familiar triangular form. This method is also appropriate because it does not rely on gradient-based optimization, which is difficult to apply to the Mamdani structure. The outcome of the Grid Search is the determination of optimal *(a, b, c)* values for each input and output membership function, achieving maximal validation accuracy.

To optimize the linear consequent parameters for each rule in the Sugeno FIS after heuristic initialization, a Genetic Algorithm (GA) was employed. Each chromosome in the GA represents a complete set of linear parameters for all 50 rules (200 total parameters), with the initial population of 100 chromosomes standardized from the heuristic initialization. The evolutionary process incorporates tournament selection (size = 3), uniform crossover between parent chromosomes (probability = 0.8), and Gaussian mutation on each gene (σ = 0.1, probability = 0.05). The fitness function used is the Mean Squared Error (MSE) between the Sugeno output (*y*) and the numerical targets ({0.00, 0.33, 0.67, 1.00}), evaluated on the validation set using 5-fold cross-validation. The termination criteria include a maximum of 150 generations or stagnation of the fitness score for 15 consecutive generations. The choice of GA is justified by its capability to globally explore large and non-convex parameter spaces, its independence from gradient information (since the objective function is difficult to differentiate analytically), and its ability to balance exploration and exploitation to refine the highly sensitive linear weights. Although stochastic in nature, the final solution remains interpretable through the analysis of the resulting linear parameter contributions. Upon completion of the GA, the optimal set of linear parameters is obtained and subsequently used for inference on the test set to compute the final performance metrics.

D. Proposed Hybrid Fuzzy Convolutional Inference Engine (FCIE)

To overcome the limitations of manual feature extraction while maintaining model interpretability, we propose the Hybrid Fuzzy Convolutional Inference Engine (FCIE). This architecture integrates a custom Convolutional Neural Network (CNN) as a trainable feature extractor with a Mamdani Fuzzy Inference System for final decision-making.

D.1. CNN-Based Feature Extraction Module

The input to the FCIE is the pre-processed MRI image $X \in \mathbb{R}^{244 \times 244 \times 1}$. As described in the preprocessing stage, real-time data augmentation (rotation $\pm 10°$, flip $P = 0.5$, zoom $0.9 - 1.1$, brightness ± 10) is applied during training to enhance robustness.

The feature extraction backbone consists of five hierarchical convolutional blocks. Each block B_i (for $i=1 \dots 5$) follows a standardized sequence. (1) Conv2D with filter sizes increasing sequentially ($32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512$),

kernel size 3×3 . (2) Rectified Linear Unit (ReLU) for non-linearity. (3) Batch Normalization (BN) to stabilize gradient flow. (4) MaxPool2D (2×2) to reduce spatial dimensions. (5) Dropout (rate = 0.25) to prevent overfitting.

Following the final convolutional block, a Global Average Pooling (GAP) layer is employed to collapse the spatial dimensions. This is followed by a fully connected (Dense) layer with 128 units (ReLU) and a Dropout rate of 0.5.The critical interface between the neural and fuzzy components is the Latent Feature Layer. This is a Dense layer with 9 linear units, designed to map the high-level CNN representations into the 9 distinct features required by the fuzzy system. These outputs are passed through a Sigmoid-like normalization to ensure they strictly fall within the range (0, 1), serving as the crisp inputs for the subsequent fuzzy module.

D.2. Fuzzy Inference and Classification Module

The fuzzification module receives 9 learned features from the CNN. It utilizes the same Triangular Membership Functions defined in Section C.2. The inference engine operates based on a concise rule base of 30 IF-THEN rules. The decision-making process follows the standard Mamdani pipeline. (1) The Minimum operator ($\min$) is used to compute the firing strength of each rule's antecedent. (2) The Maximum operator ($\max$) aggregates the truncated output fuzzy sets across all active rules. (3) The Centroid method calculates the crisp output value y^*.

