



ARTICLE

Improving Multiple Sclerosis Disease Prediction Using Hybrid Deep Learning Model

Stephen Ojo¹, Moez Krichen^{2,3,*}, Meznah A. Alamro⁴, Alaeddine Mihoub⁵,
Gabriel Avelino Sampedro⁶ and Jaroslava Kniezova^{7,*}

¹Department of Electrical and Computer Engineering, College of Engineering, Anderson University, Anderson, SC 29621, USA

²Faculty of Computing and Information, Al-Baha University, Al-Baha, 65528, Saudi Arabia

³ReDCAD Laboratory, University of Sfax, Sfax, 3038, Tunisia

⁴Department of Information Technology, College of Computer & Information Science, Princess Nourah Bint Abdul Rahman University, Riyadh, 11564, Saudi Arabia

⁵Department of Management Information Systems, College of Business and Economics, Qassim University, P.O. Box 6640, Buraidah, 51452, Saudi Arabia

⁶School of Management and Information Technology, De La Salle-College of Saint Benilde, Manila, 1004, Philippines

⁷Department of Information Management and Business Systems, Faculty of Management, Comenius University Bratislava, Bratislava 25, 82005, Slovakia

*Corresponding Authors: Moez Krichen. Email: moez.krichen@redcad.org; Jaroslava Kniezova. Email: jaroslava.kniezova@fm.uniba.sk

Received: 25 March 2024 Accepted: 31 July 2024 Published: 15 October 2024

ABSTRACT

Myelin damage and a wide range of symptoms are caused by the immune system targeting the central nervous system in Multiple Sclerosis (MS), a chronic autoimmune neurological condition. It disrupts signals between the brain and body, causing symptoms including tiredness, muscle weakness, and difficulty with memory and balance. Traditional methods for detecting MS are less precise and time-consuming, which is a major gap in addressing this problem. This gap has motivated the investigation of new methods to improve MS detection consistency and accuracy. This paper proposed a novel approach named *FAD* consisting of Deep Neural Network (DNN) fused with an Artificial Neural Network (ANN) to detect MS with more efficiency and accuracy, utilizing regularization and combat over-fitting. We use gene expression data for MS research in the GEO GSE17048 dataset. The dataset is preprocessed by performing encoding, standardization using min-max-scaler, and feature selection using Recursive Feature Elimination with Cross-Validation (RFECV) to optimize and refine the dataset. Meanwhile, for experimenting with the dataset, another deep-learning hybrid model is integrated with different ML models, including Random Forest (RF), Gradient Boosting (GB), XGBoost (XGB), K-Nearest Neighbors (KNN) and Decision Tree (DT). Results reveal that *FAD* performed exceptionally well on the dataset, which was evident with an accuracy of 96.55% and an F1-score of 96.71%. The use of the proposed *FAD* approach helps in achieving remarkable results with better accuracy than previous studies.

KEYWORDS

Multi Sclerosis (MS); machine learning; deep learning; artificial neural network; healthcare



1 Introduction

In recent years, there has been a significant surge in medical advancements, particularly in successfully treating MS [1]. These breakthroughs mark a pivotal moment in the medical landscape, offering new and innovative ideas to improve the management of MS. With continuous progress in medical research, our understanding of this complex neurological disorder is becoming more sophisticated, paving the way for more effective and targeted therapeutic methods. MS is a debilitating neurological condition that affects the central nervous system. It occurs when the immune system mistakenly attacks the protective covering of nerve fibers, known as the myelin sheath [2]. Consequently, this disrupts communication between the brain and body. The results of this disruption can include fatigue, visual impairments, tingling or numbness, weakness in muscles, coordination difficulties, and sensory abnormalities. Diagnosis of MS typically involves imaging tests and clinical evaluations. While there is currently no known cure for MS, various treatments, such as medication and physical therapy, are utilized to manage symptoms, slow disease progression, and enhance the overall quality of life for those living with MS. There exists a dire need to increase the detection efficiency of MS for efficiently dealing with its diagnosis. Previous methods exist, but with several limitations; their complexity is high to achieve higher accuracy. This high complexity and low results are the obstacles in the diagnosis of MS. Studies show that multi-sclerosis is increasing with time in different areas of the world; there is a crucial need to understand its dynamics to address this rising chronic condition. Previous approaches were conventional and not efficient for diagnosing MS, showing that these techniques are inefficient. This approach of multi-sclerosis detection, including other attempts at diagnosis, emerges as an authentic weapon, and quick detection and differentiation of the disease come at the pinnacle of success. In addition, the development of this approach led to a more precise diagnosis. In addition, early management is done so that the patient's situation can be improved, and the prognosis will be enhanced.

1.1 Motivation

Millions of people worldwide have MS, a chronic and disabling autoimmune disease [3]. Research in this area is motivated by various factors, including the need to understand the disease better, develop effective treatment strategies, and improve the quality of life for affected individuals. The subtle intricacies of MS, an enduring and chronic autoimmune disease, have caused many researchers to start exploring it due to the ever-growing rate of this condition [4]. Recent statistics indicate an increasing trend necessitating deeper inquiry into the complexities of the disease. It is not just about numbers; MS can be devastating for patients globally. This encompasses mobility issues, cognitive decline, and emotional turmoil's [5]. Hence, researchers must look beyond causation and create a new approach that will provide promising solutions that will significantly enhance the quality of life among victims. Although we have come to discover much about MS, there is still a lot of work that needs to be done in terms of improving the current therapies and understanding how the disease progresses [6]. In this way, our investigation represents a notable progression in science by integrating a FAD using a multi-model approach. ANN supports complicated pattern identification through this, while the deep-learning hybrid model offers a strong foundation for combining numerous data sources. This effective merger exploits the advantages of each model. Therefore, through concerted efforts, we expect to get profound insights into the intricacies of multi-sclerosis and provide steppingstones for targeted and successful interventions [7]. Our studies employ expert computational models to widen our comprehension of how multi-sclerosis is born and develops. This method of treatment gives hope

for better care and results in patients with this multifaceted neurological disorder, thereby initiating a new era of precision medicine specifically designed for the unique complexities associated with MS.

1.2 Contributions

- Proposed a novel approach named FAD consisting of ANN-DNN for multi-sclerosis detection that significantly improves MS detection by using Machine Learning (ML) and Deep Learning (DL) multimodal techniques. This work substantially improves the previous method as this approach increases efficiency by increasing accuracy and precision.
- The gene expression data utilized for our research on MS is available to the public through the GEO GSE17048 dataset. We preprocessed the data by encoding, standardizing, and selecting features to improve and optimize the dataset.
- This study provides. It performed better than the previous studies and the other deep-learning hybrid model approach. The results provide an impressive accuracy of 96.55%.

1.3 Organization

The article is divided into five main sections. [Section 1](#) briefly introduces MS and its effects on the human mind, and it also contains an abbreviation ([Table 1](#)) for understanding technical terms in the research being conducted. [Section 2](#) delves into the research surrounding MS detection and its advancements. The following section, [Section 3](#), outlines the specific methods and techniques used to achieve improved outcomes. Moving on, [Section 4](#) presents the Experimental Analysis and Results. [Section 5](#) provides the discussion and comparison. Lastly, [Section 6](#) wraps up the article with conclusions and recommendations for future research.

