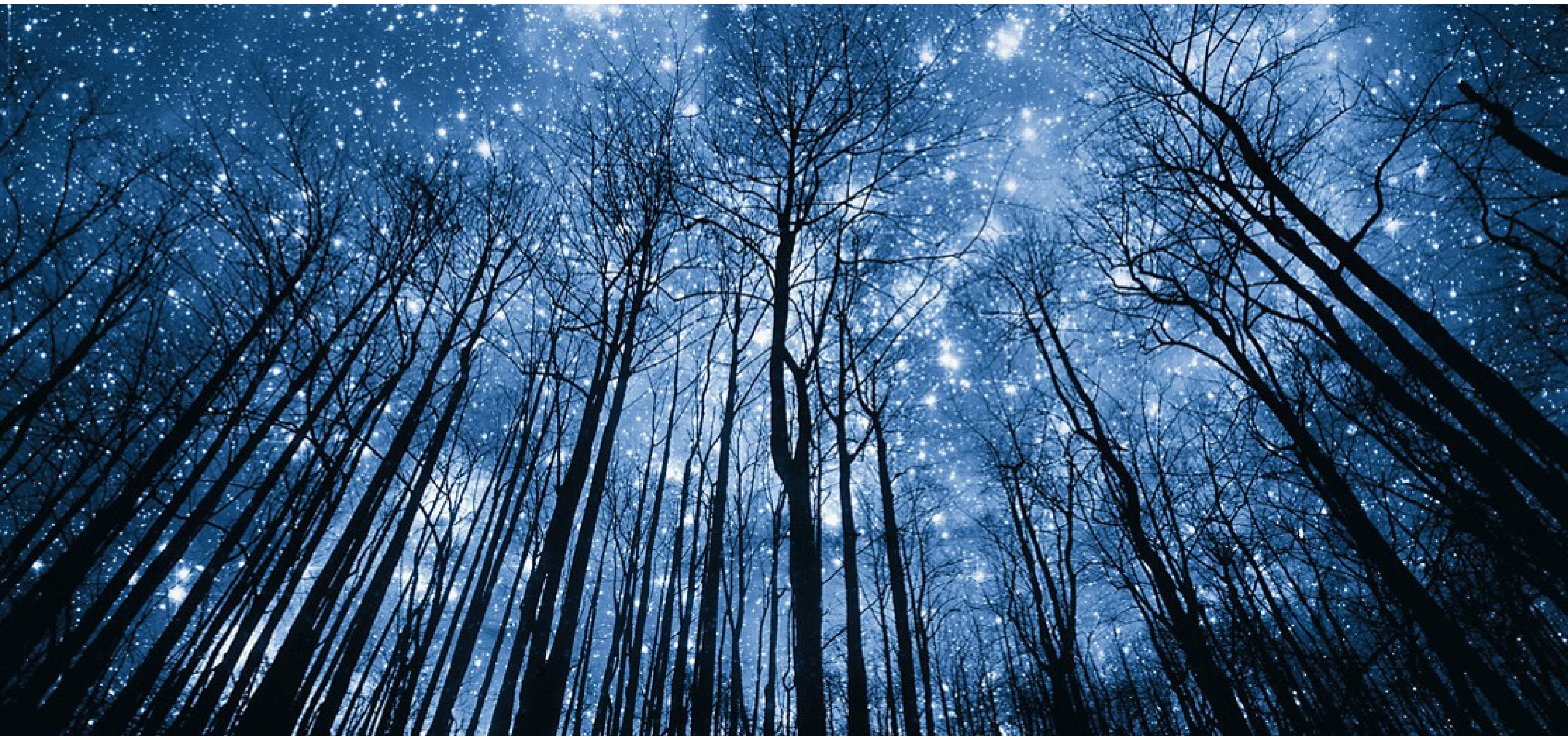
FUZZY CLASSIFICATION TECHNIQUES IN ARTIFICIAL INTELLIGENCE



Kartik Sumrani
Associate Sales Engineer Analyst
Dell Technologies

Knowledge Sharing Article

The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged and Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

Learn more at www.dell.com/certification

Table of Contents

Abstract

We are living in a data driven world, and all data is made up of bits. There are 8 bits in a byte, and 1024 bytes in a kilobyte. After kilobytes come megabytes, and so on. In 2025, we expect to collectively generate, record, copy, and process around 175 zettabytes of data, and there are no signs of slowing down. To interpret and organize this large amount of data, we now turn to Artificial Intelligence.

The future of humanity is being shaped by artificial intelligence in nearly every industry. It's a disruptive technology that is finding more and more uses every day. But with each new innovation in artificial intelligence technologies like machine learning, deep learning, neural network, the possibilities to scale a new horizon in tech widens up.

One such technique is classification using Artificial Intelligence, that is, it's ability to identify and recognize a wide variety of criteria—from image contents to the time of day. While there are many techniques at present to classify data, however, in this paper our focus will be on Fuzzy Rough Pattern Classifiers.

The algorithms of major systems are developed by using techniques and theories of the pattern recognition field, data pre-processing, feature extraction, and classification. In this whitepaper, analysis will be done by using a classic case of early detection of Alzheimer's from the medical big data. There is an attempt to explore the possibility of using fuzzy-rough pattern classifiers to identify brain pattern for predicting and analysis of Alzheimer's problems in an individual. A thorough comparative analysis will then be performed to find the most accurate classifier. This paper intends to impart practically applied knowledge in the following areas:

1. Artificial Intelligence
2. Classification Techniques
3. Case Study: Early Detection of Alzheimer's using Fuzzy Rough Pattern Classifiers
4. Comparative Study of the Classifiers
5. Future Scope in the field

1. Introduction to Artificial Intelligence

1.1 Overview

Artificial Intelligence (AI) enables machines to adapt to new inputs, learn from experience, and perform human-like tasks. Computers can be trained to perform particular tasks by processing a large amount of data and recognizing patterns. Several innovations and advancements that were previously only possible in science fiction have slowly become real over the past few years. Despite the fact that AI conjures up images of high-functioning, human-like robots taking over the world, it is not intended to do so. Its goal is to significantly increase human contributions and capabilities. It is a factor of production that has the potential to introduce new growth sources and alter the way work is done across industries. For instance, as per Gartner, 37% of organizations have implemented AI in some form. The percentage of enterprises employing AI grew 270% over the past four years. As a result, it is an extremely valuable business asset.

Over the course of the past few decades, a number of definitions of artificial intelligence (AI) have emerged. According to Wikipedia, “AI is intelligence—perceiving, synthesizing, and inferring information—demonstrated by machines, as opposed to intelligence displayed by animals and humans.” Speech recognition, computer vision, translation between (natural) languages, and other input mappings are examples of tasks where this is done. The Oxford English Dictionary defines AI as: “The theory and development of computer systems able to perform tasks normally requiring human intelligence.” However, one of my favorite definitions is by Gartner, which defined it as, “AI applies advanced analysis and logic-based techniques, including machine learning, to interpret events, support and automate decisions, and take actions”.

When it comes to AI, there are a lot of theories, approaches, and technologies. For example, the terms "Deep Learning" and "Machine Learning" are frequently used in the same sentence. However, there are differences.

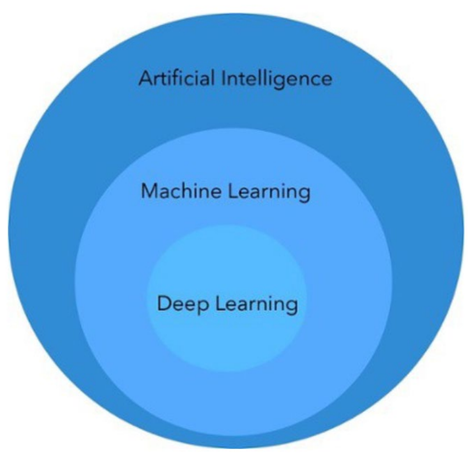


Figure 1.1: AL, ML and DL

The relationships between AI, ML, and DL are depicted by the concentric circles in the image above. ML and DL are included in AI. Since ML is a subset of AI, all ML algorithms are considered AI components. However, it does not operate in the opposite direction, and it is essential to keep in mind that not all AI-based algorithms are ML. This is comparable to the fact that every square is a rectangle but not every rectangle is a square. DL is the next level of ML. It is a subset of machine learning that draws inspiration from how human brains work.

