Iraqi Journal of Statistical Sciences

http://stats.uomosul.edu.iq

Diagnosis and Classification of Alzheimer's Disease Using Some Machine Learning Models: A Comparative Study

Saman Hussein Mahmood

Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Iraq.

Article information

Article history:
Received: October 16,2024
Revised: December 4,2024
Accepted :April 30,2025
Available online : December 1,2025

Keywords:
Support vector machines,
Decision trees,
feedforward neural networks
classification
Alzheimer's disease

Correspondence:
S.H.Mahmood
saman.mahmood@su.edu.krd

Abstract

Alzheimer's disease damages brain neurons, resulting in memory loss. Early and accurate diagnosis of the disease is crucial for implementing preventive measures. However, differentiating between Alzheimer's and healthy data in older adults is challenging due to the similarities in their brain patterns and intensities, complicating researchers' efforts to make an accurate diagnosis. Therefore, the research aims to use machine learning to improve diagnosis and classification of the disease, such as support vector machines (SVM), decision trees, and feedforward neural networks (FFNN). Classification algorithms were applied to the Alzheimer's disease dataset, including 2149 cases, and the models were evaluated through metrics (Accuracy, Precision, Recall, specificity, F1 Score, F2 Score, F3 score, and AUC).

Following data analysis and obtaining the results, we reached the Decision Tree model excels across all metrics, achieving high scores in accuracy (96.32%), precision (94.63%), recall (95%), specificity (97.05%), and AUC (94.96%). This demonstrates its ability to correctly identify true positives and negatives, and reduce false positives and negatives, makes it the most reliable model for accurately classifying Alzheimer's disease cases. In contrast, the SVM linear and FFANN models offer a good balance with accuracy (83.53% and 83.57%), specificity (89.27% and 91.72%), and AUC (89.63% and 89.84%). However, their lower recall (73.03% and 68.68%) compared to the Decision Tree may result in missed positive cases, making them less effective for classification. The SVM RBF model is the least effective option, with high precision and specificity but poor performance across all other metrics and lacks overall balance, resulting in a high number of false negatives. metrics and lacks overall balance, resulting in a high number of false negatives.

Conclusion: The decision tree model outperforms other models, making it the best choice for Alzheimer's disease classification.

DOI 10.33899/iqjoss.v22i2.54070.

1. Introduction

Alzheimer's disease is a neurological disorder that harms neurons and leads to cognitive problems, causing memory loss and other problems. It results from abnormal brain buildup and currently has no cure. While treatments are available to ease symptoms, scientists are working on ways to prevent or treat it. The disease is becoming increasingly common, affecting millions worldwide, and is expected to triple by 2050. This rise presents major challenges for patients, families, and the global economy, with ongoing research aimed at developing more effective solutions. (Breijyeh and Karaman, 2020)

Machine learning is revolutionizing various sectors, especially healthcare, by accurately predicting outcomes from large datasets. It can help diagnose diseases like Alzheimer's by analyzing medical data. However, diagnosing Alzheimer's is complex and often takes a long time. Many patients with Alzheimer's are not diagnosed correctly. This is because it requires a lot of information and can be difficult for doctors to do. Machine learning can help improve diagnosis accuracy and classifications, utilizing techniques such as support vector machines (SVM) decision trees, and feedforward neural networks (FFNN).

Supervised learning is a form of machine learning in which models are trained using labeled data to classify new items and predict their category in classification tasks. Different classification algorithms offer different advantages: SVMs excel at distinguishing between data, decision trees offer clear interpretability and adaptability, and FFNNs have adaptive learning capabilities. The unique strengths of each algorithm make them suitable for different classification challenges.

This study assesses the performance of three machine learning algorithms in classifying Alzheimer's disease data based on various patient features. By comparing the algorithms across several metrics, we aim to determine the most reliable method for enhancing the accuracy of Alzheimer's disease diagnosis in clinical settings.

2. Material and methods

2.1 Machine learning

Machine learning (ML), a branch of artificial intelligence, uses algorithms to identify patterns in data across a variety of domains, such as web searches, stock market forecasts, genetic analysis, and weather forecasting. Its role in healthcare is rapidly expanding, supporting diagnosis, treatment planning, and drug discovery. Machine learning excels at analyzing large data sets, identifying relationships, and adapting to new patterns. Unlike traditional optimization methods, machine learning improves models by learning from training errors to improve predictions, allowing computers to boost performance through experience without explicit programming. (Han, Kamber and Pei, 2012) (Choi et al., 2020)

In classification tasks, machine learning uses algorithms called classifiers, which analyze data characteristics (feature vectors) to categorize information. For example, in medical imaging, classifiers can distinguish between healthy and diseased tissue, trained on large-scale clinical datasets, and can apply the learned knowledge to new cases, dramatically improving diagnostic accuracy and enabling personalized medicine. By leveraging massive amounts of data, these models enhance early detection of diseases and enable timely interventions, revolutionizing healthcare delivery. (Alpaydın, 2010) (Ratner, 2017)

Depending on the learning methods, supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the four main categories under which machine learning techniques fall. One of the basic methods is supervised learning, which trains systems and makes predictions using labeled input data. It is further divided into classification and regression tasks. Classification algorithms, such as support vector machines (SVMs), decision trees, k-nearest neighbors (KNNs), and artificial neural networks (ANNs), each have advantages and are appropriate for certain classification issues, depending on the features of the data and the demands of the task. (Han, Kamber and Pei, 2012)