Table 1: List of abbreviations

Acronym	Abbreviations
SVM	Support Vector Machines
FAD	Fused ANN DNN
DLHM	Deep Learning Hybrid Model
RFECV	Recursive Feature Elimination and Cross-Validation
DT	Decision Tree
GB	Gradient Boosting
DT	Decision Tree
RF	Random Forest
XGB	XGBoost
KNN	K-Nearest Neighbors
MS	Multi Sclerosis
MRI	Magnetic Resonance Imaging
DL	Deep Learning
PBMC	Peripheral Blood Mononuclear Cells
OCT	Optical Coherence Tomography
ANN	Artificial Neural Network
AI	Artificial Intelligence
LSTM	Long Short-Term Memory

(Continued)

Table 1 (continued)

Acronym	Abbreviations
DNN	Deep Neural Network
NB	Naive Bayes
ReLU	Rectified Linear Unit

2 Literature Review

Traditional methods have long been the norm for diagnosing MS. However, groundbreaking research from 2011 to 2024 has uncovered the potential of using ML and DL to revolutionize the diagnostic process. This approach includes advanced algorithms like Convolutional Neural Networks and Support Vector Machines, which have shown promise in accurately analyzing Magnetic Resonance Imaging (MRI) scans and clinical data, outperforming human abilities. The goal is to improve diagnostic accuracy. Still, this technological advancement also carries the potential for creating personalized treatment plans that could lead to earlier interventions and, ultimately, reduce medical costs. Despite these exciting prospects, significant challenges still need to be addressed. To fully capitalize on the potential of ML and DL for diagnosing MS, it is crucial to continually seek out large and reliable datasets. Equally important is making these advanced models accessible and understandable for medical professionals.

De Leon-Sanchez et al. [8] introduced an innovative DL approach that utilizes a single hidden layer ANN to diagnose MS accurately. It will set our model apart in utilizing gene expression profiles, specifically 74 genes, as input features. Unlike traditional ML methods, this strategy improves accuracy and reduces complexity. To further enhance the model's performance and maintain its simplicity, we have integrated a regularization term into the single hidden layer of the ANN architecture to prevent over-fitting. By conducting a comprehensive assessment, the study effectively evaluated the proposed ANN model against four prominent ML algorithms-RF, Support Vector Machine (SVM), KNN, and Naive Bayes (NB). The dataset, consisting of 144 individuals (45 healthy controls and 99 MS patients), underwent feature selection using a powerful dimensionality reduction technique to pinpoint key genes. The results were remarkable, showcasing the ANN model's superiority in recall (80.80%), accuracy (83.33%), precision (85.71%), and F1-score (0.83%). Aslam et al. [9] delved into the multitude of possibilities that ML and DL can bring to the field of MS diagnosis while also acknowledging and addressing the challenges that must be overcome for their responsible and successful integration. This acknowledgment serves as a launching point for further research and progress, ensuring that these powerful tools are utilized ethically and effectively to enhance the lives of those affected by MS.

Maleki et al. [10] applied Artificial Intelligence (AI) to the analysis of retinal images, which is a way to improve MS management. This method is based on Optical Coherence Tomography (OCT) scans, with AI algorithms that are trained through ML and DL techniques to reveal and classify the abnormalities that are MS indicators. Also, these algorithms can predict the disease progression, individualize the treatment plans and classify the MS subtypes according to the retinal features; thus, potentially, it will be easier for the physicians. Although the research shows the possibilities of improving MS diagnosis, monitoring and treatment planning by increasing the image analysis accuracy and efficiency of OCT, the study also recognizes the difficulties, such as the ethical considerations and the need for more research to be done before the use can be applied to the whole clinical practice.

Shoeibi et al. [11] delved into the potential of ML to differentiate individuals with MS from healthy individuals utilizing data from an instrumented walkway. The researchers hypothesized that ML algorithms could more effectively classify MS patients using raw gait data instead of traditional methods relying on predetermined features. To put this theory to the test, they recruited a group of 72 participants and collected data, such as pressure distribution and foot trajectory, from the walkway's sensors. The results revealed that the ML models accurately identified MS patients. Specifically, Support Vector Machines (SVM), Logistic Regression, and XGB demonstrated remarkable accuracy, with SVM boasting an 81% success rate. Significantly, including novel features derived directly from the raw gait data boosted the accuracy rate to 88%. This approach highlights the impressive capabilities of ML in accurately detecting MS through careful analysis of raw data. This research suggests the potential for surpassing traditional approaches that rely on predetermined features.

Daqqaq et al. [12] showed how DL can be the source of the early detection of MS by training Convolutional Neural Networks (CNNs) to perform MRI scans for indicative patterns. Although the correctness of the results is guaranteed, the problems that must be solved are the big, labeled datasets, the way of explaining the decisions of the model and the proper organization of the data from different sources. The patient's age, race and MRI variations should be taken into consideration while ensuring model compatibility with varied patient demographics. Despite difficulties, methods like transfer learning and data augmentation are proving to be the future of AI. In the end, the MS diagnosis through DL can be used to detect the disease at its early stage, to design personalized treatments, and to follow the disease progression; thus, the patient outcomes are improved.

Denissen et al. [13] introduced an innovative approach to enhancing the accurate prediction of distinctive cognitive changes in individuals with MS. A multidimensional ML model was developed through federated learning methods by merging information from diverse sources, including brain MRI scans, clinical assessments, and genetic data. The team successfully integrated these different forms of data to anticipate the cognitive trajectory of individual patients in the future. This model outperformed previous ones that solely relied on singular data sources, resulting in a more comprehensive comprehension of the disease's progression and generating more precise predictions. The utilization of federated learning played a pivotal role in addressing privacy issues by facilitating efficient model training with decentralized patient data. The study's results demonstrate the potential of multimodal ML in accurately predicting the rate of cognitive decline for each patient with MS. This offers promising prospects for implementing early intervention and personalized treatment plans to minimize cognitive impairment in MS patients.

Imaging and analyzing MS through currently used methods are not very accurate, and that is why a novel method of detection is required. Conventional approaches are not precise enough; rather, they use features that are already known in advance. At the same time, ML and DL face issues with the availability of data used and their interpretability. The new technique delivers exclusive strategies, which are based on gene expression profiles for discriminating, regularization using FAD, and analyzing the raw gait data for classification purposes. A new approach is proposed that uses FAD for multi-sclerosis detection with more precise, better, reliable, and personalized ways to diagnose and prognosis MS.

3 Proposed Approach

This paper proposes a fused approach named FAD consisting of fused ANN [14] and DNN [9] algorithms to tackle the detection of MS with more efficiency and accuracy, utilizing regularization and a single hidden layer in an ANN to combat over-fitting. This innovative approach incorporates

AI to enhance MS classification accuracy, showcasing its immense potential. Using a FAD helps achieve remarkable results with better accuracy than previous studies. Fig. 1 visually represents the proposed approach. Integrating regularization techniques within the fused ANN-DNN framework effectively tackles over-fitting concerns, while its streamlined hidden layers design balances efficiency and complexity. This advanced approach has revolutionized our understanding of tackling complex neurological disorders such as MS and the potential impact of AI in medicine. The experimental validation of this approach further highlights its potential to improve diagnostic accuracy and personalizing treatments for individuals with MS.

$$final_{output} = \sigma(a2_{dropout} \cdot W_{final} + b_{final}) \quad (1)$$

In Eq. (1), the sigmoid activation function is symbolized by σ and $a2_{dropout}$ shows the output of the second hidden layer after applying dropout, and $W_{final} + b_{final}$ are the weight and bias terms of the output layer. Although previous techniques had shown success, they eventually became less precise, prompting the adoption of more advanced methods such as single-hidden-layer ANN. This cutting-edge approach aims to enhance accuracy and precision, ultimately improving results when addressing current challenges. The ultimate objective is to elevate the model's performance and generate more refined and precise outcomes than previous methods by harnessing the DL capabilities of an ANN and DNN framework. To stay ahead of the curve in the ever-evolving world of AI and neural network technology, our strategic shift reflects a firm dedication to incorporating the latest advancements in pursuit of optimal results.