Here's the key difference between Machine Learning and Deep Learning:

1. Machine Learning - It finds hidden insights in data using techniques from physics, statistics, operations research, and neural networks without being explicitly programmed to look where or draw what conclusions. More specifically, Machine Learning results is an algorithm or statistical formula that combines a number of data points into a single result. ML algorithms "learn" by "training," in which they use data patterns and correlations to make new predictions and insights.

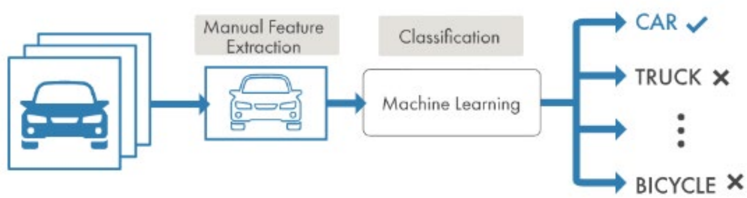


Figure 1.2: Machine Learning Approach

2. Deep Learning - Deep learning is a more advanced form of machine learning that makes use of neural networks, networks of algorithms that are modeled after the brain's structure. Each question it answers leads to a set of related questions in a deep neural network, which has nested neural nodes. By working with complex and frequently high-dimensional data, such as images, speech, and text, deep learning can outperform conventional machine techniques.

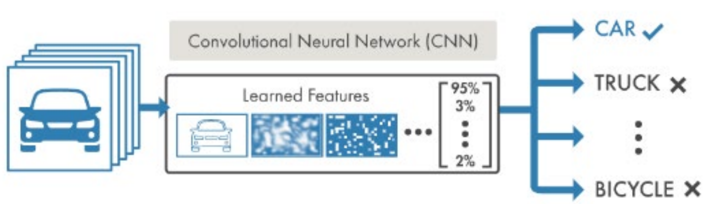


Figure 1.3: Deep Learning Approach

1.2 Using AI in Business

The presence of AI is reshaping our world, and new employment opportunities, such as those for software developers, data scientists, and information managers, are appearing daily. Businesses are finding more and more ways to make their day-to-day operations more efficient, keep in touch with customers, and gain a competitive advantage to accelerate growth using AI. The primary reason as to why AI is making an impact is due to the following two reasons:

1. Reveal better ways of doing things through advanced probabilistic analysis of outcomes
2. Interact directly with systems that take actions, enabling the removal of human-intensive calculations and integration steps

Before we deep dive into the classification techniques, let's look at why businesses are trying to realize the full potential of AI to drive value and innovation:

1. Automation: AI is able to carry out repetitive, high-volume, and reliable computerized tasks without getting tired. Businesses can focus on work that is more strategic and has an impact when AI is used to perform repetitive and time-consuming tasks.

2. Intelligence with Fast and Accurate Decisions: AI helps with decisions that are heavily influenced by data and involve a lot of complicated calculations which in turn reduces human error. The more data you can feed them, the more accurate they become. For example, your interactions with Alexa, Google Search, and Google Photos, are all based on deep learning, and the more we use them, the more accurate they become. AI methods like image classification, object recognition, and deep learning can now be used in medicine to find cancer on MRIs with the same precision as highly trained radiologists.

3. Availability and Scalable Communications: An intelligent system can work continuously throughout the day without taking breaks and swift communications are possible through the use of bots and virtual agents, which enable businesses to simultaneously offer guidance and assistance to a greater number of people.

In addition, AI is being utilized in a variety of contexts across functions, organizations, A few examples from business operations are:

1. Healthcare: Patients from all over the world have benefited from the use of artificial intelligence in healthcare in numerous ways. AI systems are assisting with routine, day-to-day administrative tasks like meeting scheduling and file system organization in an effort to reduce human error and boost productivity. It is assisting in early diagnosis of many diseases, robotic surgery, and health monitoring with a much narrower margin of error. This whitepaper will deal with one of the AI-applications in Healthcare industry.

2. E-Commerce: Personalized shopping with AI-powered virtual assistants is driving the growth of AI in e-commerce industry. Customers are provided with 24/7 self- and assisted-service options across channels by virtual customer assistants (VCAs) equipped with speech recognition, sentiment analysis, automated/augmented quality assurance, and other technologies. It can also assist in detecting and dealing with fraudulent reviews.

3. Finance: AI is assisting the financial sector in streamlining and optimizing procedures in a variety of areas, including credit decisions, quantitative trading, and financial risk management. Businesses and individuals can avoid significant losses by using artificial intelligence to spot changes in transaction patterns and other potential red flags that humans often overlook.

4. Cyber Security and IT Operations: By assisting professionals in identifying network irregularities through the analysis of user actions and patterns, AI has an impact on security. AI can now be used by security professionals to study network data and find vulnerabilities to prevent malicious attacks.

5. Social Media and Marketing: It can help with real-time personalization, content and media optimization, campaign orchestration, and other marketing processes and tasks that would otherwise be limited by human costs and capability. The ability of AI to quickly discover new customer insights and accelerate marketers' ability to deploy them at scale is the most compelling value proposition.

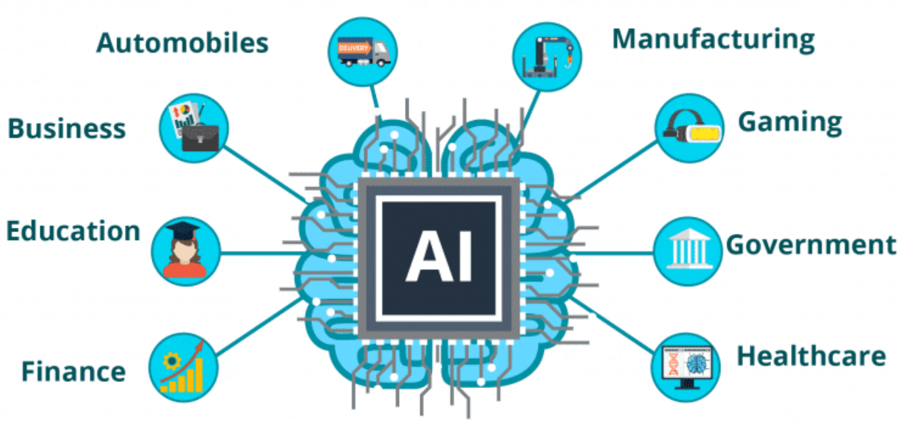


Fig 1.4: Applications of Artificial Intelligence

AI will continue to be pursued by businesses in the future to improve their decision-making processes. They will become more adaptable and more responsive to changes in the ecosystem if they adopt these strategies quickly.

2. Classification Techniques in Artificial Intelligence

2.1 Why Classification is Important?

Building intelligent machines from vast amounts of data and getting them to perform human-like tasks is the focus of artificial intelligence, which improves the speed, precision, and efficiency of human efforts. The first job for many artificial intelligence (AI) algorithms is to examine the data and find the best classification. A street sign, for instance, might be captured by an autonomous vehicle; The street sign must be read by the classification algorithm and compared to a list of known shapes and sizes for it to be understood. A phone must listen to a sound to determine if it is a wake-up command (such as "Alexa," "Siri," or "Hey Google").

An algorithm's ultimate objective may sometimes be to perform the classification task. AI algorithms are used by many data scientists to preprocess data and assign categories. Frequently, the primary job is simply observing the world and keeping track of what is going on. For instance, security cameras are now programmed to identify specific suspicious activity, radiologist uses medical images to diagnose/classify diseases, autonomous cars rely on street sign classification and object detection systems to make smart decisions.

Now, let's understand how classification is defined and how we can approach it in AI. The process of classifying a set of data into distinct categories is known as classification. Both structured and unstructured data can be used with it. "Target," "Label," or "Categories" are frequently used abbreviations for classifications. Based on what it learned and statistical calculations, it creates a model for each category that probably reflects the kind of product in that category. The model then places the new products/data points into various categories.

The following steps determine the appropriate category for a given observation:

1. The model uses a classification algorithm to find characteristics that are shared by certain classes.
2. It then relates those characteristics to the data you are attempting to categorize.
3. Finally, it makes use of that data to estimate the likelihood that an observation belongs to a specific class.

Two classical approaches for performing classification are supervised classification and unsupervised classification. The main difference is one uses labeled data to help predict outcomes, while the other does not.