2.2 Support Vector Machines

Support Vector Machines (SVMs) are supervised learning models that can be applied to classification and regression problems. What sets SVMs apart is their focus on minimizing classification errors while maximizing the separation margin between data points, which is why they are often called Maximum Margin Classifiers. This margin refers to the distance between two parallel hyperplanes surrounding a central hyperplane, which acts as the boundary separating different classes. The larger this margin, the better the model typically performs on new, unseen data. (Saradha and Pavithra, 2024)

Support vector machines are grounded in the Structural Risk Minimization (SRM) principle derived from statistical learning theory. Unlike traditional approaches that aim only to reduce training error, SRM minimizes the upper limit of generalization error by considering both the training error and a complexity term associated with the model's VC dimension. By balancing these factors, SVMs often exhibit superior generalization, leading to better results on new data. To achieve the best separation of classes, SVMs map the input data into a higher-dimensional space, making it easier to distinguish between complex patterns. (Byun and Lee, 2002)

We represent data points in the form $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, where y_n is either -1 or 1, indicating the class of each point. There are many possible hyperplanes that can separate these training samples. (Srivastava and Bhambhu, 2010), (Ratner, 2017)

Let's assume they are fully separated by a d-dimensional hyperplane, defined by the equation.

$$w^T.x_i + b = 0 \tag{1}$$

Support vector machines work by creating a boundary that differentiates between two data classes. The primary goal of SVM is to determine the values of the weight vector w and the bias b such that the hyperplane is placed as far away from the nearest data points as possible. To achieve this, SVM defines two boundary planes with the following equations:

$$\left.\begin{array}{ll} w^T.x_i + b = 1 & for \;\; y_i = 1 \\ w^T.x_i + b = -1 & for \;\; y_i = -1 \end{array}\right\} \tag{2}$$

Here, x represents the input feature vector of a data point, w is the weight vector, b is the bias (a scalar value), and yi is the class label, which can be either +1 or −1-. As shown in Figure1.

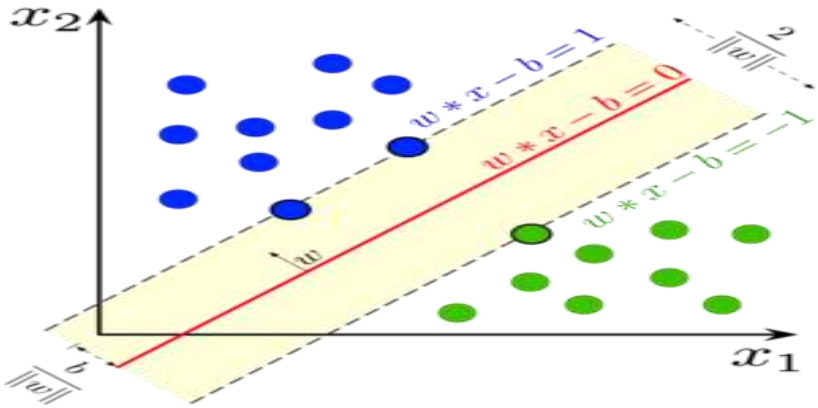


Figure 1, it is clear that the red level is the best, because it is the most robust against local disturbances when training the sample

The challenge in separation is to find the hyperplane that satisfies the conditions $w^T.x_i + b \geq 1$ for positive examples and $w^T.x_i + b \leq -1$ for negative examples. Since SVM aims to maximize the margin between these two classes, the hyperplane that achieves this can be found by minimizing $\frac{1}{2}\|w\|$.

$$\min_{w,b} \phi\,(w) = \frac{\|w\|^2}{2} \tag{3}$$

Consequently, the optimal separating hyperplane can be identified by minimizing equation (3) to effectively separate the training data as specified in equation (4).

$$y_i(w^T.x_i + b) \geq 1 \;\; i = 1,2,..I \tag{4}$$

Where *I* represent the number of training data points
To reduce the complexity of this problem, we introduce the Lagrange function, resulting in:

This optimization problem solved by the saddle points of the Lagrange's Function

$$L_p = L_{(w,b,\alpha)} = \frac{1}{2}\|w\|^2 \sum_{i=1}^{I} \alpha_i(y_i(w^T.x_i + b) - 1)$$

$$= \frac{1}{2} w^T w \sum_{i=1}^{I} \alpha_i(y_i(w^T.x_i + b) - 1) \quad (5)$$

Where αi represents a Lagrange multiplier. By taking the partial derivatives of w and b in Equation (5), setting them equal to zero, and substituting the results back into Equation (5), we obtain the dual problem.

$$L_d(\alpha) = \sum_{i=1}^{I} \alpha_i - {}^1\!/_2 \sum_{i=1}^{I}\sum_{j=1}^{I} \alpha_i\alpha_j y_i y_j x_i^T x_j \quad (6)$$

The dual Lagrangian (Ld) must be maximized concerning nonnegative αi>0 To find the optimal hyperplane. It is important to note that the dual Lagrangian Ld(α) is formulated based on the training data and depends solely on the scalar products of input patterns $(x_i x_j)$.