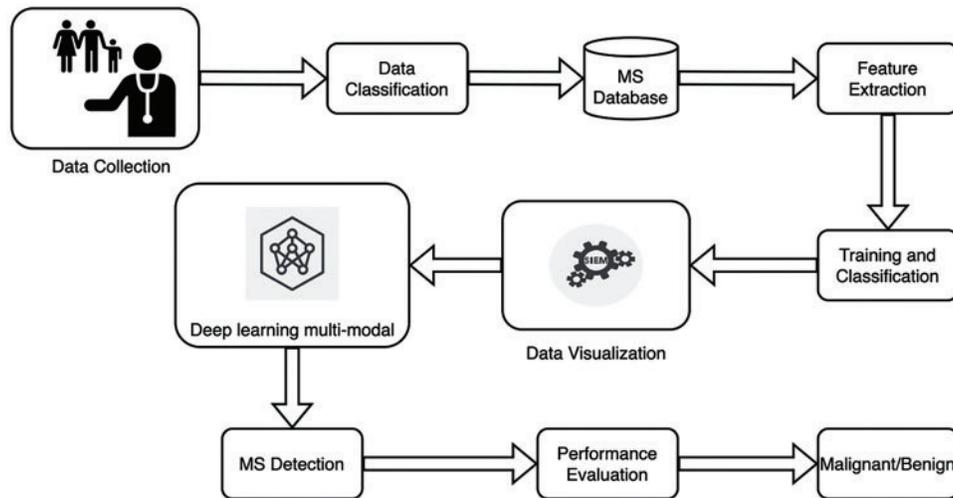


Figure 1: Workflow of proposed FAD approach for MS detection

3.1 Dataset

This study uses a microarray dataset and an external blood T-cell microarray dataset that are combined in a dataset, which pertains to Peripheral Blood Mononuclear Cells (PBMC) [15]. Gene expression data for MS research can be found in the GEO GSE17048 dataset, which is accessible to the public [16]. The dataset includes information from 144 people, 45 of whom are healthy and 99 of whom have MS. They concentrate on 74 genes, including 73 genes found in earlier studies and the HLA-DRB1 gene, which is linked to MS risk. The GEO2R network tool was utilized for analysis, while the GPL6947 platform was utilized for data acquisition.

3.2 Feature Extraction and Selection

One of the basic dimensionality reduction methods in ML and DL is feature selection. The selection of parameters has the potential to greatly enhance the learning model’s performance by improving the computational efficiency of the regression or classification algorithm. The research methodology is developed using the RFECV method, which formally applies to encourage the flexibility of the models being used [17]. The key to changing feature selection with this new method is carefully preserving features while iteratively training our model to create our feature landscape. A thorough assessment of the model’s efficacy is ensured by including cross-validation and RFECV, effortlessly integrated into our research methodology, more conventional methods, and dynamically evolving feature subsets.

The ML model employed to investigate the importance of features in this line of code is called the RF classifier, or clf. The computed logic is arranged neatly in descending order in a dataframe. Then, the less important component is removed. The curated list generated and saved in the variable *selected_feature* contains the most important features determined by the model, as shown in Fig. 2. The features are shown on x-axis and importance of each feature is shown on y-axis where highest the value the better the feature. This analytical stage can guide intelligent product selection for analysis on the next try, facilitate fine-grained result interpretation, and help understand the model’s underlying decision logic.

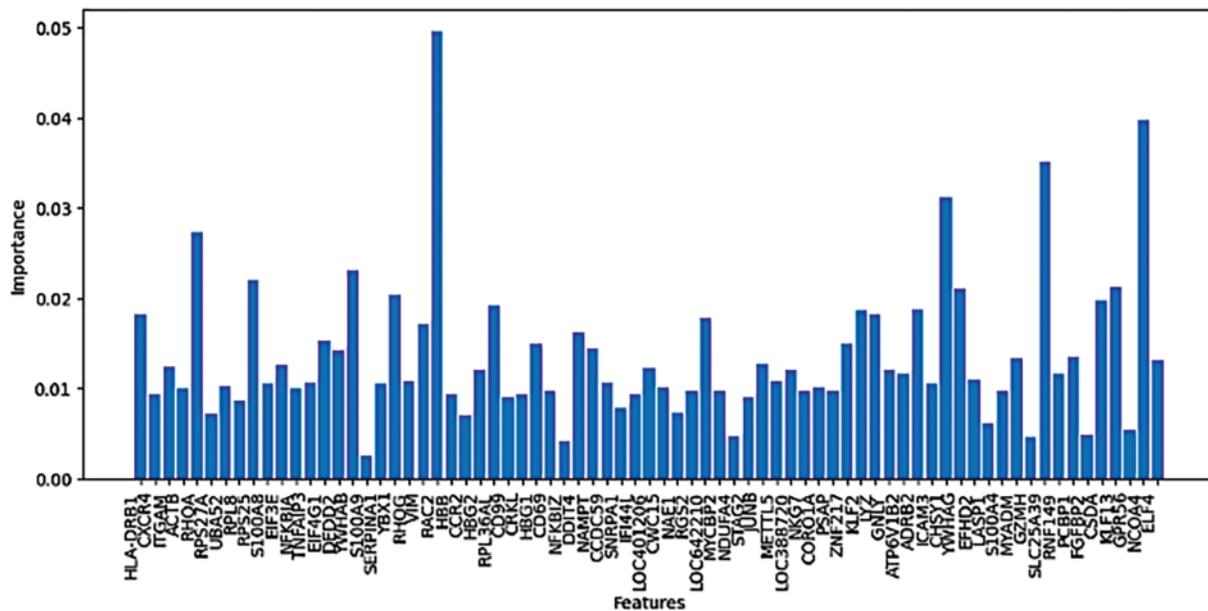


Figure 2: Important features of dataset

The RFECV model is built with 5-fold cross-validation, a step size of 1, and parameters that define accuracy as the scoring metric. The base estimator that has been chosen is RF classifier. RFECV uses a RF partition to iteratively find the least significant features and identify the best sub features. The ‘*X_selected*’ data set is then created by the code extracting the index of the least important features and removing it from the original dataset. The least important feature, representing the selected features, is removed before printing these complexes. Algorithms like backpropagation, which change the weights between neurons to lower the prediction error, are used in the training process. Because DL can learn

in isolation and represent complex data objects, it has been extremely successful in applications like natural language processing, graphic speech recognition, and healthcare diagnostics.

Algorithm 1: Pseudocode of the proposed approach

Require: X_2 (Features), Y_2 (Labels), Test Size = 0.25, Random State = 0

- 1: Split the dataset into training and testing sets: $X_{\text{train}}, X_{\text{test}}, y_{\text{train}}, y_{\text{test}} = \text{train_test_split}(X, Y, \text{test_size} = 0.25, \text{random_state} = 0)$
 - 2: Initialize empty lists: $\text{accuracies}, f1_measure_list, \text{precision_list}, \text{recall_list}$
 - 3: Number of Iterations = 3
 - 4: for i in range (Number of Iterations) do
 - 5: Initialize a model: $\text{model} = \text{training_model}()$
 - 6: Train the model using the current labeled data: $\text{model.fit}(X_{\text{train}}, y_{\text{train}})$:
 - 7: Get the uncertainty scores
 - 8: Get the index of the most uncertain_sample
 - 9: Add the uncertain sample to labeled data
 - 10: Remove the uncertain sample from the test
 - 11: $X_{\text{test}} = X_{\text{test}}.\text{drop}(\text{index} = \text{uncertain_sample})$
 - 12: $y_{\text{test}}.\text{pop}(\text{uncertain_sample})$
 - 13: $y_{\text{pred}} = \text{model.predict}(X_{\text{test}})$
 - 14: Calculate evaluation metrics:
 - 15: $\text{accuracy} = \text{accuracy_score}(y_{\text{test}}, y_{\text{pred}})$
 - 16: $f1_measure = f1_score(y_{\text{test}}, y_{\text{pred}})$
 - 17: $\text{prec} = \text{precision_score}(y_{\text{test}}, y_{\text{pred}})$
 - 18: $\text{recall} = \text{recall_score}(y_{\text{test}}, y_{\text{pred}})$
 - 19: Append the metrics to the respective lists:
 - 20: $\text{accuracies.append}(\text{accuracy})$
 - 21: $f1_score_list.append(f1_Score)$
 - 22: $\text{precision_list.append}(\text{prec})$
 - 23: $\text{recall_list.append}(\text{recall})$
 - 24: Print ($\text{Iteration}i+1$: Accuracy, F1-score, Precision, Recall)
 - 25: end for
 - 26: Get the iteration with the best accuracy, F1-score, precision, and Recall
 - 27: Print the best accuracy, F1-score, Precision, Recall
-

3.3 ML with Hybrid Model

In this segment on ML for MS detection, we employ an array of algorithms, including RF, XGB, GB, KNN, and DT with a hybrid model. This comprehensive approach enables us to thoroughly medical data, thus enhancing the accuracy and effectiveness of MS detection. We aspire to advance diagnostic techniques and improve patient outcomes by combining these diverse models. Its architecture consists of three layers: a merged features layer, a layer for RF predictions, and an input layer for original features. A Long Short-Term Memory (LSTM) layer comes first in the sequence, followed by two dense layers and an embedding layer. Finally, a sigmoid activation function in the output layer makes binary categorization easier, as shown in [Fig. 3](#).