1. Supervised Learning: Supervised learning is a machine learning approach that's defined by its use of labeled datasets. The purpose of these datasets is to "supervise" or train algorithms to accurately classify data or predict outcomes. The model is able to measure its accuracy and learn over time thanks to labeled inputs and outputs. There are currently a number of supervised classification techniques, some of which are common: Linear Regression, Logistic Regression, Gaussian Naive Bayes, Decision Trees, Support Vector Machine (SVM), and Random Forest.
2. Unsupervised Learning: Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled data sets. The term "unsupervised" refers to the fact that these

algorithms do not require human intervention to uncover hidden patterns in data. When there is less information available prior to classification, unsupervised classification can be utilized. Some of the common unsupervised classification methods are K-Means Clustering, KNN, and Hierarchical clustering.

Choosing the right approach for your situation depends on how your data scientists assess the structure and volume of your data, as well as the use case. Do the following before making a decision:

1. Evaluate the input data and identify if it's labeled or unlabeled.
2. Define your goals and predict if your data is static or dynamic in terms of parameters.
3. Review the options available and decide the best algorithm that suits your needs.

Classifying big data can be a real challenge in supervised learning, but the results are highly accurate and trustworthy.

2.2 Fuzzy Logic

The field of fuzzy logic has yet to be fully explored, but it can be applied to a wide range of fields. Fuzzy refers to something that is a bit vague. The computer may not be able to produce a True or False result when a situation is unclear. However, a Fuzzy Logic algorithm considers all the uncertainties of a problem, where there may be other possible values. A value on the range [0.0, 1.0] is used to indicate membership values in fuzzy sets or truth values in fuzzy logic, with 0.0 denoting absolute falsehood and 1.0 denoting absolute truth, respectively. This is the fundamental idea behind fuzzy systems. It has shown to be an effective tool for making decisions as well as for handling and manipulating erratic and noisy data.

Fuzzy Logic Architecture:

i. Rules: It contains all of the expert-provided rules and if-then conditions to control the decision-making process. Fuzzy controller design and tuning can now be accomplished in a variety of efficient ways thanks to the most recent revision of the theory. Usually, these developments reduce the number of fuzzy rules.

ii. Fuzzification: In this step, crisp numbers are transformed into fuzzy sets. A set of elements with identical properties is called a crisp set. An element can either belong to the set or not, according to some logic.

iii. Inference Engine: The degree to which the rules and the input values agree is determined by the inference engine. The input values are used as the basis for the rules. After that, control actions are developed using the rules. The inference engine and the knowledge base together are called a controller in a Fuzzy Logic system.

iv. Defuzzification: The fuzzy sets are transformed into a crisp value by the Defuzzification procedure. This can be accomplished using a variety of defuzzification techniques, but the most effective one is chosen based on the input.

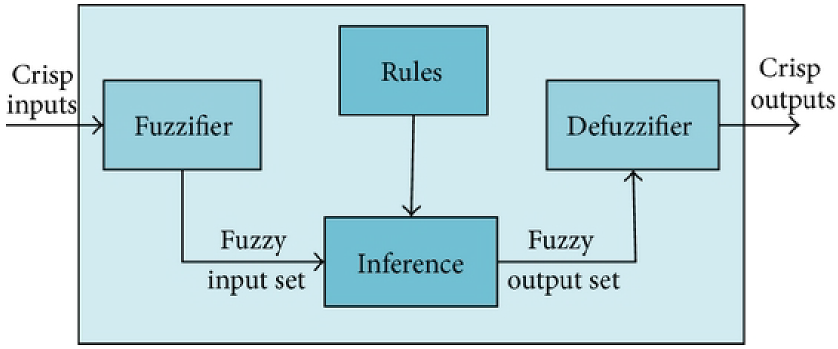


Fig 2.1: Fuzzy Logic Architecture

Now that we know the architecture of Fuzzy Logic, let's understand the significance of Membership Function and the role it plays. The fundamental building blocks of the fuzzy logic controller's Fuzzy Set Theory are membership functions (MFs). It shows the degree of truth in the fuzzy logic and the fuzziness that is described by the fuzzy system. In mathematical terms, it's a graph that shows how each input space point is mapped to a membership value between 0 and 1 and allows you to quantify linguistic terms with a key objective is to plot the non-fuzzy input values to the fuzzy linguistic terms, and vice versa.

A membership function for a fuzzy set A on the universe of discourse X is defined as $\mu A:X \rightarrow [0,1]$, where each element of X is mapped to a value between 0 and 1. This value, called membership value or degree of membership, quantifies the grade of membership of the element in X to the fuzzy set A.

Different types of membership functions are Singleton Fuzzy Logic System, Trapezoidal, Gaussian Fuzzifier, and Triangular Fuzzifier.

In both supervised and unsupervised classification, a fuzzy approach can be utilized. Fuzzy logic helps to regulate machines and provides a range of adequate thinking that can occur as a subset of the human decision-making process. It also helps to understand how people will perform in a dynamic environment in an unidentified way. In this article, the focus is given on application of fuzzy in images for pattern recognition and classification of medical images. This technique is slowly gaining popularity since one of its major advantages is that it describes the problems in a more natural way rather than trying to establish a relationship between numerical values.

3. Case Study: Fuzzy Pattern Classifier for Neurological Disorder Detection

3.1 Introduction Statement

Alzheimer's disease is a neurologic condition in which the brain shrinks (atrophy), and brain cells die. Alzheimer's disease is the most prevalent cause of dementia, which is characterized by a progressive loss of cognitive, behavioral, and social abilities that impairs a person's capacity to operate independently. Early diagnosis – and access to the right services and support – can help people take control of their condition, plan and live well with the disorder. It will help to eliminate the possibility of other, potentially treatable, conditions with dementia-like symptoms being responsible for memory, communication, behavior, and other problems. According to research, pharmacological and non-pharmacological Alzheimer's disease treatments may be more efficacious in the early stages of the disease than in later stages. Early detection has multiple benefits, including, but not limited to, gain access to information, resources, and support, and improve the quality of life.

Diagnosis of neurological disorders like Alzheimer’s require high level of precision, effort, and experience. Current diagnosis of Alzheimer’s involves a variety of approaches and tools, and is based on patient’s medical history, mental status tests, neurological exams, and brain imaging. Experts usually look for structural changes in the brain and high levels of beta-amyloid deposits to confirm the presence of Alzheimer’s disease. Recently, wide variety of advanced diagnostic technologies have been used to detect, manage, and treat neurological diseases such as Computerized Tomography (CT) Scan, Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET Scan), which have revolutionized our understanding of structure and functions of the brain. These technologies produce huge quantity of big medical data, and visual inspection is a challenging process for an individual to collect, manage, analyses, and integrate such vast volumes of data. As a result, experts have been demanding computerized diagnostic tools which can detect neurological disorders using medical big data automatically.

A large number of groups are conducting research in the area of applied computer-aided tools to study medical imaging data, but its accurate evaluation and diagnostic ability still remains a challenge. Soft computing techniques have been attracting researchers over the past years due to their ability to process and evaluate complex medical records and images. In this article, we’ll discuss three different forecasting techniques – Adaptive Neuro-Fuzzy Inference System, Fuzzy Min-Max Classifier, and Fuzzy KNN – that have been deployed to classify and predict different stages of Alzheimer’s disease based on structural brain magnetic resonance imaging (MRI) data, in order to aid experts by increasing the consistency and accuracy of the diagnosis and reduce the analysis time. The data features connected to diseases can be defined based on the projected findings, which can serve as a foundation for clinical and basic research, as well as etiology and pathological alterations. A thorough comparative analysis is then performed to find the most accurate forecasting methodology for Alzheimer's disease diagnosis.

3.2 Conceptual Overview

3.2.1 Understanding Alzheimer's Disease

Alzheimer's disease symptoms develop throughout time, yet the rate at which the disease advances vary. A person with Alzheimer's disease survives on average four to eight years following diagnosis, although it can last up to 20 years depending on other circumstances. Alzheimer's disease causes changes in the brain that begin years before symptoms appear. Since Alzheimer's affects people in different ways, each person may experience symptoms — or progress through the stages — differently. The stages below are intended to give a basic concept of how abilities change after symptoms manifest.

This paper involves the study of four stages, namely, Cognitive Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD).