Moreover, $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is referred to as the kernel function, there are several kernel functions that are commonly used in SVM like (**Linear Kernel** , Polynomial Kernel, Gaussian (RBF) Kernel and Sigmoid Kernel). (Kar et al., 2024) (Byun and Lee, 2002)

2.3 Decision Tree

Decision trees are powerful supervised learning algorithms used for both classification and prediction tasks. They are regarded as one of the leading machine learning techniques for decision analysis and the classification of unknown cases. By organizing data into a hierarchical structure of decisions and outcomes, decision trees offer strong predictive power while remaining easy to interpret, making them ideal for a range of data-driven applications. Decision trees are especially popular due to their straightforward interpretation and efficiency when processing large datasets, making them faster and more effective than many other methods. (James et al., 2013) (Rokach and Maimon, 2015)
The key elements of a decision tree that contribute to its effectiveness in analysis are: (Tan et al., 2019)
Root Node: The tree begins with a root node that represents the main question or problem to be addressed.
Branches: These branches extend from the root nodes, and symbolize the different options and actions available at each decision point.
Nodes: These are decision points that lead to branches in two or more directions, based on the choices made among different options or considerations.
Leaves (Leaf Nodes): Indicates the outcome or decision and shows the outcome of the path taken through the tree.

Decision Tree Splitting

There are various measures that can be used to determine the most effective way to split the data points. These measures are defined based on the class distribution of the data both before and after the split. In a decision tree, each observation is assigned to the most frequent class within its group. The classification error rate is calculated as the proportion of training observations in that group that do not belong to the most common class, expressed as (James et al., 2013)

$$E = 1 - \max_k(\hat{p}_{mk}) \quad (7)$$

In this context, $\hat{\boldsymbol{p}}_{\boldsymbol{mk}}$ denotes the proportion of training observations in the mth group that belong to the kth class.

However, classification error alone is not sensitive enough for effective tree growth, leading to the preference for two alternative measures. One such measure is the Gini index, which assesses total variance across the K classes. It yields a low value when the $\hat{\boldsymbol{p}}_{\boldsymbol{mk}}$ values are near zero or one, indicating node purity. A small Gini index suggests that a node predominantly contains observations from a single class. (Hastie, Tibshirani and Friedman, 2013)

However, it has been found that classification error alone is not sensitive enough for effective tree growth, and in practice, two alternative measures are often preferred.

The Gini index is defined by

$$\text{Gini Index} = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \quad (8)$$

Another measure is cross-entropy or deviance, defined as:

$$\text{Entropy} = -\sum_{k=1}^{K} \hat{p}_{mk} log\, \hat{p}_{mk} \quad (9)$$

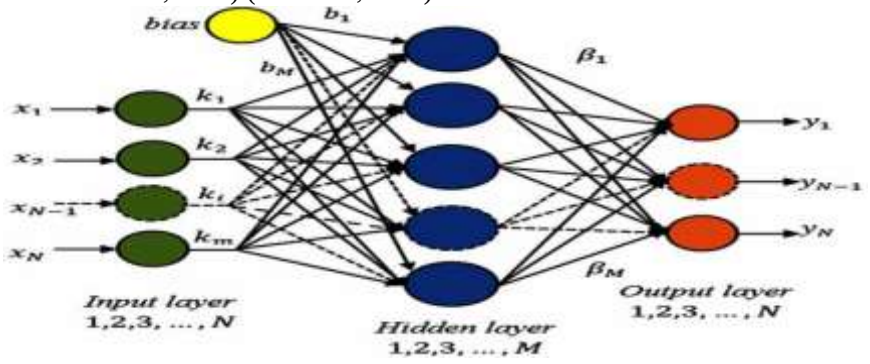
Given that $0 \leq \hat{\boldsymbol{p}}_{\boldsymbol{mk}} \leq \mathbf{1}$, it follows that $0 \leq -\hat{\boldsymbol{p}}_{\boldsymbol{mk}} \boldsymbol{log} \hat{\boldsymbol{p}}_{\boldsymbol{mk}}$. Cross-entropy approaches zero when the $\hat{p}_{mk}$ values are close to zero or one, leading both cross-entropy and the Gini index to produce small values for pure nodes.

When building decision trees, these measures are favored for assessing split quality due to their sensitivity to node purity, while classification error is preferred for improving final prediction accuracy during pruning. (Han, Kamber and Pei, 2012)

2.4 Artificial Neural Networks

Artificial neural networks are computer models inspired by the neural architecture of the brain, and learn from experience to solve complex problems more efficiently than traditional computers. Artificial neural networks offer a flexible approach to computing through parallel networks and pattern recognition, focusing on learning, self-organization, and problem solving without relying on traditional programming methods. (del-Pozo-Bueno et al., 2023) Artificial neural networks come in many different forms, each designed for different tasks. This study focuses on the use of feedforward neural networks for data classification.

Data in a feedforward neural network flows in one direction, starting from the input layer, which represents inputs from different attributes. Numeric or binary attributes are typically denoted by a single node. These inputs are routed to intermediate layers known as hidden layers, which consist of processing units called hidden nodes. Each hidden node processes signals from the input or previous hidden nodes, generating activation values that are sent to the next layer. The final layer, known as the output layer, processes these activation values to produce predictions for the output variables, with a single node representing the binary class label in binary classification tasks. This architecture is called feedforward neural networks because of the forward propagation of signals Figure(2). The network is fully connected, allowing each unit to contribute to the computations of the subsequent layer, enabling efficient modeling of complex functions. (Han, Kamber and Pei, 2012) (Tan et al., 2019)



Finding the optimal architecture often involves trial and error and experimentation, commonly starting with a simple design and gradually increasing complexity until performance levels off.