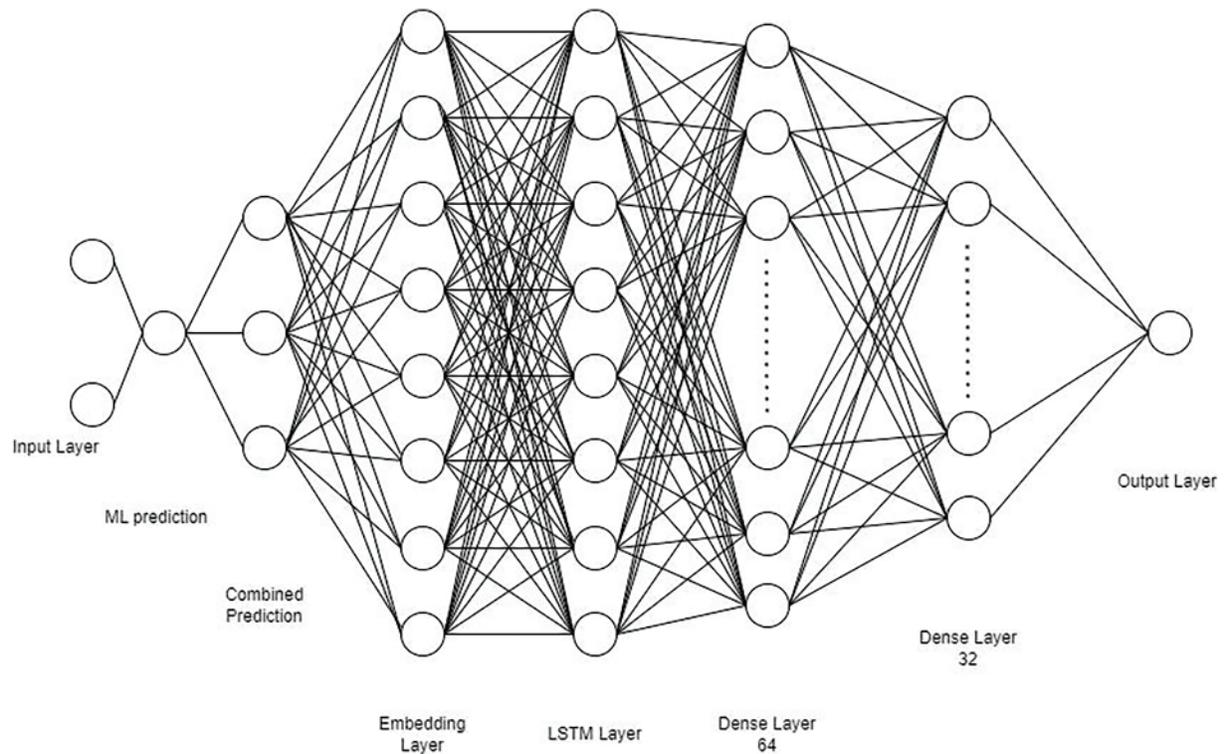


Figure 3: Model architecture of ML with DLHM model

3.3.1 Random Forest (RF) with Deep Learning Hybrid Model (DLHM)

This approach, known as RF with DLHM, blends the substantial representations gleaned by DNN with the cooperative learning capabilities of RF. This hybrid model capitalizes on the strengths of both methods to flexibly adapt to evolving data. Notably, it excels in optimizing diagnostic precision in the context of MS detection, particularly when confronted with limited labeled data. The model continually enhances its understanding by incorporating ambiguous predictions via DL, proving to be a powerful tool for the accurate and timely identification of MS patterns in clinical settings. The synergy of RF and DL in this hybrid model leads to superior performance in MS detection. The RF classifier has been trained with default parameters While the neural network that we have used comprises an LSTM layer with an 8-unit Rectified Linear Unit (ReLU) activation function, a dense layer with 64 ReLU activation function, another dense layer with 32 ReLU activation function and output layer with a single Sigmoid activation neuron. Although multi-layered structures are proposed, dropout layers are recommended but not applied. This model is trained using Adam optimizer and binary cross entropy loss across 50 epoch and batch size of 32 and then evaluated using accuracy, precision, recall, F1-score criteria.

3.3.2 Gradient Boosting (GB) with DLHM

This approach blends with DLHM to enhance the boosting capabilities with refined feature learning. The hybrid model's flexibility boosts diagnostic accuracy, particularly when handling limited labeled data for MS detection. The model iteratively refines its understanding by incorporating uncertain predictions through active learning. It is a powerful tool for precise and timely recognition of

MS patterns in clinical settings. This combination showcases a concise yet effective technique to elevate the performance of ML models in medical contexts. The gradient boosting classifier is employed with default parameters and the output generated from this is used as feature engineering. All these features are padded and reshaped into a matrix of shapes that can be fed into the LSTM layer. This neural network consists of an LSTM layer with 8 units, ReLU activation, and fully connected layers with 64 and 32 units, ReLU activation, and the output layer with one neuron, sigmoid activation. It was recommended to use dropout layers, but this layer was not applied in the network. This model used the Adam optimizer together with binary cross-entropy loss function with the training process being performed for 50 epochs with a batch size of 32, and performance evaluation metrics included accuracies, precision, recolonization, and F1-score.

3.3.3 *XGB with DLHM*

This approach merges the strengths of two methods: XGB's enhanced boosting techniques and DLHM enriched feature representations. The result is a hybrid model that adapts dynamically, leading to better diagnostic accuracy, especially when dealing with a scarcity of labeled data for MS detection. The model's ability to integrate uncertain predictions through active learning continuously enhances its understanding, making it a reliable tool for accurately and timely identifying MS patterns in clinical settings. Combining XGB and DLHM improves the performance and versatility of ML models. Specifically, the XGB classifier is trained by default, and its outputs are utilized as extra categories. These combined features are padded and reshaped as required for LSTM input. They applied an LSTM layer with 8 neurons; the next layers were dense layers, with 64 and 32 neurons activating ReLU; and the output layer had a neuron with sigmoid activation. However, the recommended dropout layers are not utilized here. The proposed model uses the Adam optimizer and the binary cross-entropy loss function implemented for 50 epochs for 32 batch size and uses accurate, precise, recall, and F1-score measurement frameworks.

3.3.4 *K-Nearest Neighbors (KNN) with DLHM*

This approach seamlessly blends the strengths of KNN and DLHM to achieve superior performance in MS detection. The model's adaptability allows it to excel even in scenarios with limited labeled data. Its continuous learning capability enhances diagnostic accuracy by actively incorporating uncertain predictions, making it a dependable tool for the timely and precise identification of MS patterns in clinical settings. This ingenious integration of KNN and DLHM showcases a straightforward yet effective technique for boosting the efficiency and adaptability of ML models, demonstrating its potential for widespread use across various domains. The KNN model is trained with default hyper parameters, and the predictions resulting from this are added as new variables. All these features are padded and reshaped to suit the input dimension expected of the LSTM layer. The architecture of the neural network comprises of an LSTM layer that has 8 neurons with ReLU activation, two fully connected layers with ReLU activation, one with 64 neurons and the other with 32 neurons, and an output layer with one neuron having sigmoid activation. While it may be daunting to see that dropout layers are recommended, none of it is implemented in the final model. The neural network is compiled with Adam optimizer, binary cross-entropy loss function to train the model through 50 epochs, 32 batch size, and then test the model on the test data set using accuracy, precision, recall, F1-score.