To enable end-to-end classification compatible with the categorical cross-entropy loss function, the scalar crisp output y^* is mapped to class probabilities using a scaled Softmax function,

$$P(C_k) = \text{Softmax}\big(\alpha \cdot d(y^*, C_k)\big) \tag{14}$$

where $\alpha = 10$ is a scaling factor that sharpens the probability distribution and $d(\cdot)$ represents the distance to the centroids of the four Alzheimer's classes (ND, VMD, MD, MoD).

The FCIE model was implemented using the TensorFlow/Keras framework. Training was executed on an NVIDIA T4 GPU, with convergence typically achieved in approximately 15 hours. The hyperparameters and training configuration are summarized as follows, (1) Optimizer using Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of 1×10^{-3}. (2) Scheduler using ReduceLROnPlateau (factor=0.5, patience=3) triggered by validation loss stagnation. (3) Loss Function using Categorical Cross-Entropy. (4) Batch Size 32. (5) Stopping Criterion using Early Stopping with a patience of 10 epochs.

D.3. Baseline Convolutional Neural Network Architectures

To benchmark the performance of the proposed FCIE/Hybrid model, we implemented two established deep learning architectures: VGG16 (representing deep, uniform structures) and EfficientNet-B0 (representing modern, efficient compound scaling).

D.3.1. Modified VGG16 Architecture

As the input MRI scans are grayscale ($H \times W \times 1$) while the pre-trained VGG16 expects RGB inputs ($H \times W \times 3$), we applied a pseudo-RGB conversion by replicating the grayscale channel three times: $I_{RGB} = \left[I_{gray}, I_{gray}, I_{gray}\right]$. Pixel intensities were normalized to the range (0-1).

We modified the standard VGG16 architecture to improve regularization on the small MRI dataset. The network comprises five convolutional blocks with sequentially increasing filter depths ($64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 512$). Unlike the original implementation, we introduced Batch Normalization and Dropout (rate = 0.25) after each block to mitigate internal covariate shift and overfitting. The output is flattened via GlobalAveragePooling2D into a 512-dimensional vector. A Dense layer (512 units, ReLU, Dropout 0.5) followed by a Softmax output layer (4 units). The

model was trained using the Adam optimizer ($LR = 1 \times 10^{-4}$) for 75 epochs. A dynamic learning rate scheduler (ReduceLROnPlateau) halves the LR if validation loss stagnates for 5 epochs.

D.3.2. EfficientNet-B0 with Transfer Learning

For EfficientNet-B0, preprocessing involved the specific preprocess input function, which scales pixel values to the range (-1-1). Similar to VGG16, grayscale images were converted to 3-channel inputs to leverage ImageNet pre-trained weights. The architecture utilizes Mobile Inverted Bottleneck Convolution (MBConv) blocks. We employed a Two-Stage Transfer Learning Strategy to adapt the model to the Alzheimer's domain, (1) Feature Extraction, All base layers were frozen. Only the custom classification head (Dense 256, ReLU, Dropout 0.5) was trained for 10 epochs with a learning rate of 1×10^{-3}. This stabilizes the weights of the new top layers. (2) Fine-Tuning, The final MBConv block (Stage 7) and the top Conv2D head were unfrozen. The model was retrained for an additional 20–30 epochs with a reduced learning rate of 5×10^{-5} to fine-tune the high-level features without destroying the pre-trained weights.

Both models utilized Data Augmentation exclusively on the training set to ensure robustness. The specific augmentation parameters were empirically tuned for each architecture as detailed in Table 2.