1. Cognitive Normal (CN): In the first stage, a person with Alzheimer's disease has no memory impairment with no evident symptoms of dementia. At this stage, Alzheimer's disease is undetectable. This stage is also sometimes called No Cognitive Decline.
2. Early Mild-Cognitive Impairment (EMCI): Individuals in this stage start experiencing increased forgetfulness as well as slight difficulty with focus or concentration. They may get lost or begin to struggle to find the right words in communication.
3. Mild Cognitive Impairment (MCI): Major memory deficiencies are present beginning this stage, with individuals often forgetting prominent bits of information that affect their daily lives. They may not be able to identify where they are (orientation to place) or what time of day it is (orientation to time).
4. Alzheimer's Disease (AD): At this stage, most people will have lost their ability to speak or communicate. They often require assistance with most of their activities. People in this seven often lose psychomotor capabilities, they may be unable to walk or require significant assistance with ambulation.

3.2.2 Understanding Brain MRI Images

A brain MRI is one of the most regularly used medical imaging procedures. It allows clinicians to investigate the anatomy and disease of different areas of the brain using different MRI sequences, such as T1w, T2w, or FLAIR. The most common MRI sequences are T1-weighted and T2-weighted scans. T1-weighted images are produced by using short Time to Echo (TE) and Repetition Time (TR) times, while T2w have long TE and TR. The contrast and brightness of the image are predominately determined by T1 properties of tissue.

Tissue	T1-Weighted	T2-Weighted
CSF	Dark	Bright
White Matter	Light	Dark Gray
Cortex	Gray	Light Gray
Fat (within Bone Marrow)	Bright	Light
Inflammation	Dark	Bright

Table 1: Difference between T1w and T2w [1]

Another important concept with Brain MRI Images relates to orientation of the scans. The following are the basic terms pertaining to MRI orientation, A sagittal plane would be viewed from the side, a coronal plane from the front, and a transverse plane from the top down.

i. Sagittal plane, (also known as median plane) is an y-z plane, perpendicular to the ground, which separates left from right. The mid-sagittal plane is the specific sagittal plane that is exactly in the middle of the body.
ii. Coronal plane, (also known as frontal plane) is an x-z plane, perpendicular to the ground, which (in humans) separates the anterior from the posterior, the front from the back, the ventral from the dorsal.
iii. Transverse plane, (also known as axial or horizontal plane) is an x-y-z plane, parallel to the ground, which (in humans) separates the superior from the inferior, or put another way, the head from the feet.

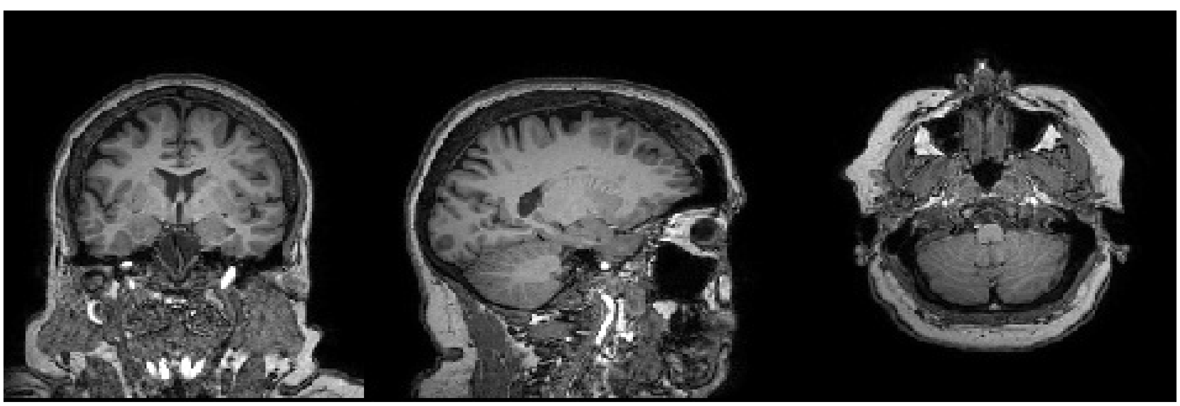


Figure 3.1: Different Orientation of Brain MRI Scan (left: Coronal, center: Sagittal, right: Axial)

3.3 Data Collection

The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset was used to compile data for this study. Under the direction of Dr. Michael W. Weiner, ADNI was established in 2004 as a public-private collaboration with $27 million from 20 corporations and two foundations through the Foundation for the National Institutes of Health and $40 million from the National Institute of Aging. Subjects from more than 50 sites aged 55 to 90 were recruited, including elderly people with normal cognitive abilities. In this study, 3-D, T-1 weighted brain images of 53 AD patients (28 males and 25 females), 57 MCI patients (25 males and 32 females), 50 EMCI patients (25 males and 25 females), and 50 CN patients (25 males and 25 females), acquired in the sagittal plane were studied. Data balance and gender bias was maintained throughout. Key demographic information about the database has been presented in Table 2. As seen, Table 2 depicts that gender bias and age range across samples was maintained throughout with little/no variation.

Stage	**Total Images**	**Gender (M/F)**	**Age Range (Years)**
CN	50	25/25	55-80
EMCI	50	25/25	55-80
MCI	57	25/32	60-90
AD	53	28/25	60-90

Table 2: Demographic Information of Subjects used in this study

Data from ADNI Database is collected in NifTI format which is processed and sliced with the help of NiLearn Library. Mid-level slicing of the sagittal section post pre-processing was done because brain pattern and features are more visible and provide maximum information. Figure 3.2 shows the mid-level slicing of the sagittal view.

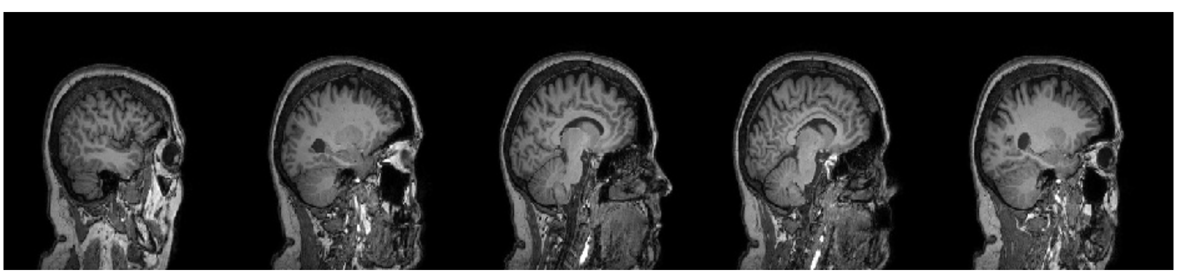


Figure 3.2: Mid-level slicing of sagittal view

3.4 Data Preprocessing

Preprocessing the raw images is the initial stage in every data-driven investigation. Preprocessing eliminates noise by ensuring that all images have a degree of parity, which improves the effectiveness of the segmentation and feature extraction stages that follow. To address the contrast differences in studies where images are taken from multiple sources and machines, images undergo normalization which is extremely necessary when employing machine-learning techniques. Before normalization, noise on scans of any modality, including the signal from the patient's skull, may need to be removed.

Data pre-processing is an important step when dealing with MR images as they may contain noise problem, intensity inhomogeneity, undesired parts in the background, and the quality of MR images often degrades during its acquisition process or afterwards. In this research, all MR images collected from ADNI Database were processed with a number of techniques to optimize, enhance, and eradicate certain details of the image, make it suitable for further analysis and evaluation. Each of MRI data has been preprocessed using various techniques which includes skull stripping, image registration, noise reduction, intensity normalization, and image enhancement.

Skull Stripping was performed with FSL-BET to eliminate skull and all the non-brain tissues such as fat and skin from the MR images. It can further eliminate inner and outer skull surfaces; all the images were then resampled to MNI152 sample space. Once the image was resampled, it was registered to a reference image. Image registration is a critical since it aligns multiple medical images adapted to a reference atlas (MNI152) to ensure that the spatial correspondence of same anatomical structures on the same regions of the image. MR Images also contain a significant amount of noise due to various reasons related to equipment and environment, field strength, voxel volume, or receiver bandwidth. Finally, images were de-noised and Z-score normalization was performed, which is defined as the image-signal intensity on a per-pixel basis.

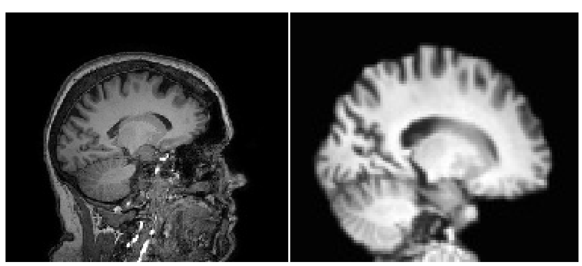


Figure 3.3: T1w MR Image of subject collected from ADNI database. a: Raw MR image, b: pre-processed MR image.