3.3.5 *Decision Tree (DT) with DLHM*

This improved hybrid model merges DT and DLHM for diagnosing MS efficiently and accurately. The model's strength lies in combining the interpretability of DT with the sophisticated feature

learning abilities of DLHM. Its flexibility allows it to handle sporadic labeled data and generate reliable predictions by incorporating uncertain predictions through active learning. This makes it a valuable tool for clinical settings, where precise detection and clear-cut, understandable methods are crucial. It offers a powerful machine-learning solution that aids in accurate MS detection while being comprehensible for medical professionals. The DT classifier is then fitted with the default parameters and the predictions obtained are used as the features. These combined features are padded and reshaped to fit the dimensions of the LSTM layer's input. The network is comprised of LSTM with 8 neurons (ReLU activation), and then two dense layers, with 64 and 32 neuronal units (ReLU activation), as well as the final output layer with a single neuron (sigmoid activation). However, for this setup there are some dropout layers suggested but they are not used in this context. To train the model we use the Adam optimizer and binary cross-entropy loss, we train the model for 50 epochs utilizing a batch size of 32 and the model is evaluated by the accuracy, precision, recall, F1-score metrics.

3.4 Deep Learning with Hybrid Model

In ML, DL is a crucial task that entails developing and deploying applications, particularly deep roots. Adding additional layers of deep architectures to these networks is what the “Deep” aspect means. Because of the way the human brain is structured, neurons are interconnected, multilayered networks of inputs, stores, and outputs. Enabling the model is the main objective of DL. These models are very good at extracting the hierarchy of objects and images through deep architecture, making it possible to view intricate shapes.

Neurons in the input layer of the architecture are oriented according to the dataset's properties. The 53 neurons in Hidden Layers 1 and 2 add to the intricacy. Overfitting is prevented by dropout layers in both layers with a 0.1 rate. One neuron in the output layer designed for binary classification has a sigmoid activation function. This approach uses dropout layers to improve generalization while balancing resilience and complexity, as shown in Fig. 4. It has two hidden layers, which have 53 neurons, each having ‘tanh’ activation with normal kernel initializer and has $L2 = 0.0$ on the regularizer. This is followed by adding ‘1’ to each of the nodes in the succeeding hidden layers: The hybrid model used the last fully connected layer followed by a ‘sigmoid’ activation function for binary classification. They use the Adam optimizer with a learning rate of 0.01, binary cross-entropy loss function for training and 50 epochs with a batch size of 32. Model selection is performed using model checkpoint through the best validation accuracy, and the metrics for the best model include accuracy and log loss.

3.4.1 Long Short-Term Memory (LSTM)

DL approach, particularly with Long Short-Term Memory (LSTM) networks, is a game-changing tool for handling data that unfolds over time. LSTMs, a specialized form of recurrent neural networks, excel at capturing long-term dependencies and patterns hidden within sequential data. This powerful technology is being leveraged in the medical realm to advance the diagnosis of MS. LSTM, an ANN, improves the ability to detect MS by optimizing sensitivity and precision. This research establishes a new benchmark for using LSTMs and DL in the intricate task of MS pattern recognition. It demonstrates how using these sophisticated models can greatly enhance the accuracy of neurodegenerative disease diagnoses.

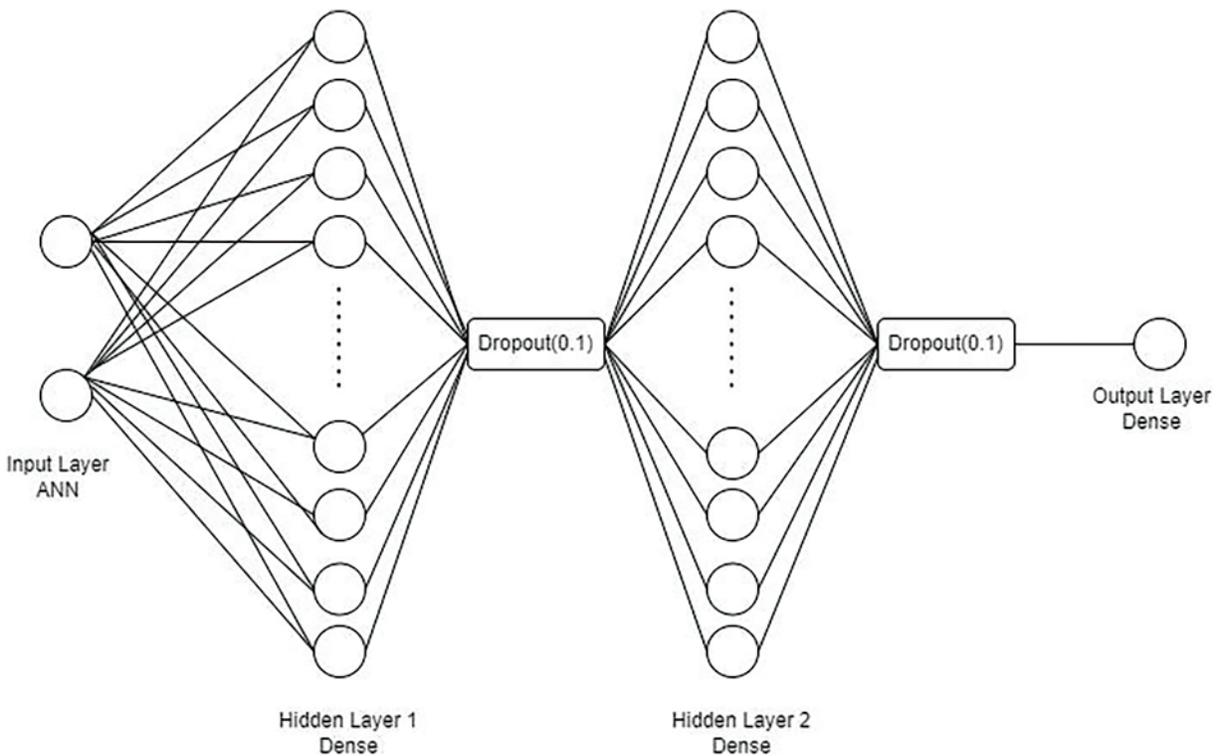


Figure 4: Model architecture of ANN with hybrid model

3.4.2 Fused ANN-DNN (FAD)

FAD is a fusion of two DL models, ANN and DNN. The combination of these two models helps in better data processing and gives better results than other ML models used for MS detection. In this technique, the output of the DNN model is the input of the ANN model. This technique helped us achieve better accuracy because of its robustness and efficiency. LSTM layers within ANN architecture enable the FAD model transformation of temporal data since they can detect sequential patterns in the data, thereby helpful when dealing with time-stamped or sequential data sets. Moreover, proper utilization of innovations like L2 regularization and drop-out properly addresses the problem of over fitting and, therefore, the development of a model that has a high degree of generality. Finally, the ideas of the careful standardization or padding of inputs need to be adhered to for uniform processing while preserving temporal interactions and dependencies, as it was mentioned before. The interaction between these factors leads to the FAD model's greater accuracy and F1-scores than the other models.

ANN and DNN have a significant role by using the characteristic whose objective is to identify patterns in medical data. Another aspect illustrates how these neural networks can be applied to increase the level of MS identification. First, the features are normalized utilizing the standard scaler, which helps balance the measures' contributions to the model's performance. The next step in the construction of the model is to create the ANN model with two hidden layers, each of the hidden layers contains 53 neurons and uses the 'tanh' activation function and dropout layer to reduce overfitting of the model. To increase generalization regularized methods are used. The last dense layer activation function, sigmoid function is used in the final layer to produce binary classification such as presence of MS after the integrated outputs of ANN have been further processed. The given hybrid model is

created using the Adam optimizer which modifies the learning rate during training, and the binary cross entropy loss function which is perfect for the classification problems. To rightly select the best model, a Model Checkpoint callback is applied during the training process for the viewport of saving the model with the maximum validation accuracy. From the training process, the validation performance is checked, and the weights must be tweaked to minimize the loss and thus maximize the accuracy. After training the model which has the best accuracy is selected and is tested on the test set to check its efficiency. The performance of the model is evaluated using metrics of accuracy and of the log loss in the case of the binary classification.