Common activation functions include the sigmoid, the rectified linear unit (ReLU), the hyperbolic tangent, the exponential linear unit, the leaky ReLU, and the gradient exponential linear unit. The specific task will determine the type of activation function used. Loss functions assess the discrepancy between the predicted and actual outputs, and guide the training process to minimize this discrepancy by adjusting the connection weights using optimization algorithms. (Bouraya and Belangour, 2024)

Backpropagation is a common method that computes the gradient of the loss function with respect to the update weights. Different algorithms can be used, such as stochastic gradient descent (SGD) and adaptive moment estimation (ADAM). To learn the weights of a feedforward neural network (FFNN), an efficient algorithm is necessary to converge to the correct solution with sufficient training data. One approach is to treat each hidden node or output node as an independent realization and apply a weight update formula. However, this approach is limited by the lack of

prior knowledge of the true outputs of the hidden nodes, which complicates the determination of the error term$(y - \hat{y})$. (Han, Kamber and Pei, 2012) (Jalil and Mahmood, 2012)

Given our focus on FFNNs in this work, we can represent their operation mathematically as follows:

$$Output = \text{activation}(W * \text{input} + b) \tag{10}$$

In this equation, W is the weight matrix that links neurons in the current layer to those in the previous layer. The term "input" denotes the vector from the preceding layer, b is the bias vector for each neuron in the dense layer, and "activation" refers to the applied activation function.
The learning algorithm of the FFNN aims to find a set of w weights that minimizes the total sum of squared errors:

$$E(w) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{11}$$

Here, the sum of squared errors is conditioned by www, where $\hat{y}$ is affected by the weights assigned to the hidden nodes and the output nodes. The outputs of FFNNs are typically a nonlinear function of their coefficients, which complicates the search for a globally optimal solution to w. To address this optimization problem, gradient descent algorithms are used, using the weight update formula: (Tan et al., 2019)

$$w_j = w_j - \lambda\frac{E(w)}{\partial w_j} \tag{12}$$

Where λ represents the learning rate, which guides weight adjustments to minimize the overall error, however, the non-linearity of the error function means that the gradient descent method can become trapped in local minima. (Mahmood and Haji Khider, 2023)

2.5 Performance Metrics

Comparing various algorithms is crucial in this study. Model evaluation helps assess the effectiveness and dependability of machine learning-based predictive models through systematic performance analysis.

There are fundamental components of a confusion matrix (table 1) necessary for calculate the performance metrics:

Table1 : Confusion Matrix

Actual Value	**Predicted Value**	
	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{13}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

Accuracy measures the percentage of correct predictions a classifier makes on a given dataset and Precision is defined as the ratio of true positive forecasts to the proportion of true negative forecasts that are incorrectly classified as positive (FP).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{15}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{16}$$

The recall includes all correctly identified positive cases, even those that were mistakenly classified as negative. Specificity measures the proportion of true negatives that are correctly identified by the classifier.

$$F_\beta = \frac{(1+\beta^2) * \text{Precision} * \text{Recall}}{(\beta^2 * \text{Precision}) + \text{Recall}} \quad (17)$$

The scores of F1, F2, and F3 are calculated by setting β to 1, 2, and 3 and taking a positive real number.
Precision and recall are vital metrics for evaluating performance in medical research focused on classification and diagnosis. While the F-measure represents the average of precision and recall, combining them into a single, easier-to-comparison metric. Since false negatives can be more detrimental than false positives, recall is typically regarded as the most important metric in the medical field. Specificity indicates the model's ability to accurately identify individuals who do not have the disease. (Tan et al., 2019) (Assiri, Nazir and Velastin, 2020)
AUC: The area under the curve (AUC) measures a classifier's performance, ranging from 0.50 (random guessing) to 1.00 (perfect accuracy). A higher AUC signifies a better model, allowing comparison across different classifiers by considering all possible decision thresholds. (Tan et al., 2019)

3. Results

In this paper we used a dataset related to Alzheimer's disease that was taken from the publicly available Kaggle dataset (downloadable from [19]), which includes 2149 cases. The target or dependent variable is binary, with 1 indicating the presence of Alzheimer's disease (760 cases) and a value of 0 indicating its absence (1,389 cases). The independent variables consist of 32 attributes, including:

Demographic Details [Age(x1), Gender(x2), Ethnicity(x3), Education Level(x4)]

Lifestyle Factors [BMI (x5), Smoking (x6), Alcohol Consumption (x7), Physical Activity (x8), Diet Quality (x9), Sleep Quality (10)].

Medical History [Family History] Alzheimer's (x11), Cardiovascular Disease (x12), Diabetes (x13), Depression(x14),Head Injury(x15), Depression (x14), Head Injury (x15), Hypertension (16)].

Clinical Measurements [Systolic BP (x17), Diastolic BP (x18), Cholesterol Total (x19), Cholesterol LDL (x20), Cholesterol HDL (x21), Cholesterol Triglycerides (x22)].

Cognitive and Functional Assessments [MMSE (Mini-Mental State Examination score) (x23), Functional Assessment (x24), Memory Complaints (x25), Behavioral Problems (x26), ADL (x27)].

When analyzing data with three algorithms, we split the dataset into two segments: 70% for training and validation, and 30% for testing the model's performance. In the analysis, we used Matlab(R2022b) to analyze.

3.1 Support Vector Machine Analysis

In our analysis and classification of Alzheimer's disease using SVM, we started by splitting the dataset into 70% for training and 30% for testing. Since the target of this dataset is to classify it into two groups (Alzheimer's disease and non-Alzheimer's disease), we chose binary SVM. We applied and compared different kernel functions, including linear kernel and radial basis function (RBF), to determine the most suitable model for this type of data. After training the models, we evaluated their accuracy using relevant metrics, as shown in Table (2).
We used binary SVM to classify Alzheimer's disease data into two groups (Alzheimer's disease and non-Alzheimer's disease). We applied two different kernel functions (linear and RBF) and compared their performance. We then evaluated the accuracy of the models using various metrics Tables 2 and 3.