Table 2. Hyperparameter and Augmentation Configuration

Parameter	VGG16 (Modified)	EfficientNet-B0	FCIE (Hybrid)
Input Shape	224 × 224 × 3 (RGB)	224 × 224 × 3 (RGB)	224 × 224 × 1 (Grayscale)
Rotation Range	± 15°	± 15°	± 10°
Zoom Range	[0.8, 1.2]	[0.9, 1.1]	[0.9, 1.1]
Shear / Shift	0–10° / 0.1	0–10° / 0.1	N/A
Flip Strategy	Horizontal & Vertical	Horizontal & Vertical	Horizontal (50%)
Brightness	± 20%	± 20%	± 10%
Optimizer	Adam	Adam	Adam
Learning Rate	1×10^{-4}	Phase 1: 1×10^{-3} Phase 2: 5×10^{-5}	1×10^{-3}
Training Time	~8 Hours	~9 Hours	~15 Hours

RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the five proposed models: two traditional Fuzzy Inference Systems (Mamdani and Sugeno), two baseline Deep Learning architectures (VGG16 and EfficientNet-B0), and the proposed Hybrid FCIE model. The performance is assessed using quantitative metrics including Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC). Additionally, computational efficiency is evaluated based on training duration. The overall performance comparison is summarized in Table 3.

Table 3. Metrics Model

Model	Accuracy	Preccision	Recall	F1-Score	AUC
FIS-Mamdani	0.32	0.33	0.32	0.32	0.48
FIS-Sugeno	0.35	0.36	0.35	0.35	0.51
FCIE	**0.98**	**0.98**	**0.98**	**098**	**0.99**
VGG16	0.93	0.95	0.95	0.95	0.99
EfficientNet-B0	0.96	0.97	0.95	0.97	0.99

A. Performance of Traditional Fuzzy Systems

A.1. FIS-Mamdani Evaluation

The FIS-Mamdani model demonstrated suboptimal performance in classifying the four stages of Alzheimer's disease. As detailed in Table 3, it achieved a validation accuracy of 32.01%, with an F1-Score of 0.32 and a cumulative AUC of 0.48. While the model is computationally inexpensive, requiring only 2 minutes for training on a standard Intel i7 CPU, its diagnostic capability is close to random guessing for a 4-class problem.

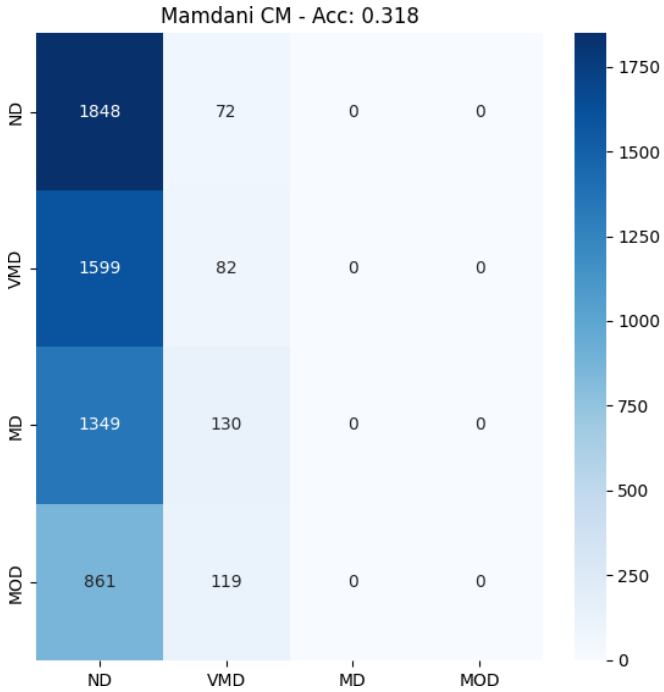

Figure 5. Confusion Matrix for evaluation of FIS Mamdani

The confusion matrix in Figure 5 elucidates the underlying cause of this performance. While the model shows a marginal ability to identify extreme cases (Non-Demented and Moderate Demented), it fails significantly in distinguishing intermediate stages (Very Mild and Mild Demented). This is corroborated by the per-class AUC values ($ND = 0.50, VMD = 0.47, MD = 0.45, MoD = 0.49$), which hover around 0.5. This failure suggests that the fuzzy rule base, derived solely from simple statistical and texture features, lacks the complexity required to capture the subtle non-linear transitions between early-stage dementia.