3.4.3 *Artificial Neural Network (ANN)*

ML has transformed significantly with the introduction of ANNs, drawing inspiration from the intricate framework of the human brain. In this section, we delve into the fundamental components of ANNs, unraveling their tiered arrangement encompassing input and output layers, hidden layers, weights, biases, and activation functions. Our attention is then turned towards the training mechanisms, highlighting the iterative adjustments made through algorithms like backpropagation. Understanding these mechanisms is crucial in grasping the adaptability of ANNs. The versatility of ANNs is evident through their numerous applications. They have greatly improved image recognition techniques and have been critical in developing natural language processing with sentiment analysis and machine translation tools. Beyond that, ANNs have made significant strides in facial recognition and medical image analysis, transforming how we approach these tasks. Furthermore, their presence is felt in the financial world, where they play a crucial role in evaluating credit risks, predicting stock prices, and detecting fraudulent activity.

3.4.4 *Deep Neural Network (DNN)*

DNNs are a revolutionary development in AI characterized by a complex, multilayered hierarchical structure. In the framework of my work, this section sheds light on the architectural details of DNNs, describing them as a mathematical assembly of non-linear transformations that can extract subtle patterns and features from large, intricate datasets. The investigation goes into the revolutionary uses of DNNs in various fields, which aligns with my research's particular emphasis. DNNs play a key role in computer vision tasks like medical image analysis, object detection, and image classification, all of which are directly related to the main goals of my research. DNNs have proven to be an interesting subject in the research, with applications in robotics, healthcare, finance, and natural language processing offering promising avenues for future investigation. By exploring the intricacies of DNNs, including interpretability and ethical issues, I can better match the possible difficulties of using these cutting-edge technologies with my research objectives. This subsection thoroughly explains DNN applications and demonstrates their relevance to my research. Understanding the complex nature of DNNs is essential for researchers as the field of DL and AI develops, and the knowledge acquired from this section will surely be a useful tool in navigating this cutting-edge area within the research framework being conducted.

4 **Experimental Analysis and Results**

The section dedicated to the experimental analysis and results evaluates the effectiveness of different classifiers for MS detection. It investigates conventional models such as DLHM-RF, DLHM-XGB, and a recently developed FAD. The analysis compares the performance of these classifiers and provides insights into their advantages and disadvantages. It aids in the ongoing development

of accurate and reliable tools for MS diagnosis. In ML, a performance matrix encompasses an array of metrics that guide evaluating and gauging a model's ability to predict outcomes on a given dataset. These metrics include accuracy, precision, recall, F1-score, and other relevant measures, using which one can also analyze the model's generalization capabilities on novel data. Performance metrics aim to objectively assess a model's strengths and weaknesses, enabling practitioners to make informed decisions. Ultimately, this empowers them to ensure that the model fulfills the requirements of the given application. Model accuracy measures how well an ML model classifies instances, expressed as the percentage of correctly classified ones. Model loss is a numerical value reflecting the discrepancy between a model's predicted outputs and targets. Lower values indicate less error and better model performance. In ML, particularly in detection tasks, the performance of models is measured by their accuracy (correct predictions) and loss (total errors). Striking a balance between high accuracy and low loss is crucial for building detection systems that are both accurate and efficient. The model's loss is minimized, as shown in Fig. 5, resulting in a smaller difference between predicted and actual values. This dual trend suggests the FAD continuously learns and optimizes its parameters, leading to improved overall performance.

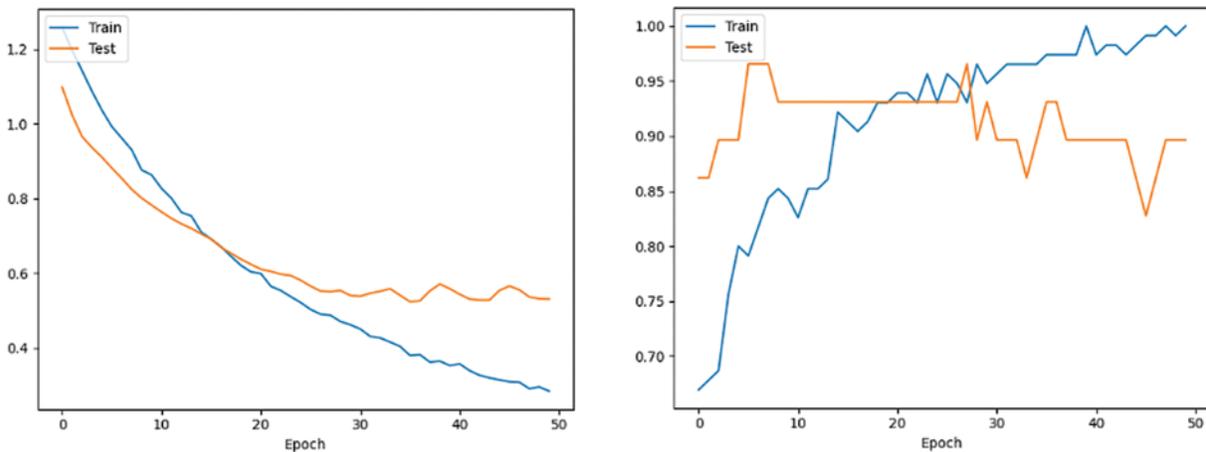


Figure 5: Accuracy and loss curves

Evaluating the performance of a model is crucial, where accuracy serves as a fundamental indicator that measures the percentage of accurately predicted cases. Throughout the model's training process, it is guided by the concept of model loss. This metric represents the difference between the expected and actual values and aids in improving its precision. The Receiver Operating Characteristic (ROC) curve visually represents a model's ability to distinguish between classification thresholds, highlighting the delicate balance between true positives and false positives. Additionally, the confusion matrix offers a comprehensive breakdown of the model's predictions, providing insights on false positives, false negatives, true positives, and true negatives.

This approach relies on the confusion matrix as a crucial tool for evaluating the model's classification performance. The confusion matrix offers a detailed breakdown of the predicted and actual classifications, allowing for a thorough understanding of the strengths and weaknesses of the model. Notably, observations reveal incidences of misclassifications of DLHM-RF in Fig. 6. DLHM-GB and DLHM-RF models contributing to a lower overall accuracy rate. It could be attributed to the data's complexity or specific challenges these models face in accurately recognizing certain patterns. The data in the dataset consists of binary classification, consisting of 2 classes for multi-sclerosis detection. FAD

is remarkable for its unusually high precision in classification. It can discern and adapt to complex data patterns better than DLHM-RF and DLHM-GB in this application. The multi-sclerosis detection framework using FAD underscores its potential as a robust and accurate tool for classification tasks. Understanding the unique features contributing to FAD’s exceptional performance can offer valuable insights for optimizing model selection in similar diagnostic scenarios. In Fig. 7, Receiver Operating Characteristic (ROC) curves explain model performance at various classification levels. For instance, the ROC curve for FAD showed strong performance for all classes.