3.5 Feature Extraction using VGG16 and Dimensionality Reduction

Feature Extraction is the process of translating raw data into numerical features that can be processed while preserving the information in the original dataset. The aim is to extract quantitative information from an image that acts as an input to the classifier by considering the description of relevant properties of image into feature space. One way of extracting features from images is the use of deep learning-based models. Keras provides a set of deep learning models that are made available alongside pre-trained weights on ImageNet dataset. These models can be used for prediction, feature extraction, and fine-tuning. Such a process is also known as transfer learning. These models are computationally less expensive and has the benefit of decreasing the training time for a neural network and can result in lower generalization error. A wide range of high-performing models are available which can be easily integrated into a new model for classification purposes, such as VGG (VGG16 or VGG19), GoogLeNet (InceptionV3), or Residual Network (ResNet50).

In this study, VGG16 is used for automatic feature extraction. From the input layer to the last max pooling layer (labeled by 7 x 7 x 512) is regarded as feature extraction part of the model, while the rest of the network is regarded as classification part of the model. In this study, we have taken VGG16 model with the pre-trained weights for ImageNet dataset for feature extraction from pre-processed MRI images. Architecture of VGG16 is presented using block diagram in Figure 3.4. MRI images were loaded into the model with 224x224 as the input size, image features or pooled feature map (7x7x512) was extracted from the last max pooling layer. The next step involves flattening our feature vector into 2-dimensional array (i.e., sample size x 25088). This indicates that we now have successfully extracted 25088 features for each sample in the data.

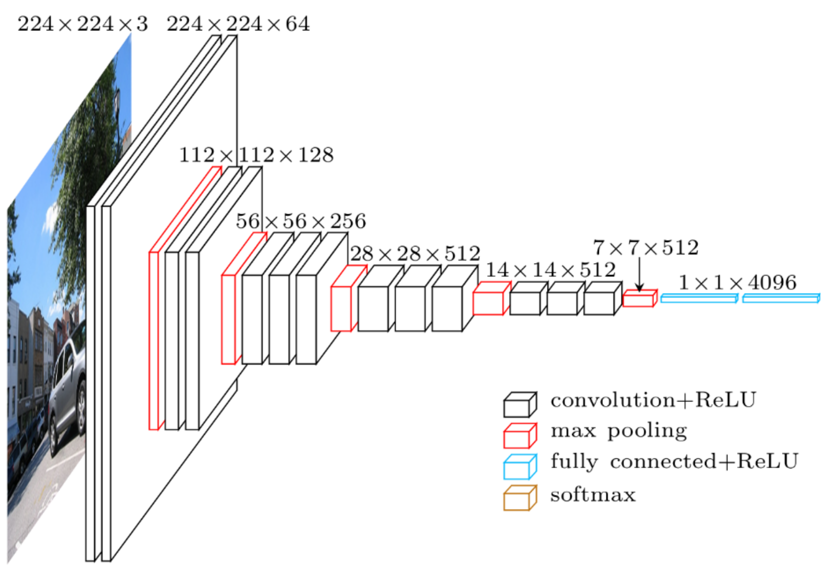


Figure 3.4: VGG16 Architecture

Practically, its computationally very expensive to input 25088 features per sample into the classification model, so we have used Dimensionality Reduction to be able to process these features. Dimensionality reduction strategies, as the name implies, lower the number of dimensions (i.e., variables) in a dataset while maintaining as much data as possible. There are various techniques which aid in dimensionality reduction such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). In this study, we have used Linear Discriminant Analysis (LDA) technique, which is a supervised dimensionality reduction algorithm that projects n-dimensional feature into a lower dimensional sub-space k (where $k \leq n-1$), while maintaining the class-separability. It takes the information from both characteristics and creates a new axis on which the data is projected in such a way that the variance is minimized and the distance between the two classes' means is maximized. Using LDA, our 25088-feature space vector was reduced to a lower dimension sub-space of 3 features only which was then passed to three models – ANFIS, Fuzzy Min-Max Classifier, and Fuzzy k-NN.

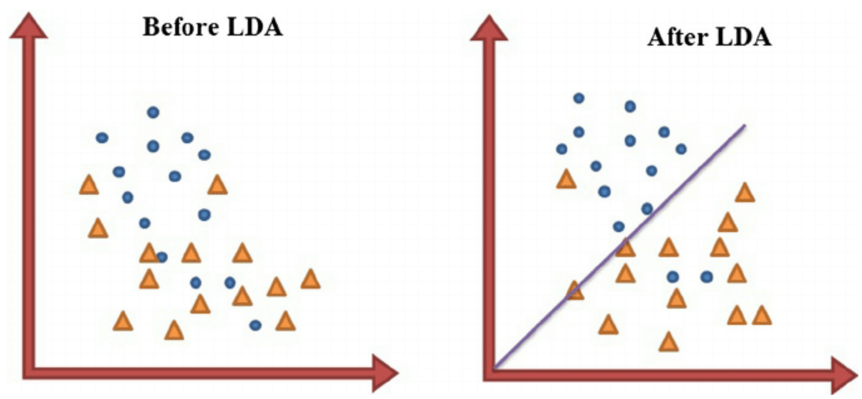


Figure 3.5: Linear Discriminant Analysis

Let's discuss some of the advantages of LDA and why we have deployed this technique in our study.

i. LDA performs dimensionality reduction by maximizing the class separability of classification datasets unlike PCA which depends upon maximizing the variance.
ii. The maximum number of components that PCA can find is equal to the number of input features (original dimensionality) of the dataset. We frequently prefer to locate a significantly smaller number of components that capture as much of the original data's variance as possible. The maximum number of components that LDA can find is equal to the number of classes minus one in the classification dataset. For instance, PCA analysis reduced the number of features to ~1200, which is still computationally expensive while LDA reduced the number to 3 only.

3.6 Model Building

Three models, namely, Adaptive Neuro-Fuzzy Inference System (ANFIS) with Particle Swarm Optimization (PSO), Fuzzy Min-Max Classifier, and Fuzzy k-NN was developed and tested.

Adaptive Neuro-Fuzzy Inference System with Particle Swarm Optimization

For pattern recognition and classification, fuzzy logic and neural networks are complementary to each other rather than competitive. As a result, rather than using these approaches individually, it is preferable to deploy them in combination. Adaptive Neuro-Fuzzy Inference System is based on this hybrid approach which combines the computational capabilities of both, the neural network which has the ability to recognize patterns and adapt to the changing environment, and the fuzzy logic, which takes into consideration the system's imprecision and uncertainty. A shared framework termed adaptive network, which unites the neural network with the fuzzy model, is an important aspect of neuro-fuzzy modelling. The neural network in the resulting hybrid intelligent system has the ability to recognize patterns and adapt to changing environments. The fuzzy inference system, on the other hand, incorporates human knowledge and executes inference. The neural network offers the system a sense of flexibility, while the fuzzy logic takes into account the system's imprecision and uncertainty.

The adaptive neuro-fuzzy inference system (ANFIS) is a five-layered architecture that is mathematically characterized as follows:

Let consider two inputs x and y and one output f with two fuzzy if-then rules defined as under:

Rule 1: If x is A_1 and y is B_1, then $f_1=a_1x + b_1y + c_1$

Rule 2: If x is A_2 and y is B_2, then $f_2=a_2x + b_2y + c_2$

Here, A_i, B_i are fuzzy sets in the antecedent, and a_i, b_i, and c_i are the design parameters determined during the training process. The individual layers of the fuzzy inference systems are described below:

Layer 1: Every node i^{th} node is an adaptive node with node function:

$$O_i^1 = \mu_{A_i}(x)$$

where, x is the input to the node, A_i is the linguistic variable associated with this node function and μ_{Pi} is the membership function of A_i. It can be bell shaped, gaussian or any other function. This node performs fuzzification operation on the linguistic variables.