A. 2. FIS-Sugeno Evaluation

The FIS-Sugeno model exhibited a slight performance improvement over the Mamdani approach, yielding a validation accuracy of 35.00% and an AUC of 0.51. The training duration remained highly efficient at approximately 2.5 minutes.

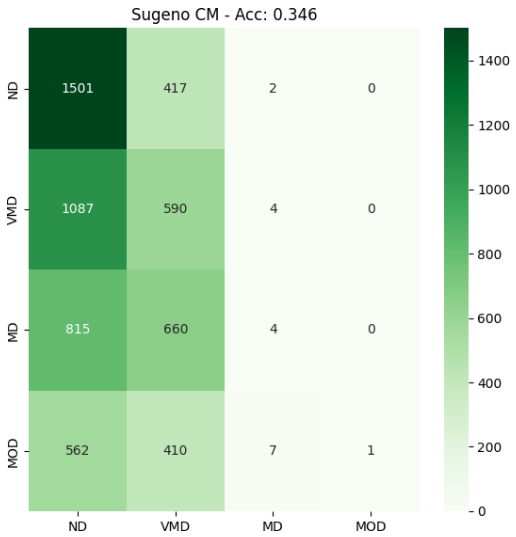

Figure 6. Confusion Matrix for evaluation of FIS Sugeno

Analysis of Figure 6 indicates that the use of linear consequent functions in the Sugeno model allows for a better approximation of the decision boundaries compared to the constant linguistic terms used in Mamdani. The per-class AUCs ($ND = 0.52, VMD = 0.50, MD = 0.48, MoD = 0.54$) reflect this marginal gain. However, the overall accuracy remains below 40%, confirming that the primary bottleneck lies not in the inference mechanism (Mamdani vs. Sugeno), but in the input feature representation. The nine hand-crafted features (GLCM and statistics) are insufficient to represent the complex morphological heterogeneity of Alzheimer's brain atrophy, leading to poor generalization. Nevertheless, these models retain the advantage of high interpretability, offering a transparent decision-making process that "black-box" models lack.

B. Evaluation of Deep Learning Benchmarks

B.1. VGG16 Performance Analysis

The VGG16 architecture exhibited robust capabilities in extracting spatial features from grayscale MRI scans. As a baseline deep learning model, it achieved a validation accuracy of 93.14% and a combined class of AUC of 0.99, significantly outperforming traditional fuzzy systems.

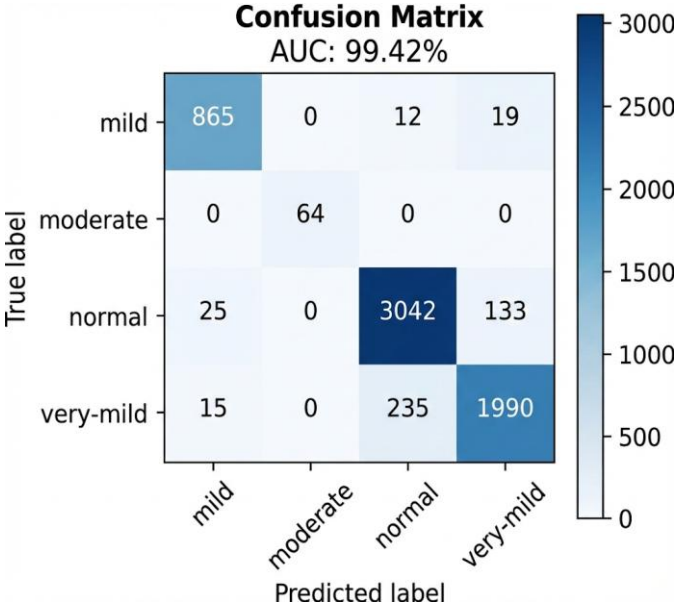

Figure 7. Confusion Matrix for evaluation of VGG16

The confusion matrix (Figure 7) indicates relatively strong classification consistency. However, per-class ROC analysis reveals slight variations in sensitivity, with AUC values of 0.96 (ND), 0.94 (VMD), 0.93 (MD), and 0.97 (MoD). While effective, the discussion highlights two limitations: (1) Mild overfitting was observed after the 70th epoch, necessitating aggressive dropout and augmentation strategies, and (2) Parameter inefficiency, which limits its scalability on larger datasets. Nevertheless, the primary advantage of VGG16 remains its architectural simplicity, facilitating straightforward debugging and feature visualization.