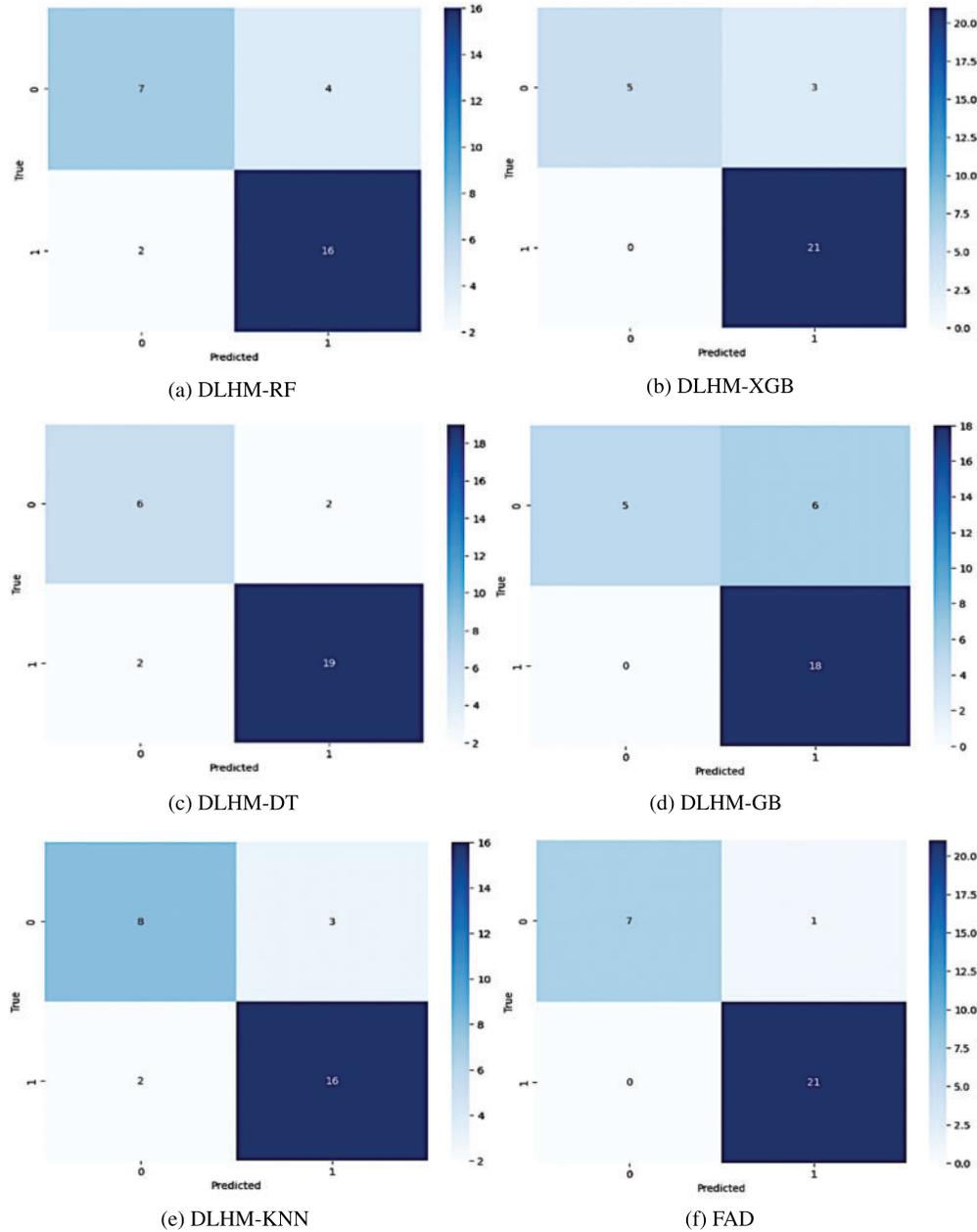


Figure 6: Confusion matrix for ML classifiers

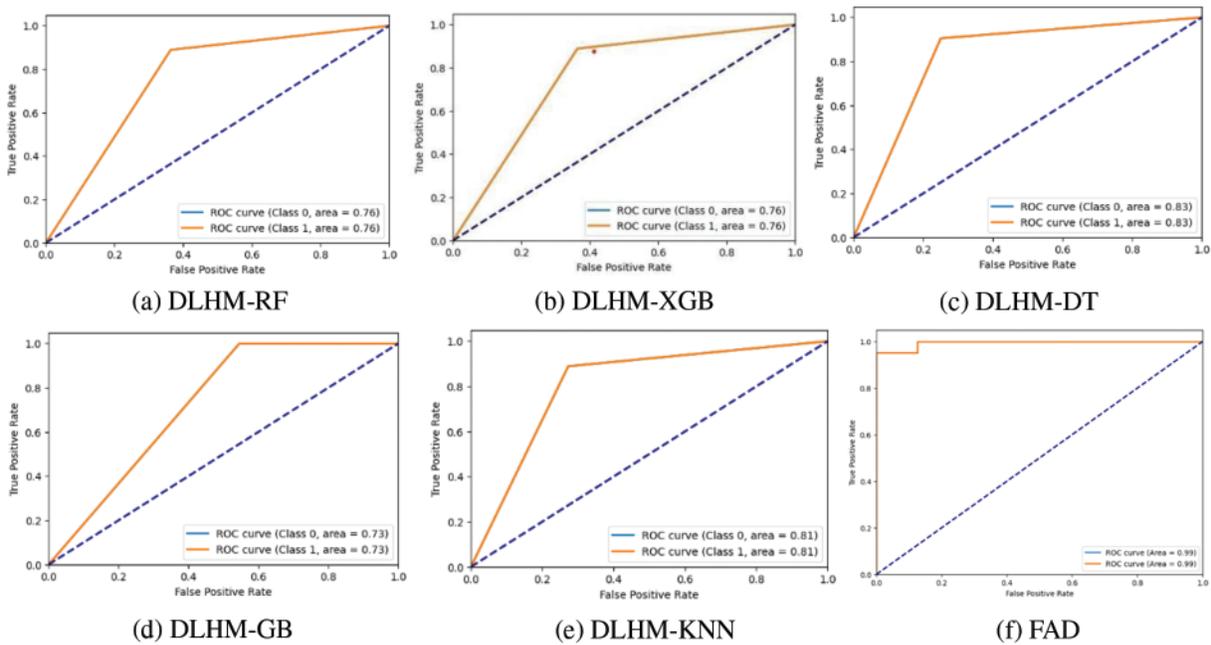


Figure 7: ROC for ML classifiers

We assessed the efficiency of various classifiers in detecting MS using a MS dataset. Models like RF, GB, XGB, KNN and DT were integrated with a deep-learning hybrid model that provided the highest accuracy of 89.65% by XGB. Meanwhile, ANN was fused with another DNN (FAD), which provided even better results, as shown in Table 2. The accuracy, precision, recall, and F1-scores were used to measure each classifier's performance. The results revealed that the efficiency of the classifiers differed. Impressively, the FAD exceeded traditional ML classifiers, achieving an outstanding accuracy of 96.55%. This outcome shows how DL approaches can enhance the accuracy and effectiveness of MS detection systems.

Table 2: Evaluation results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RF-DLHM	79.31	80.00	88.89	84.21
GB-DLHM	79.31	75.00	100	85.71
XGB-DLHM	89.65	87.5	100	93.33
KNN-DLHM	82.76	84.21	98.89	86.49
DT-DLHM	86.21	90.48	90.48	90.48
FAD	96.55	96.48	96.55	96.71

With an accuracy of 96.55%, FAD demonstrated the best overall performance among the traditional classifiers. Given its high recall (96.55%) and precision (96.48%), it is a good option for MS detection. With a balanced precision-recall trade-off and an accuracy of 89.65% accuracy, XGB-DLHM also demonstrated competitive performance.