Layer 2: Each node in this layer is a fixed node which calculates the firing strength of rule and is calculated by multiplying the input signals and the output.

$$O_i^2 = w_i = \mu_{A_i}(x) x \mu_{B_i} \quad \text{where, i =1,2}$$

Layer 3: Every node in this layer is a fixed node. This layer performs the normalization of the firing strengths.

$$O_i^3 = \overline{w_i} = \frac{w_i}{W_1 + W_2} \quad \text{where, i = 1,2}$$

Layer 4: Every node in this layer is an adaptive node with a node function given by,

$$O_i^4 = \overline{w_i} f_i = w_i(a_i x + b_i y + c_i) \ where, i = 1,2$$

Layer 5: Only one fixed node exists in this layer, which calculates the overall output as the total of all incoming signals, i.e.,

$$O_i^5 = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

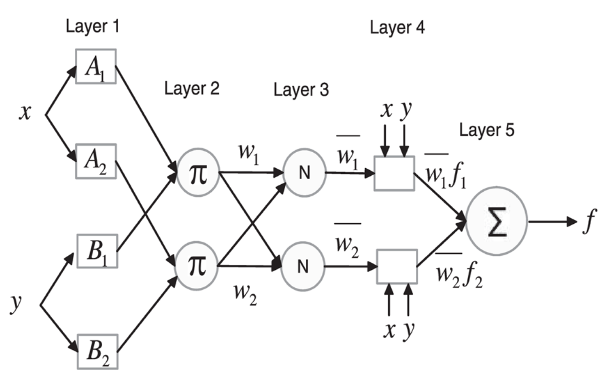


Figure 3.6: ANFIS Architecture

In the present study, the AD diagnostic system has been implemented using ANFIS with PSO trained for 500 epochs. Gauss membership function has been used and ANFIS Layout was initialized to [3,3,3] layout. Sigmoid and cross-entropy function are computed using a numerically stable implementation and Generalized Bell curves are plotted depending on three parameters (mean, standard deviation, and exponent). A particle swarm optimizer (PSO), which is a metaheuristic algorithm that solves complicated mathematics problems in engineering, is used to solve the minimization problem.

Particle swarm optimization (PSO) is a population-based optimization methodology that is inspired by the behavior of flocks of birds and schooling fish. PSO resembles evolutionary computation approaches in a number of ways. The system starts with a population of random solutions and updates generations to find the best solution. Potential solutions, referred to as particles in PSO, move across the problem space by following the current optimum particles. Both in terms of performance and memory needs, PSO is more computationally efficient.

Metric	**Value**
Number of Samples	2013
Number of Inputs	3
Number of Outputs	4
Classes	[0,1,2,3]
ANFIS Layout	[3,3,3]
Number of Premise Function (Sum)	9
Number of Consequent Function (Product)	27

Table 3: ANFIS Parameters

Fuzzy Min-Max Classifier

Fuzzy logic makes a significant contribution by providing a paradigm for computing with words that can deal with imprecision and granularity. The human brain is capable of deciphering and processing inaccurate and partial sensor data acquired from the perceptual organs. Similarly, fuzzy set theory can give a systematic strategy to dealing with such information from a linguistic standpoint.

The use of fuzzy sets as pattern classes in a supervised learning neural network classifier is presented. Each fuzzy set is made up of a collection of fuzzy set hyper boxes (union). A fuzzy set hyper box is an n-dimensional box with a membership function specified by a min point and a max point. The fuzzy min-max learning algorithm, an expansion-contraction process that can learn nonlinear class boundaries in a single pass through the data and allows for the incorporation of new and refined classes without retraining, is used to identify the min-max points. Inherently, the use of a fuzzy set method to pattern classification provides a level of membership information that is particularly important in higher-level decision making.

Fuzzy KNN

One of the most effective methods in supervised learning issues is the k-Nearest Neighbors (KNN) classifier. It classifies cases that haven't been seen before by comparing their similarity to the training data. Nonetheless, it assigns the same importance to each labelled sample when it comes to classification. There are various methods for improving precision, with the Fuzzy k-Nearest Neighbors (Fuzzy-KNN) classifier being one of the most effective. Fuzzy-KNN calculates each instance's fuzzy degree of membership to the problem's classes. As a result, the borders between classes are smoother. In FKNN, the fuzzy membership values of samples are assigned to different categories. The fuzzy strength parameter (m) is used to determine how heavily the distance is weighted when calculating each neighbor’s contribution to the membership value. After calculating all of the membership values of a query sample, it is assigned to the class with which it has the highest degree of membership.

3.7 Results and Discussion

After doing the tests, it was shown that while all three systems can diagnose AD, the ANFIS with PSO is more effective at predicting AD than Fuzzy Min-Max Classifier and fuzzy k-NN system. The training and test accuracies are mentioned in the Table 4.

Method	Train Accuracy	Test Accuracy
ANFIS	94.90%	96.77%
Fuzzy Min-Max Classifier	100%	91.81%
Fuzzy k-NN	93.20%	94.04%

Table 4: Comparison of classification accuracies

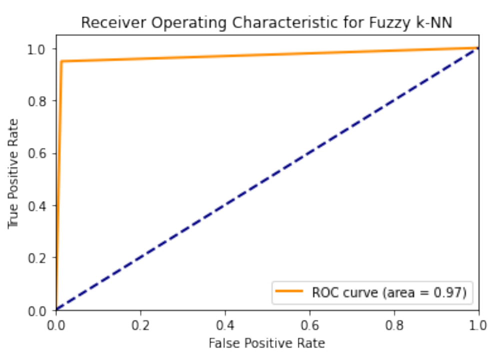
The area under the curve (AUC) and the receiver operating characteristic (ROC) curve are effective measures of accuracy with meaningful meanings. When analyzing diagnostic capabilities of tests to discern the true condition of subjects, identifying the appropriate cut off values, and comparing two different diagnostic tasks when each task is done on the same subject, this curve plays a critical role. ROC Operating Curves are shown in Figure 3.7.

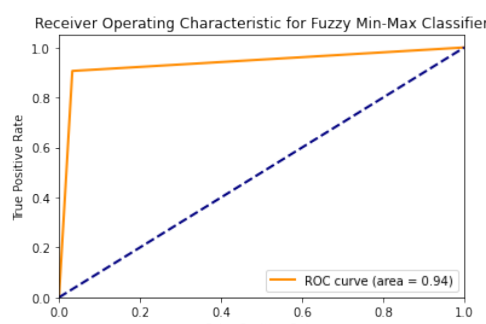
The intent of the ROC Curve is to show how well the model works for every possible threshold, as a relation of TPR vs FPR. So basically, to plot the curve we need to calculate these variables for each threshold and plot it on a plane. The location where TPR = FPR is depicted by the blue dotted line on the plots below, while the classifier's ROC curve is depicted by the blue line. The classifier has the same predictive power as flipping a coin if the ROC curve is exactly on the blue line.

AUC provides an aggregate measure of performance across all possible classification thresholds. The probability that the model ranks a random positive example higher than a random negative example is one way to interpret AUC. The AUC value ranges from 0 to 1. An AUC of 0.0 indicates a model whose predictions are 100% incorrect; an AUC of 1.0 is one whose predictions are 100% accurate. AUC is a desirable because:

1. It is Scale-invariant AUC. Instead of measuring predictions' absolute values, it measures how well they rank.
2. AUC is invariant at the classification threshold. No matter what classification threshold is selected, it measures the quality of the model's predictions.

Below are the AUC curves for individual models ranging from 0.94 to 0.98 and for individual classes used in the study:





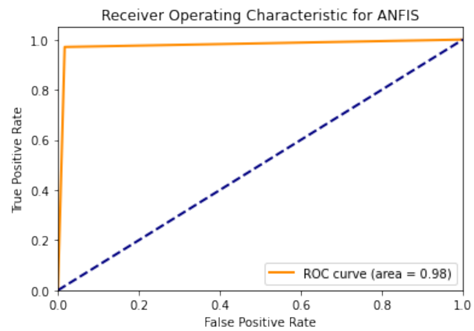


Figure 3.7: ROC Operating Characteristic for Models Proposed

ROC Operating Characteristic for individual classes is displayed in figure 3.8.