B.2. EfficientNet-B0 Performance Analysis

Leveraging the compound scaling method, EfficientNet-B0 demonstrated superior performance metrics compared to VGG16. It achieved a validation accuracy of 96.40% (+3.26% over VGG16), with precision, recall, and F1-score all reaching 0.97.

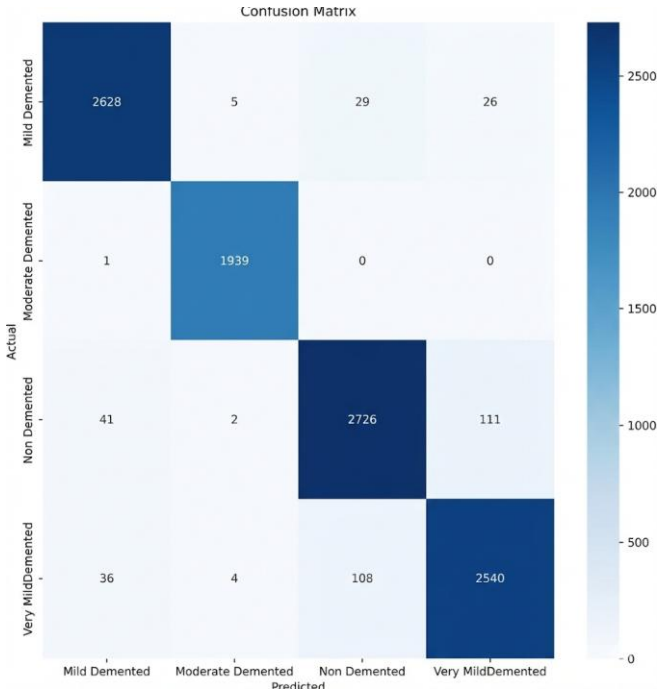

Figure 8. Confusion Matrix for evaluation of EfficientNet-B0

The confusion matrix (Figure 8) and per-class ROC curves confirm this enhancement, yielding high AUC values: ND (0.97), VMD (0.95), MD (0.94), and MoD (0.98). The superior performance is attributed to the model's ability to balance network depth, width, and resolution, resulting in more stable training convergence. Crucially, EfficientNet-B0 achieved these results with significantly fewer parameters than VGG16, positioning it as the optimal choice for resource-constrained environments where GPU memory is limited but high accuracy is required.

C. Performance of the Proposed Hybrid FCIE

The proposed Hybrid FCIE model achieved state-of-the-art performance in this experiment, recording the highest validation accuracy of 98.03% and a near-perfect combined AUC of 0.99. This represents a substantial improvement over both standalone deep learning models and traditional fuzzy systems.