Table 2 : Confusion Matrix for Alzheimer's Disease Classification

A: **SVM Linear Model**

Actual	Predicted Value		Total
	Disease	**No Disease**	
Disease	555	205	760
No Disease	149	1240	1389

B: SVM RBF Model

Actual	Predicted Value		Total
	Disease	**No Disease**	
Disease	270	490	760
No Disease	36	1353	1389

Table 2 can be used to calculate a number of performance metrics, including:

Table 3: Comparison of the percentage results between the SVM linear and the SVM RBF models.

Metric	SVM linear	SVM RBF	Explanation
Accuracy	83.53	75.52	SVM Linear is more accurate, correctly classifying more instances. This indicates that it is better at distinguishing between Alzheimer's patients.
Precision	78.84	88.24	Higher precision indicates that the SVM RBF model has a higher chance of being accurate for predicting Alzheimer's disease. It appears that the RBF model produces fewer false positives due to its increased precision.
Recall	73.03	35.53	The linear SVM model has a much higher recall, meaning that it identifies a higher proportion of true positive cases (Alzheimer's disease). This is essential for medical diagnostics.
Specificity			The RBF SVM model is significantly more specific (97.4%) than the Linear SVM model (89.27%), demonstrating its superior ability to accurately identify healthy individuals
F1 Score	75.82	50.66	SVM Linear has a better balance between precision and recall, finding true Alzheimer's cases while avoiding misdiagnosis.
F2 Score	74.12	40.35	Recall is given greater weight than precision, particularly the F3 score, and the results further support the idea that the linear SVM model is more effective at identifying true positives.
F3 Score	73.57	37.73	
AUC	89.63	59.86	SVM Linear has a much higher AUC, indicating superior ability to differentiate between classes.

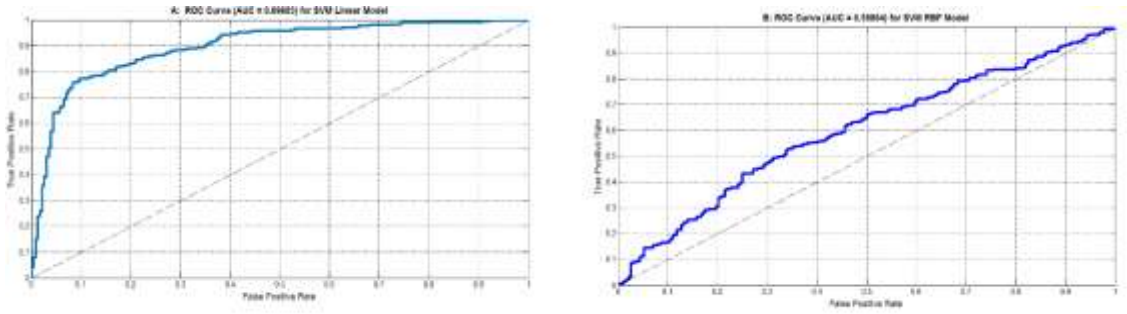


Figure 3 (A and B) :ROC Curve (AUC) for SVM linear model and SVM RBF model.

Based on Table 3 and Figure 3, we can conclude that SVM Linear is the preferred model for classifying Alzheimer's disease because of its higher accuracy, precision, F1 score, and AUC. It offers a more dependable diagnosis by reducing false positives while still accurately identifying Alzheimer's patients. In contrast, while SVM RBF performs well in terms of recall, it is overall less reliable due to its lower precision and AUC

3.2 Decision Tree Analysis

We used decision trees to analyze Alzheimer's disease data. We measured how accurate the models were using different metrics like (accuracy, precision, recall ...) table4. We used the Gini Index and Entropy to decide how to split the data into branches. The best split is the one that makes the branches the purest.

Table 4: Confusion Matrix for Decision Tree Model

Actual	Predicted Value		Total
	Disease	**No Disease**	
Disease	722	38	760
No Disease	41	1348	1389

Table 4 can be used to calculate a number of performance metrics, including:

Table 5: Percentage results of the Decision Tree model

Metric	Accuracy	Precision	Recall	Specificity	F1 Score	F2 Score	F3 Score	AUC
Decision Tree	96.32	94.63	95.0	97.05	94.81	94.93	94.96	93.06

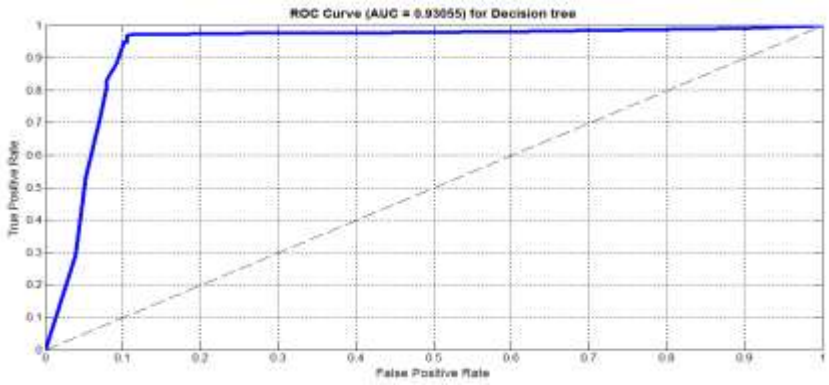


Figure 4: ROC Curve (AUC) for Decision Tree model

The resulting decision tree model accurately classified the majority of cases (96.32%) as shown in Table 5 and Figure 4, and this is confirmed by the results for high precision (94.63%) and recall (95.00%), meaning it is good at finding people who actually have the disease. It also has high precision specificity (97.05%), meaning that it is good at not mistakenly identifying people who do not have the disease. The F-score and AUC (93.06%) are also very good, indicating that the model is well suited for this task. Overall, the decision tree model is a reliable choice for classifying people with and without Alzheimer's disease.

