4. Results & Discussion

4.1. Introduction

This chapter presents the outcomes of the research study, which focused on predicting human performance in upcoming cognitive tasks using EEG data. A comprehensive overview of the data collection process, the analysis conducted on the EEG data and the development of a machine learning model for predicting human performance has also been presented in this besides highlighting the key findings and discussing their implications.

The research study involved the collection of EEG data from participants prior to their engagement in various cognitive tasks. The EEG data were recorded to capture neural activity associated with cognitive processes, such as attention and working memory. This data served as the basis for predicting human performance in upcoming cognitive tasks.

Following the description of the methodology, the chapter presents the results of the analysis and prediction models. It discusses the performance of the machine learning model in predicting human performance in upcoming cognitive tasks based on the collected EEG data. The chapter incorporates relevant tables, figures and analyses data to support and illustrate the findings.

The results obtained from this study hold significant implications as they shed light on the potential of EEG data for predicting human performance in cognitive tasks. By understanding the relationship between neural activity captured by EEG and subsequent task performance, insights can be gained into cognitive processes and factors influencing human performance. The chapter concludes by discussing the implications of the results, their limitations and suggesting directions for future research to further enhance the accuracy and applicability of human performance prediction using EEG data.

4.2. Captured Data:

The data collection process for this study involved conducting two sessions: session A and session B. Session A consisted of a 5-minute idle period, while session B involved a 5-minute test of cognitive ability. The EEG data were recorded during these sessions from a total of 10 students, each with an average age of 27 ± 5 years. The recording process was conducted twice a day for 15 days. During session A, the EEG signals were recorded using three electrodes (Fp1, Fpz and Fp2) with a sampling rate of 250 Hz. A sample of this record is given in Appendix 1 and the summary is given in Table 4.1. This session served as a baseline measurement of neural activity during an idle state.

The selection of 10 participants for this study was driven by practical considerations, specifically the immediate availability of regular male students for recording sessions. Due to resource and time constraints, a smaller sample size was chosen, focusing on a specific group to explore particular characteristics. While the participants were not randomly drawn from a broader target population, the study hones in on insights specific to male participants. It is crucial to acknowledge that the findings may have limited generalisability beyond this particular demographic.

In session B, the participants underwent a cognitive ability test using Lumosity software. The test specifically focused on attention and working memory. The participants' performance in this test was recorded, including the number of failures, successful attempts made and the time taken to complete the test. The whole recordings are provided in Appendix 2.

As a result, the collected data consisted of ten variables for each participant. These variables included subject ID, recordings from the three EEG electrodes (fp1, fp2 and fp3), the number of failures and successful attempts in the attention test along with the corresponding time taken to complete the test. Similarly, the data also included the number of failures and successful attempts in the working memory test along with the time taken to complete it.

Table 4.1. Data Collected

Subject ID	1	2	3	4	5	6	7	8	9	10	Total
No. of recordings	15	18	28	30	30	29	29	23	20	14	236

This dataset provides a comprehensive set of measurements capturing both the EEG data and cognitive performance metrics of the participants. The combination of EEG recordings and cognitive test results offers an opportunity to investigate the relationship between neural activity and cognitive ability, paving the way for the development of a predictive model for human performance based on EEG data.

4.3. Feature extraction of recorded EEG data:

The raw EEG data was processed using the Brainstorm software to extract relevant features for further analysis. Average power values of different frequency bands, including delta, theta, alpha, beta and gamma, were computed from the EEG data. Additionally, the power values were averaged across different electrodes to obtain a comprehensive representation of neural activity.

In addition to the power values of frequency bands, blink-related features were extracted using the Blink software. The software provided the starting, total closure and ending time stamps of each blink during the session A. From these time stamps, the blink duration and blink intervals were derived. Finally, from these values, the average blink duration and average blink interval across the entire session A were calculated.

Furthermore, cognitive performance metrics were computed from the session B data. The number of failures and successful attempts in the attention and memory tests were used to calculate the success rate for each test. Also, the average response time for each attempt in both the attention and memory tests was calculated from session B recordings.

As a result of the feature extraction process, a total of 12 features as shown in Table 4.2 were obtained for each participant. The entire data is presented in Appendix 7.

Table 4.2. Features obtained for each participant

Features from Session A	**Features from Session B: Attention test**
Average power in delta frequency band	Success Rate
Average power in theta frequency band	Average response time for each attempt
Average power in alpha frequency band	
Average power in beta frequency band	**Features from Session B: Memory test**
Average power in gamma frequency band	Success rate
Average blink duration	Average response time for each attempt
Average blink interval	Subject ID

By extracting these features, the raw EEG data and cognitive performance metrics were transformed into meaningful numerical representations. These features serve as the input for further analysis and the development of a predictive model to assess human performance based on the collected data.

4.4. Data preparation apropos to machine learning models:

In order to develop a machine learning model based on the collected data, several steps were taken to prepare the data for analysis. The input variables considered for the model include the subject ID, the average power values in the delta, theta, alpha, beta and gamma frequency bands, the average blink duration, the average blink interval and the average response time. The output variable is the success rate, which ranges from 0 to 1.

To facilitate the analysis, the success rate values were classified into two groups. One group contains all the values equal to 1, representing successful performance, while the other group contains the remaining values.

In order to handle outliers in the data, a clipping technique based on quartiles was applied. This approach replaces extreme values with the corresponding quartile values, effectively mitigating the impact of outliers on the analysis.