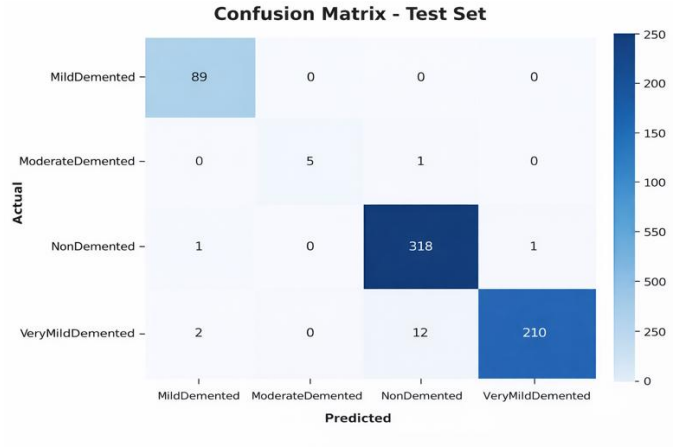

Figure 9. Confusion Matrix for evaluation of FCIE

The confusion matrix (Figure 9) demonstrates exceptional classification of purity, with negligible false positive/negative rates. This is further corroborated by the per-class AUC values: ND (0.995), VMD (0.990), MD (0.989), and MoD (0.997).

The success of the FCIE is attributed to the synergistic integration of CNNs and Fuzzy Logic. While the CNN component effectively extracts complex high-level spatial features, the Mamdani fuzzy inference module acts as a robust decision stabilizer, particularly in handling the uncertainty and noise inherent in the transition phases of Alzheimer's (e.g., distinguishing Very Mild from Mild Dementia).

However, this performance comes with a computational cost. The training time for FCIE was approximately 15 hours on an NVIDIA T4 GPU, which is nearly double the time required for VGG16 or EfficientNet. Despite the increased implementation of complexity and training duration, FCIE is recommended for clinical applications where diagnostic precision is paramount, and computational resources are sufficient.

CONCLUSION

This study comprehensively evaluated five MRI-based Alzheimer's classification models, revealing significant performance disparities between traditional rule-based systems and advanced deep learning architectures. While Fuzzy Inference Systems (Mamdani and Sugeno) offered high interpretability, they failed to achieve sufficient diagnostic accuracy (<35%) due to the limitations of hand-crafted statistical features in capturing complex morphological patterns. In contrast, deep learning benchmarks demonstrated superior performance, with VGG16 achieving 93.14% accuracy and EfficientNet-B0 offering an optimal balance of high accuracy (96.4%) and computational efficiency through compound scaling. Notably, the proposed Hybrid Fuzzy Convolutional Inference Engine (FCIE) emerged as the most robust architecture, achieving state-of-the-art validation accuracy of 98.03% and an AUC of 0.99 by successfully synergizing automated CNN feature extraction with the decision-stabilizing capabilities of fuzzy logic. Consequently, this study recommends the FCIE model for high-precision referral hospital settings and EfficientNet-B0 for resource-constrained clinical environments, with future research directed toward integrating Explainable AI (XAI) to further enhance clinical trust in these automated diagnostic tools.

ACKNOWLEDGEMENT

This research is supported by New Lecturer Scheme Research Grant Period 1 Batch 4 2025 from Telkom University, under Contract No. 667/LIT06/PPM-LIT/2025.