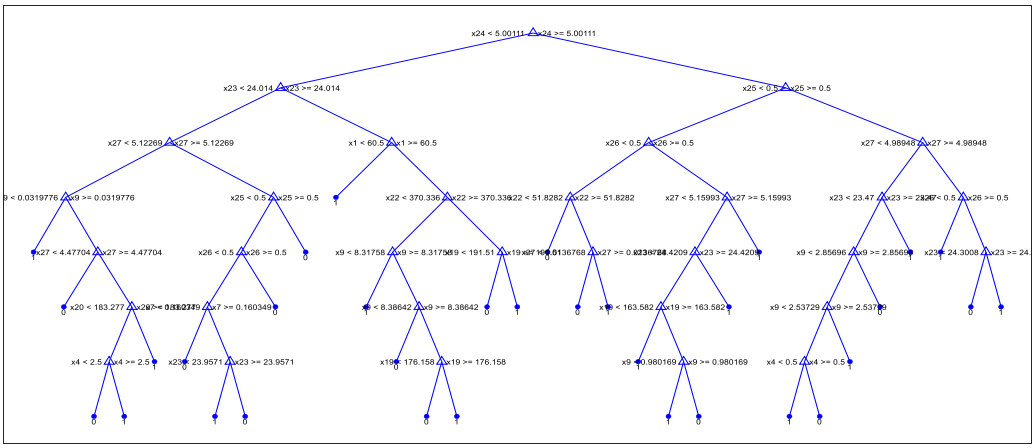


Figure 5: Classification Decision Tree display

The decision tree in Figure 5 shows which factors are most important in predicting Alzheimer's disease. The factors at the top of the tree are the most significant predictors. These factors are likely to be strong indicators of the disease. In this specific decision tree, cognitive assessments and age seem to be particularly important. The tree also considers secondary factors, such as lifestyle and medical history. By tracing the decision paths from the top to the bottom of the tree, we can see how the model arrives at its final classifications for each individual.

3.3 Artificial Neural Networks Analysis: -

This study employs multilayer feedforward neural networks to detect and classify Alzheimer's disease. The multilayer perceptron (MLP) neural networks are commonly trained using the backpropagation (BP) algorithm. The proposed multi-layer backpropagation neural network classifier consists of three layers: input, hidden, and output. The output layer's nodes correspond to the number of classes (dependent variables), while the number of hidden layer nodes is determined through trial and error. Each connection between neurons has an associated weight, which is adjusted during training based on the input and output data. The BP algorithm minimizes network errors using gradient methods or other numerical optimization techniques.

Table 6: Confusion Matrix for FFNN Model

Actual	Predicted Value		Total
	Disease	**No Disease**	
Disease	522	238	760
No Disease	115	1274	1389

Table 6 can be used to calculate a number of performance metrics, including:

Table 7 : Percentage results of the FFNN Model

Metric	Accuracy	Precision	Recall	Specificity	F1 Score	F2 Score	F3 Score	AUC
FFNN	83.57	81.89	68.68	91.72	74.73	70.89	69.81	89.84

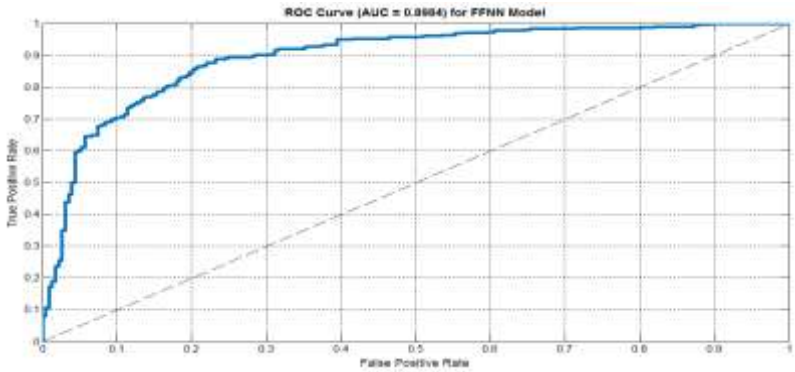


Figure 6: ROC Curve (AUC) for FFNN model

The results presented in Table 7 demonstrate the performance metrics of the FFNN model in classifying Alzheimer's disease. The FFNN model achieves good accuracy (83.57%) and commendable specificity (91.72%), as well as showing good precision (81.89%). However, its recall value of 68.68% indicates that the model has some limitations in identifying all actual positive cases of Alzheimer's disease. The F1 score reflects reasonable performance, while the F2 and F3 scores suggest adequate classification across different thresholds. With a high AUC in Figure6, the model shows strong overall performance, indicating its potential effectiveness for detecting and classifying Alzheimer's disease. This model can be considered a valuable tool in medical diagnostics, where accurately identifying cases is of utmost importance.

The results in Table 7 illustrate the performance metrics of the FFNN model in classifying Alzheimer's disease. The FFNN model achieves good accuracy and excellent specificity, as well as showing commendable precision.

4. Discussion

We evaluated the effectiveness of SVM, Decision Tree, and FFNN based on the metrics presented in Table 8 and Figure 7.

Table 8: Percentage results of all models.