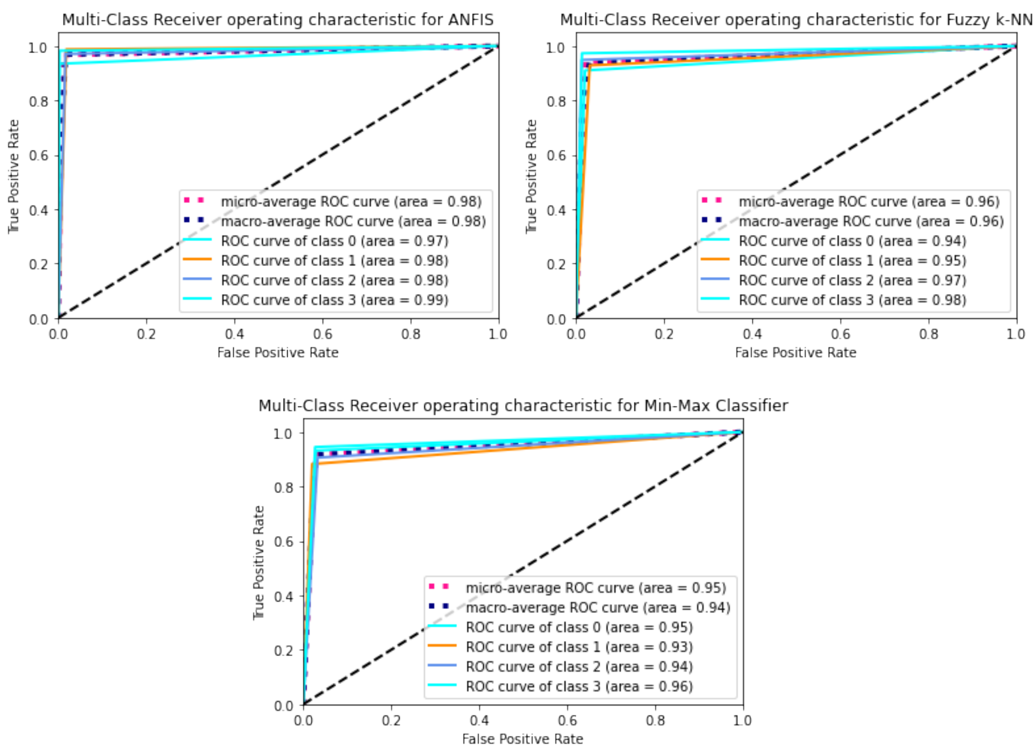


Figure 3.8: Multi-Class ROC AUC Curve

A classification report is used to assess the accuracy of a classification algorithm's predictions. On a per-class level, the report displays the primary classification metrics precision, recall, and F1-score. Classification Report for models is presented in this report.

Method	Precision	Recall	F1-Score	Support
AD	0.99	0.93	0.96	107
CN	0.93	0.99	0.96	80
EMCI	0.95	0.97	0.96	102
MCI	0.99	0.98	0.96	114
Accuracy	NA	NA	0.97	403
Macro Avg.	0.97	0.97	0.97	403
Weighted Avg.	0.97	0.97	0.97	403

Table 5: Classification Report for ANFIS

Method	Precision	Recall	F1-Score	Support
AD	0.94	0.91	0.93	111
CN	0.89	0.93	0.91	84
EMCI	0.96	0.95	0.95	97
MCI	0.96	0.97	0.97	111
Accuracy			0.94	403
Macro Avg.	0.94	0.94	0.94	403
Weighted Avg.	0.94	0.94	0.94	403

Table 6: Classification Report of Fuzzy k-NN

Method	Precision	Recall	F1-Score	Support
AD	0.92	0.93	0.92	132
CN	0.93	0.88	0.91	119
EMCI	0.89	0.91	0.90	117
MCI	0.93	0.94	0.94	145
Accuracy			0.92	513
Macro Avg.	0.92	0.92	0.92	513
Weighted Avg.	0.92	0.92	0.92	513

Table 7: Classification Report of Fuzzy Min-Max

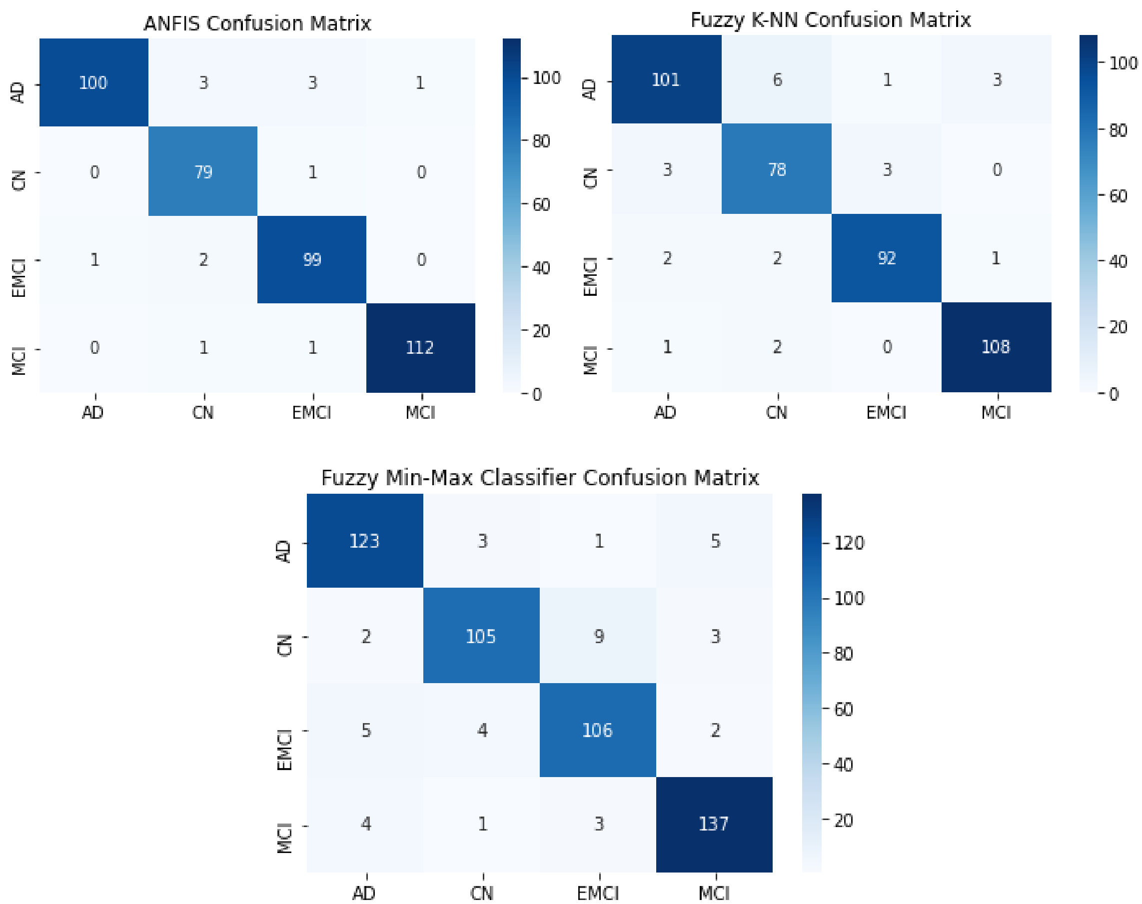


Figure 3.9: Confusion Matrix for the Models Proposed

3.8 Model Evaluation and Heatmap

The proposed models were evaluated using 10-k-fold cross validation. In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples, here k=10. Of the k (10) subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 (i.e., 9) subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once.

The result of 10-fold cross validation is presented in Figure 12.

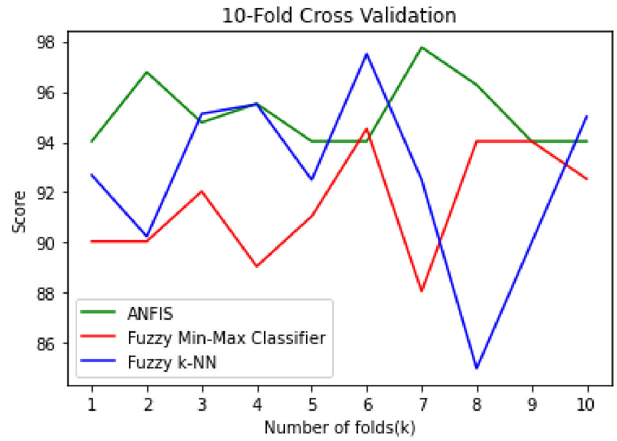


Figure 3.10: Result of 10-fold Cross Validation

Method	Average Accuracy	Best Accuracy
ANFIS	95.12%	97.76%
Fuzzy Min-Max-Classifier	91.54%	94.53%
Fuzzy k-NN	93.05%	100%

Table 8: Comparative Analysis of 10-fold Cross Validation

Grad-Cam Feature Heatmap:

While deep learning has enabled exceptional accuracy in image classification, object recognition, and picture segmentation, model interpretability, a key component in model interpretation and debugging, is one of their largest issues. To help deep learning practitioners visually debug their models and properly understand where it's "looking" in an image, we introduce the idea of Model Interpretability in order to identify features and the focus of our AI model for a particular decision. It includes all of the methods that make our models' decisions clear and easy to understand. In this study, we take our focus to Gradient-weighted Class Activation Mapping, or more simply, Grad-CAM. Grad-CAM was introduced by researchers in 2017 in the paper 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization' so as to generate class activation maps (CAMs) for certain output of a model.