REFERENCES

1. Weidner WS. World Alzheimer Report 2022 - How strong are global health systems: Lessons learned and case studies from across the globe. Alzheimer's & Dementia. 2023;19(S23):e073714. doi:https://doi.org/10.1002/alz.073714
2. Helaly HA, Badawy M, Haikal AY. Deep Learning Approach for Early Detection of Alzheimer's Disease. Cognit Comput. 2022;14(5):1711–27. doi:10.1007/s12559-021-09946-2
3. Yao Z, Wang H, Yan W, Wang Z, Zhang W, Wang Z, et al. Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images. Eur J Radiol. 2023;165:110934. doi:https://doi.org/10.1016/j.ejrad.2023.110934
4. Li B, Li H, Li J, Zhang Y, Wang X, Zhang J, et al. Relaxation enhanced compressed sensing three-dimensional black-blood vessel wall MR imaging: Preliminary studies. Magn Reson Imaging. 2015;33(7):932–8. doi:https://doi.org/10.1016/j.mri.2015.03.009
5. van Oostveen WM, de Lange ECM. Imaging Techniques in Alzheimer's Disease: A Review of Applications in Early Diagnosis and Longitudinal Monitoring. Int J Mol Sci. 2021;22(4). doi:10.3390/ijms22042110
6. Aditya Shastry K, Sanjay HA. Artificial Intelligence Techniques for the effective diagnosis of Alzheimer's Disease: A Review. Multimed Tools Appl. 2024;83(13):40057–92. doi:10.1007/s11042-023-16928-z
7. Cao J, Zhou T, Zhi S, Lam S, Ren G, Zhang Y, et al. Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. Inf Sci (N Y). 2024;662:120212. doi:https://doi.org/10.1016/j.ins.2024.120212
8. Kutlu F, Ayaz İ, Garg H. Integrating fuzzy metrics and negation operator in FCM algorithm via genetic algorithm for MRI image segmentation. Neural Comput Appl. 2024;36(27):17057–77. doi:10.1007/s00521-024-09994-3
9. Yao Z, Mao W, Yuan Y, Shi Z, Zhu G, Zhang W, et al. Fuzzy-VGG: A fast deep learning method for predicting the staging of Alzheimer's disease based on brain MRI. Inf Sci (N Y). 2023;642:119129. doi:https://doi.org/10.1016/j.ins.2023.119129
10. Aborokbah M. Alzheimer's Disease MRI Classification using EfficientNet: A Deep Learning Model. In: 2024 4th International Conference on Applied Artificial Intelligence (ICAPAI). 2024. p. 1–8. doi:10.1109/ICAPAI61893.2024.10541281
11. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. Cognit Comput. 2024;16(1):45–74. doi:10.1007/s12559-023-10179-8
12. Islam J, Zhang Y. A Novel Deep Learning Based Multi-class Classification Method for Alzheimer's Disease Detection Using Brain MRI Data. In: Zeng Y, He Y, Kotaleski JH, Martone M, Xu B, Peng H, et al., editors. Brain Informatics. Cham: Springer International Publishing; 2017. p. 213–22.
13. Ahamed MdKU, Hossen R, Paul BK, Hasan M, Al-Arashi WH, Kazi M, et al. A hybrid filtering and deep learning approach for early Alzheimer's disease identification. Sci Rep. 2025;15(1):27694. doi:10.1038/s41598-025-03472-z
14. EL-Geneedy M, Moustafa HED, Khalifa F, Khater H, AbdElhalim E. An MRI-based deep learning approach for accurate detection of Alzheimer's disease. Alexandria Engineering Journal. 2023;63:211–21. doi:https://doi.org/10.1016/j.aej.2022.07.062

15. Tanveer M, Sajid M, Akhtar M, Quadir A, Goel T, Aimen A, et al. Fuzzy Deep Learning for the Diagnosis of Alzheimer's Disease: Approaches and Challenges. IEEE Transactions on Fuzzy Systems. 2024;32(10):5477–92. doi:10.1109/TFUZZ.2024.3409412
16. Rashed BM, Popescu N. Medical Image-Based Diagnosis Using a Hybrid Adaptive Neuro-Fuzzy Inferences System (ANFIS) Optimized by GA with a Deep Network Model for Features Extraction. Mathematics. 2024;12(5). doi:10.3390/math12050633
17. Abdkhaleq MHG. Hybrid Deep Learning and Neuro-Fuzzy Approach for COVID-19 Diagnosis Using CT Scan Imaging: Integration of CNN, ANFIS, and PCA. International Journal of Computational and Electronic Aspects in Engineering. 2025 Jun 26;6(2):81–8. doi:10.26706/ijceae.6.2.20250404
18. Vo TN, Ho BN. 3D Brain MRI Classification for Alzheimer Diagnosis Using CNN with Data Augmentation [Internet]. 2025. Available from: https://arxiv.org/abs/2505.04097
19. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data. 2019;6(1):60. doi:10.1186/s40537-019-0197-0
20. Goceri E. Medical image data augmentation: techniques, comparisons and interpretations. Artif Intell Rev. 2023;56(11):12561–605. doi:10.1007/s10462-023-10453-z