5 Discussion

The hybrid model's success in integrating ANN can be attributed to its combination of depth and traditional DL strengths. Two methods are proposed for binary classification tasks. First is a traditional neural network approach with dropout layers and standardization for better adaptation and generalization, The FAD model aims to reduce over-fitting, meanwhile achieving an exceptional accuracy of 96.55% surpassing the base approach with the margin of 10.35%. The second method presents a hybrid model that combines different ML models with deep learning hybrid model, which can detect subtle patterns for classification. The choice of method depends on the dataset's features, allowing for a customized approach based on specific requirements. This combination allows the Hybrid Model to identify complex patterns in MS-related data, adapting and providing consistently high performance across various datasets. The ML models implemented with the hybrid model include RF-DLHM, GB-DLHM, XGB-DLHM, KNN-DLHM and DT-DLHM with the highest achieved accuracy of 79.31%, 79.31%, 89.65%, 82.76% and 86.21%. The XGB-DLHM results also surpassed the base paper results with a margin of 3.45%. The model's resilience and adaptability make it more reliable and accurate, cementing its role as a valuable tool in MS detection. The model's results are evident as shown in Table 3. In essence, the progress of MS detection relies on continuous advancements in modelling, embracing new technologies, and forging collaborations to guarantee our findings' clinical relevance and impact. Through these unified efforts, our goal is to significantly contribute to developing accurate and reliable methods that aid in the early diagnosis and management of MS. The provided approach has the potential to be improved for more accurate results by using this approach with other various approaches. It can be implemented in different domains, including healthcare, finance, natural language processing, and others. The proposed algorithms also provided better results than other papers as well that include the results of de Leon-Sanchez's article [8] that yielded results of 86%, which our proposed models passed efficiently with an accuracy of 96.55% and 89.65% from FAD and XGB-DLHM, respectively.

Table 3: Performance comparison with base study [8]

Reference	Accuracy
[8]	86.20%
FAD	96.55%

6 Conclusion and Future Scope

This paper proposed a FAD approach to diagnose MS based on the data collected from the patients. Various machine-learning models and one deep-learning model were integrated with the hybrid deep-learning model approach, including RF, GB, XGB, KNN, DT and ANN. After evaluation of the dataset on these models, few of them performed better than the previous approaches, which were FAD, XGB-DLHM and DT-DLHM, with accuracies of 96.55%, 89.65% and 86.21%. The other models, including GB-DLHM, KNN-DLHM and RF-DLHM, provided an accuracy of 79.31%, 82.76%, and 79.31%, respectively. The results imply that our approach can provide more accurate results on this dataset. The disadvantage is that if a user applies combining ML models and LSTM neural networks, then the possible computation complexity will bring along universal training or inference actions. Thus, it will take a long time and many resources to finish. On the same front, pre-trained models or ensemble approaches also receive interpretability challenges. Hence, they need

in-depth re-training to make them redundant on newly inputted data and domains. These problems can only be solved by the implementation of optimization techniques, which are very complicated and require considerable effort. The approach can be further improved by using other ensemble methods and by implementing different sizes of the bidirectional LSTM network as well as their parameters according to the changed dataset. We provide an opportunity to perfect the knowledge gained in the part of consistently refining and expanding the application range of this innovative method in a world where a variety of situations are possible.

Acknowledgement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R503), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: This study was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R503), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Stephen Ojo, Moez Krichen, Meznah A. Alamro, Jaroslava Kniezova; data collection: Stephen Ojo; analysis and interpretation of results: Stephen Ojo, Moez Krichen, Meznah A. Alamro, Alaeddine Mihoub Author; draft manuscript preparation: Stephen Ojo, Moez Krichen, Meznah A. Alamro, Alaeddine Mihoub, Gabriel Avelino Sampedro, Jaroslava Kniezova. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] J. J. Cerqueira *et al.*, “Time matters in multiple sclerosis: Can early treatment and long-term follow-up ensure everyone benefits from the latest advances in multiple sclerosis?,” *J. Neuro., Neuro. Psy.*, vol. 89, pp. 844–850, 2018. doi: [10.1136/jnnp-2017-317509](https://doi.org/10.1136/jnnp-2017-317509).
- [2] H. Tahernia, F. Esnaasharieh, H. Amani, M. Milanifard, and F. Mirakhori, “Diagnosis and treatment of MS in patients suffering from various degrees of the disease with a clinical approach: The original article,” *J. Pharm. Negat. Results*, pp. 1908–1921, 2022.
- [3] E. Waubant *et al.*, “Environmental and genetic risk factors for MS: An integrated review,” *Anna. Clin. Trans. Neuro.*, vol. 6, no. 9, pp. 1905–1922, 2019.
- [4] B. Solomon, P. Boruta, D. Horvath, and J. MacKellar, “The multiple sclerosis stress equation,” *J. Med. Stat. Inf.*, vol. 11, p. 1, 2023. doi: [10.7243/2053-7662-11-1](https://doi.org/10.7243/2053-7662-11-1).
- [5] K. O. Mohammed Aarif, A. Alam, Pakruddin, and J. Riyazulla Rahman, “Exploring challenges and opportunities for the early detection of multiple sclerosis using deep learning,” *Arti. Intell. Auto. Dis.: Appl. Diagn., Prognosis, Ther.*, pp. 151–178, 2024.
- [6] C. S. Constantinescu, N. Farooqi, K. O’Brien, and B. Gran, “Experimental Autoimmune Encephalomyelitis (EAE) as a model for Multiple Sclerosis (MS),” *Br. J. Pharmacol.*, vol. 164, no. 4, pp. 1079–1106, 2011. doi: [10.1111/j.1476-5381.2011.01302.x](https://doi.org/10.1111/j.1476-5381.2011.01302.x).
- [7] C. M. Hersh, *Multiple Sclerosis, An Issue of Neurologic Clinics*. vol. 42. Philadelphia, Pennsylvania: Elsevier Health Sciences, 2023.

- [8] E. R. P. De Leon-Sanchez, A. M. Herrera-Navarro, J. Rodriguez-Resendiz, C. Paredes-Orta, and J. D. Mendiola-Santibañez, "A deep learning approach for predicting multiple sclerosis," *Micro*, vol. 14, no. 4, 2023, Art. no. 749.
- [9] N. Aslam *et al.*, "Multiple sclerosis diagnosis using machine learning and deep learning: Challenges and opportunities," *Sensors*, vol. 22, no. 20, 2022, Art. no. 7856. doi: [10.3390/s22207856](https://doi.org/10.3390/s22207856).
- [10] S. F. Maleki *et al.*, "Artificial Intelligence for multiple sclerosis management using retinal images: Pearl, peaks, and pitfalls," *Semi. Ophtha., Tay. Fra.*, vol. 39, no. 4, pp. 271–288, 2024.
- [11] A. Shoeibi *et al.*, "Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review," *Comp. Bio. Med.*, vol. 136, 2021, Art. no. 104697. doi: [10.1016/j.combiomed.2021.104697](https://doi.org/10.1016/j.combiomed.2021.104697).
- [12] T. S. Daqqaq, A. S. Alhasan, and H. A. Ghunaim, "Diagnostic effectiveness of deep learning-based MRI in predicting multiple sclerosis: A meta-analysis," *Neuro. J.*, vol. 29, no. 2, pp. 77–89, 2024. doi: [10.17712/nsj.2024.2.20230103](https://doi.org/10.17712/nsj.2024.2.20230103).
- [13] S. Denissen *et al.*, "Towards multimodal machine learning prediction of individual cognitive evolution in multiple sclerosis," *J. Person. Med.*, vol. 11, no. 12, 2021, Art. no. 1349. doi: [10.3390/jpm11121349](https://doi.org/10.3390/jpm11121349).
- [14] R. Liu, R. Yan, and Z. Peng, "A review of medical artificial intelligence," *Glo. Heal. J.*, vol. 4, no. 2, pp. 42–45, 2020. doi: [10.1016/j.glohj.2020.04.002](https://doi.org/10.1016/j.glohj.2020.04.002).
- [15] A. Gautam *et al.*, "Investigating gene expression profiles of whole blood and peripheral blood mononuclear cells using multiple collection and processing methods," *PLoS One*, vol. 14, no. 12, 2019, Art. no. e0225137. doi: [10.1371/journal.pone.0225137](https://doi.org/10.1371/journal.pone.0225137).
- [16] M. Navaderi *et al.*, "Identification of multiple sclerosis key genetic factors through multi-staged data mining," *Mult. Scle. Rel. Dis.*, vol. 39, 2020, Art. no. 101446. doi: [10.1016/j.msard.2019.101446](https://doi.org/10.1016/j.msard.2019.101446).
- [17] C. Chio and D. Freeman, *Machine Learning and Security: Protecting Systems with Data and Algorithms*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, Inc., 2018.