We can visually confirm where our network is looking with Grad-CAM, ensuring that it is looking at the right patterns in the image and activating around them. Grad-CAM produces a coarse localization map showing the essential regions in the image for predicting the idea by using the gradients of any target concept flowing into the final convolutional layer.

Fig 3.11 shows the feature heatmap of one of the data points used in this study.

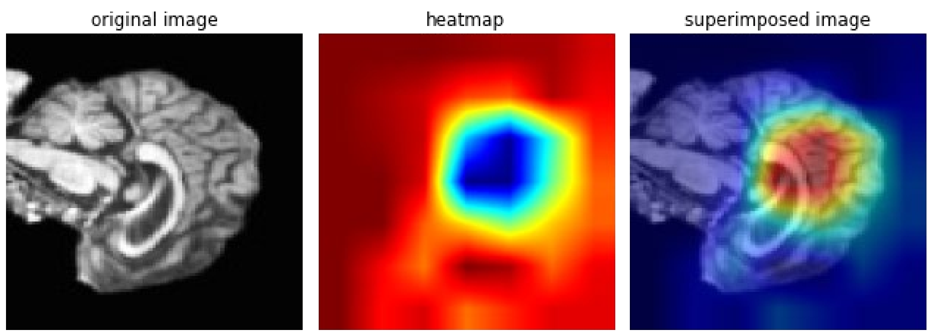


Figure 3.11: Feature Heatmap using Grad-CAM

4. Future Scope and Conclusion

This study introduces the basic concepts of Artificial Intelligence, classification and learning forms of deep learning and fuzzy logic followed by the analysis of fuzzy classification methods on a real example from the medical field. The findings of the experiments suggest that ANFIS, Fuzzy Min-Max Classifier, and Fuzzy k-NN may all be employed as expert systems in medical diagnostics/or in classification techniques. However, after k-fold cross validation, the accuracy of ANFIS and Fuzzy K-NN is found to be superior to that of the Fuzzy Min-Max Classifier. These methods could make fuzzy if-then rules directly from training patterns without having to spend time tuning them. Because the data set is small, the system can be tested for a larger data set to see how it behaves. The multi-class classification problem was used in this comparative study. This research could be expanded to look at the progression of Alzheimer's disease and compute its efficiency in this regard.

Further, although the current study focuses on the comparative analysis, however, it can be converted into pipeline using the most effective model and relevant front-end and back-end frameworks to diagnose Alzheimer's Disease using real-time data. This model can assist the experts in accurately interpreting medical big data so that the accuracy and consistency of diagnosis can be improved and reduce the analysis time.

Fuzzy classification techniques can be deployed in real word environments and tech giants can use it to strengthen their current AI environments. This is mainly due to the following reasons:

1. Fuzzy logic solves complex problems by enhancing its capability to accomplish human-like decision-making and reasoning tasks.
2. It addresses engineering ambiguities and can accommodate a variety of inputs, including data that is unclear, distorted, or inaccurate.

Nonetheless, AI will continue to be pursued by businesses in the future to improve their decision-making processes. They will become more adaptable and more responsive to changes in the ecosystem if they adopt these strategies quickly.

References

Acknowledgment: The enduring pages of the case study are the cumulative sequence of extensive guidance and arduous work. I wish to acknowledge and express my personal gratitude to Mr. Venkatesh Gauri Shankar (Assistant Professor, Information Technology, Manipal University Jaipur), for his exemplary guidance and encouragement.

[1] *Neurological Disorders*. (n.d.). Montana DPHSS from https://dphhs.mt.gov/schoolhealth/chronichealth/neurologicaldisorders

[2] Facts and Figures. (n.d.). Alzheimer's Disease and Dementia from https://www.alz.org/alzheimers-dementia/facts-figures

[3] What is Alzheimer's Disease? | CDC. (n.d.). Centers for Disease Control and Prevention from https://www.cdc.gov/aging/aginginfo/alzheimers.htm#burden

[4] What is Alzheimer's? (n.d.). Alzheimer's Disease and Dementia from https://www.alz.org/alzheimers-dementia/what-is-alzheimers

[5] Medical Tests. (n.d.). Alzheimer's Disease and Dementia. from https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests

[6] ADNI | About. (n.d.). ADNI from http://adni.loni.usc.edu/about/

[7] Fan, Z. (2019d, September 13). Classification of Alzheimer's disease based on brain MRI and machine learning. Neural Computing and Applications. https://link.springer.com/article/10.1007/s00521-019-04495-0

[8] Farhan, S. (2014, September 9). An Ensemble-of-Classifiers Based Approach for Early Diagnosis of Alzheimer's Disease: Classification Using Structural Features of Brain Images. Hindawi. https://www.hindawi.com/journals/cmmm/2014/862307/

[9] S.M. Smith. Fast robust automated brain extraction. Human Brain Mapping, 17(3):143-155, November 2002.

[10] A review on noise reduction methods for brain MRI images. (2015, January 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/7058284

[11] Ellingson, B. M. (2012, June 1). Comparison between intensity normalization techniques for dynamic susceptibility contrast (DSC)â□□MRI estimates of cerebral blood volume (CBV) in human gliomas. Wiley Online Library. https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.23600

[12] Brownlee, J. (2020, August 18). Transfer Learning in Keras with Computer Vision Models. Machine Learning Mastery. https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/

[13] Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017, May 16). Linear discriminant analysis: A detailed tutorial. ResearchGate. https://www.researchgate.net/publication/316994943_Linear_discriminant_analysis_A_detailed_tutorial

[14] Lodha, P., Talele, A., & Degaonkar, K. (2018, August 1). Diagnosis of Alzheimer's Disease Using Machine Learning. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8697386

[15] Vadakkepat, P. (2010, September 15). Boosting Based Fuzzy-Rough Pattern Classifier. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-642-15810-0_39

[16] Oh, K., Chung, Y., Kim, K. W., Kim, W., & Oh, I. (2019, December 3). Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning. Scientific Reports. https://www.nature.com/articles/s41598-019-54548-6

[17] Niu, K. (2020, January 31). Multichannel Deep Attention Neural Networks for the Classification of Autism Spectrum Disorder Using Neuroimaging and Personal Characteristic Data. Hindawi. https://www.hindawi.com/journals/complexity/2020/1357853/

[18] Kour, H. (2019, March 1). Evaluation of Adaptive Neuro-Fuzzy Inference System with Artificial Neural Network and Fuzzy Logic in Diagnosis of Alzheimer Disease. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8991423

[19] Gao, J. (2020). Multisite Autism Spectrum Disorder Classification Using Convolutional Neural Network Classifier and Individual Morphological Brain Networks. PubMed Central (PMC). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7877487/

[20] Mayo Clinic. (2020, December 29). Alzheimer's disease - Diagnosis and treatment - Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/diagnosis-treatment/drc-20350453

[21] Holeček, P., Talasová, J. and Stoklasa, J. (PDF) fuzzy classification systems and their applications, Research Gate. https://www.researchgate.net/publication/282777911_Fuzzy_classification_systems_and_their_applications

[22] Goyal, K. (no date) Fuzzy logic in artificial intelligence: Architecture Upgrad. Available at: https://www.upgrad.com/bloguzzy-login-in-artificial-intelligence/

[23] Wayner, P. (2022) What is artificial intelligence classification? VentureBeat. VentureBeat. Available at: https://venturebeat.com/ai/what-is-artificial-intelligence-classification/

[24] What is Artificial Intelligence (AI) (no date) Gartner. Available at: https://www.gartner.com/en/topics/artificial-intelligence

[25] Gavrilova, Y. (2020) AI vs. ML vs. DL: What's the difference, Serokell Software Development Company. Serokell. Available at: https://serokell.io/blog/ai-ml-dl-difference