Metric	Accuracy	Precision	Recall	Specificity	F1 Score	F2 Score	F3 Score	AUC
SVM linear	83.53	78.84	73.03	89.27	75.82	74.12	73.57	89.63
SVM RBF	75.52	88.24	35.53	97.41	50.66	40.35	37.73	59.86
Decision Tree	96.32	94.63	95.0	97.05	94.81	94.93	94.96	93.06
FFNN	83.57	81.89	68.68	91.72	74.73	70.89	69.81	89.84

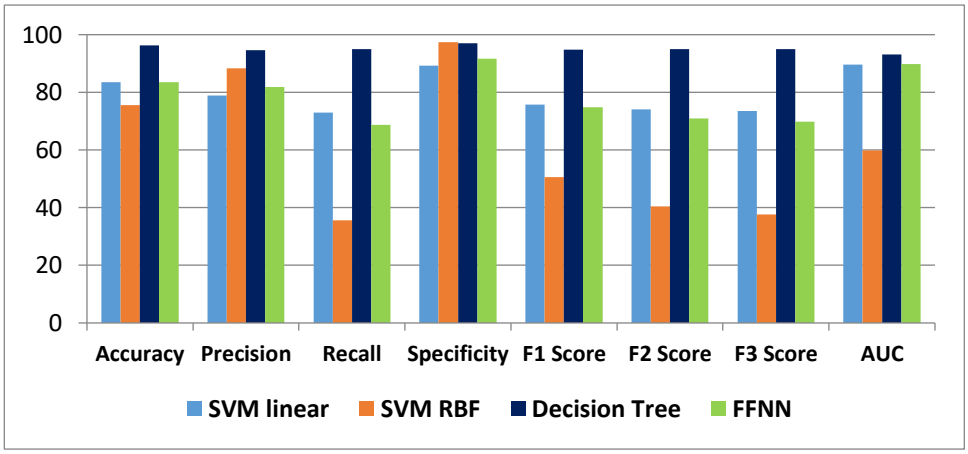


Figure 7: Compare results for all models

Table 8 and Figure 7 show the performance metrics of the various models used to classify Alzheimer's disease. The decision tree model appears to perform best, achieving high accuracy, precision, recall, and specificity. This indicates its exceptional ability to correctly identify true positives and true negatives, and reduce false positives and false negatives. While the SVM RBF model shows high precision and specificity, its low Accuracy, recall and AUC limit its practical application in medical diagnosis, where detecting all disease cases is critical. Although the SVM Linear and FFNN models provide moderate accuracy and precision, they lack recall compared to the decision tree. These models can be considered secondary options, especially in scenarios where low sensitivity is acceptable.

5. Conclusion

Alzheimer's is an incurable brain disease, and early detection helps families plan for the future. Using a number of machine learning models, we concluded that the decision tree model performs best across all metrics, making it the most reliable model for classifying Alzheimer's disease. Both the SVM Linear and FFANN models show good and balanced performance in terms of accuracy, precision, and recall, although they are lower than the decision tree results. The SVM RBF model, despite its strong accuracy, is weak in recall and lacks overall balance, which reduces its effectiveness in diagnosis.

6. Recommendations:

In the future, we could test this model on different Alzheimer's disease datasets. We could also try new ways machine learning to combine models to classify Alzheimer's disease. We could also use this model with other medical information, like brain scans and tests, to make it even better.

References

1. Alpaydın,E. (2010) Introduction to machine learning, 2nd edition, Cambridge, Massachusetts,London, England: The MIT Press.
2. Alshebly, O. Q., & Ahmed, R. M. (2019). Prediction and Factors Affecting of Chronic Kidney Disease Diagnosis using Artificial Neural Networks Model and Logistic Regression Model. Iraqi Journal of Statistical Science, 16(1), 140-159. https://doi.org/10.33899/iqjoss.2019.0164186
3. Altalabani, D., & Erdogan, F. (2024). Classification of Diabetes Data Set from Iraq via Different Machine Learning Techniques.Iraqi Journal of Statistical Sciences 21(1), 168-186. https://doi.org/10.33899/iqjoss.2024.183258
4. Assiri, A.S., Nazir, and Velastin, S.A. (2020) 'Breast Tumor Classification Using an Ensemble Machine Learning Method', journal of Imaging, vol. 6, no. 39, pp. 1-13. https://www.mdpi.com/2313-433X/6/6/39
5. Bouraya, and Belangour, (2024) 'A comparative analysis of activation functions in neural networks: unveiling categories', Bulletin of Electrical Engineering and Informatics, vol. 13, no. 5, pp. 3301-3308. https://beei.org/index.php/EEI/article/view/7274
6. Breijyeh,Z. and Karaman, R. (2020) 'Comprehensive Review on Alzheimer's disease: Causes and Treatment', Molecules, vol. 25, no. 4, p. 5789. https://www.mdpi.com/1420-3049/25/24/5789
7. Byun, H. and Lee, S.-W. (2002) 'Applications of support vector machines for pattern recognition: A survey', in International workshop on support vector machines, Springer Berlin Heidelberg.
8. Choi, R.Y., Coyner, A.S., Kalpathy-Cramer, J., Chiang, M.F. and Campbell, J.P. (2020) 'Introduction to Machine Learning, Neural Networks and Deep Learning', Translational vision science & technology, vol. 9, no. 2, pp. 1-12. https://doi.org/10.1016/j.ultramic.2023.113828
9. del-Pozo-Bueno, D., Kepaptsoglou, D., Peiro, F. and Estrade, S. (2023) 'Comparative of machine learning classification strategies for electron energy loss spectroscopy: Support vector machines and artificial neural networks', Ultramicroscopy, vol. 253, p. 113828. https://doi.org/10.1016/j.ultramic.2023.113828
10. Hamad, K. F., Celik, B., & Ahmed, R. M. (2024). Classification of Circular Mass of Breast Cancer Using Artificial Neural Network vs. Discriminant Analysis in Medical Image Processing. Iraqi Journal of Statistical Sciences, 21(1), 46-58. https://doi.org/10.33899/iqjoss.2024.183231
11. Han,J. , Kamber, M. and Pei, J. (2012) Data Mining:Concepts and Techniques, 3rd edition, Waltham, USA: Elsevier Inc.
12. Hastie,T., Tibshirani,R. and Friedman,J. (2013) The Elements of Statistical Learning :Data Mining, Inference and prediction, 2nd edition, New York: Springer.
13. James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) An Introduction to Statistical Learning: with Applications in R, New York: Springer Science and Business Media.
14. Kar, A., Nath, N., Kemprai, U., Aman (2024) 'Performance Analysis of Support Vector Machine (SVM) on Challenging Datasets for Forest Fire Detection', International Journal of Communications, Network and System Sciences, vol. 17, pp. 11-29. https://www.scirp.org/journal/paperinformation?paperid=131446
15. Obi, J. CH. (2023) A Comparative Study of Several Classification Metrics and Their Performances on Data, World Journal of Advanced Engineering Technology and Sciences, vol. 8 No. 1,pp. 308–314. https://doi.org/10.30574/wjaets.2023.8.1.0054
16. Ratner,B. (2017) Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data, 3rd edition, New York: Chapman and Hall/CRC.
17. Rokach, L. and Maimon, O. (2015) Data Mining with Decision Trees: Theory and Applications., 2nd edition, Singapore: World Scientific Publishing Co.
18. Pavithra, C. and Saradha, M. (2024) 'Classification And Analysis Of Clustered Non-Linear Separable Data Set Using Support Vector Machines', Migration Letters, vol. 21, no. 54, pp. 901-913.https://migrationletters.com/index.php/ml/article/view/7365

19. Srivastava, D.K. and Bhambhu, L. (2010) 'Data Classification Using Support Vector Machine', Journal of Theoretical and Applied Information Technology, vol. 12, no. 1, pp. 1-7. https://www.jatit.org/volumes/Vol12No1/1Vol12No1.pdf
20. Tan, P. N., Steinbach, M., Kumar, V., & Karpatne, A. (2019) Introduction to Data Mining, 2nd edition, New York: Pearson Education Limited.
21. https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset, [Online].

تشخيص وتصنيف مرض الزهايمر باستخدام بعض نماذج التعلم الآلي: دراسة مقارنة

سامان حسين محمود

قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة صلاح الدين–أربيل، العراق

الخلاصة:يتسبب مرض الزهايمر في تلف الخلايا العصبية في الدماغ، مما يؤدي إلى فقدان الذاكرة. يعد التشخيص المبكر والدقيق للمرض أمرًا بالغ الأهمية لتنفيذ التدابير الوقائية. ومع ذلك، فإن التمييز بين بيانات مرضى الزهايمر والبيانات الصحية لدى كبار السن أمر صعب بسبب أوجه التشابه في أنماط الدماغ وكثافتها، مما يعقد جهود الباحثين للتوصل إلى تشخيص دقيق. لذلك، يهدف البحث إلى استخدام التعلم الآلي لتحسين تشخيص وتصنيف المرض، مثل آلات الدعم المتجهة (**SVM**) وأشجار القرار والشبكات العصبية المغذية (**FFNN**). تم تطبيق خوارزميات التصنيف على مجموعة بيانات مرض الزهايمر، بما في ذلك 2149 حالة، وتم تقييم النماذج من خلال المقاييس (الدقة والإحكام والتذكر والخصوصية ودرجة **F1** ودرجة **F2** ودرجة **F3** و**AUC**).وبعد تحليل البيانات والحصول على النتائج، توصلنا إلى أن نموذج شجرة القرار يتفوق في جميع المقاييس، محققًا درجات عالية في الدقة (96.32٪)، و**Precision** (94.63٪)، والتذكر (95٪)، والخصوصية (97.05٪)، و**AUC** (**94.96%**). إن قدرته على تحديد الإيجابيات والسلبيات الحقيقية بشكل صحيح، وتقليل الإيجابيات والسلبيات الكاذبة، يجعله النموذج الأكثر موثوقية لتصنيف حالات مرض الزهايمر بدقة. في المقابل، يوفر نموذجا **SVM** الخطي و**FFANN** توازنًا جيدًا مع الدقة (83.53٪ و83.57٪)، والخصوصية (89.27٪ و91.72٪)، و**AUC**(**89.63**٪ و89.84٪). ومع ذلك، فإن تذكرهم المنخفض (73.03٪ و68.68٪) مقارنة بشجرة القرار قد يؤدي إلى حالات إيجابية مفقودة، مما يجعلها أقل فعالية للتصنيف. نموذج **SVM RBF** هو الخيار الأقل فعالية، مع دقة عالية وخصوصية ولكن أداء ضعيف عبر جميع المقاييس الأخرى ويفتقر إلى التوازن العام، مما يؤدي إلى عدد كبير من النتائج السلبية الخاطئة.الاستنتاج: يتفوق نموذج شجرة القرار على النماذج الأخرى، مما يجعله الخيار الأفضل لتصنيف مرض الزهايمر.

الكلمات المفتاحية: آلات الدعم المتجهة، أشجار القرار، الشبكات العصبية الأمامية، التصنيف، مرض الزهايمر.