Following the outlier treatment, the data was normalised. Normalisation is a standardisation process that scales the values of different variables between 0 and 1, ensuring that variables with larger numerical ranges do not dominate the analysis besides allowing for fair comparison and interpretation of their impact on the model's predictions.

By completing these data preparation steps, the collected data was transformed into a suitable format for the development of a machine learning model. The prepared data serves as the foundation for further analysis and training of the predictive model, enabling to predict human performance based on the provided input variables.

4.5. Development and evaluation of predictive models:

In the process of developing machine learning models, it is crucial not only to train machine learning models using available data but also to evaluate their performance to assess their suitability for real-world deployment. Evaluation helps understand whether the model requires further improvement or it can effectively handle classification tasks. To gauge the performance of these models, various evaluation metrics were employed that provided insights into their accuracy, precision, recall, specificity and overall predictive power.

Sensitivity: Also known as the true positive rate or recall, sensitivity is a performance metric that measures the proportion of actual positive instances that are correctly identified as positive by a classification model. In other words, sensitivity indicates how well a model can "sensitively" detect the positive class or identify true positives. Mathematically, sensitivity is calculated using the equation 4.1.

$$\text{Sensitivity} = TP / (TP + FN)$$

Equation (4.1)

In this equation, TP represents the number of true positives (correctly classified positive instances) and FN represents the number of false negatives (positive instances incorrectly classified as negative).

Sensitivity is typically expressed as a percentage and ranges from 0% to 100%. A sensitivity of 100% indicates that the model correctly identified all positive instances, while a sensitivity of 0% means that the model failed to detect any positive instances.

Specificity: It is a performance metric that measures the proportion of actual negative instances that are correctly identified as negative by a classification model. In other words, specificity indicates how well a model can "specifically" distinguish the negative class or identify true negatives. Mathematically, specificity is calculated using the equation 4.2.

$$\text{Specificity} = TN / (TN + FP)$$

Equation (4.2)

In this equation, TN represents the number of true negatives (correctly classified negative instances) and FP represents the number of false positives (negative instances incorrectly classified as positive).

Similar to sensitivity, specificity is typically expressed as a percentage and ranges from 0% to 100%. A specificity of 100% indicates that the model correctly identified all negative instances, while a specificity of 0% means that the model failed to distinguish any negative instances.

Precision: It is a performance metric that measures the proportion of positive predictions made by a classification model that are actually correct. In other words, precision indicates how "precise" or accurate the model's positive predictions are. Mathematically, precision is calculated using the equation 4.3.

$$\text{Precision} = TP / (TP + FP)$$

Equation (4.3)

Precision is typically expressed as a percentage and ranges from 0% to 100%. A precision of 100% indicates that all positive predictions made by the model are correct, while a precision of 0% means that all positive predictions are incorrect.

Accuracy: It is a performance metric that measures the overall correctness or accuracy of a classification model. It represents the proportion of correctly classified instances (both positive and negative) out of the total number of instances. Mathematically, accuracy is calculated using the equation 4.4.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Equation (4.4)

Accuracy is typically expressed as a percentage and ranges from 0% to 100%. An accuracy of 100% indicates that the model correctly classified all instances, while an accuracy of 0% means that the model failed to classify any instance correctly.

F1 score: It is a performance metric that combines precision and recall (sensitivity) into a single value. It provides a balanced measure of a classification model's accuracy, considering both the positive and negative class predictions. The F1 score is calculated using the harmonic mean of precision and recall, giving equal importance to both metrics. Mathematically, it is calculated using the equation 4.5.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity})$$

Equation (4.5)

The F1 score ranges from 0 to 1, with 1 representing perfect precision and recall and 0 representing the worst possible score. A high F1 score indicates that the model has high precision and high recall, striking a good balance between correctly identifying positive instances and minimising false positives and false negatives.

Receiver Operating Characteristic (ROC) curve: It is a graphical representation that illustrates the performance of a binary classification model at various classification thresholds. It plots the true positive rate (sensitivity) against the false positive rate (1-specificity) as the classification threshold is varied. Each point on the ROC curve represents a different threshold and the curve provides insights into the trade-off between sensitivity and specificity for the model.

The Area Under the ROC Curve (AUC): It is a scalar value that quantifies the overall performance of a classification model based on the ROC curve. It represents the probability that a randomly chosen positive instance will be ranked higher than a randomly chosen negative instance by the model. AUC ranges from 0 to 1, where AUC value of 0.5 indicates a model that performed no better than random guessing, whereas AUC value of 1 indicates a perfect model with flawless classification. Higher AUC values suggest better discrimination and predictive power of the model.

4.5.1. Nearest Neighbour Classification Model:

As part of this research, two K-Nearest Neighbours (KNN) models were developed to predict different classes of performance in attention and memory tasks. The first model, called attention model, aimed to predict classes of perfect performance and other performance groups in the attention task. Optimal performance was achieved by setting the number of neighbours to 4 and using the Jaccard distance metric, which measures the dissimilarity between two samples based on the presence or absence of features. The model was evaluated using leave-one-out cross-validation, a technique where each sample is left out once as a validation set while training on the remaining samples. The results of this evaluation are given in the Table 4.3 and Figure 4.1.

Table 4.3. Nearest Neighbour Attention Model Evaluation Metrics

Evaluation Metrics	Value
Sensitivity	95.40%
Specificity	27.42%
Precision	78.67%
Accuracy	77.54%
F1 score	0.86
Area Under the ROC Curve (AUC)	0.77

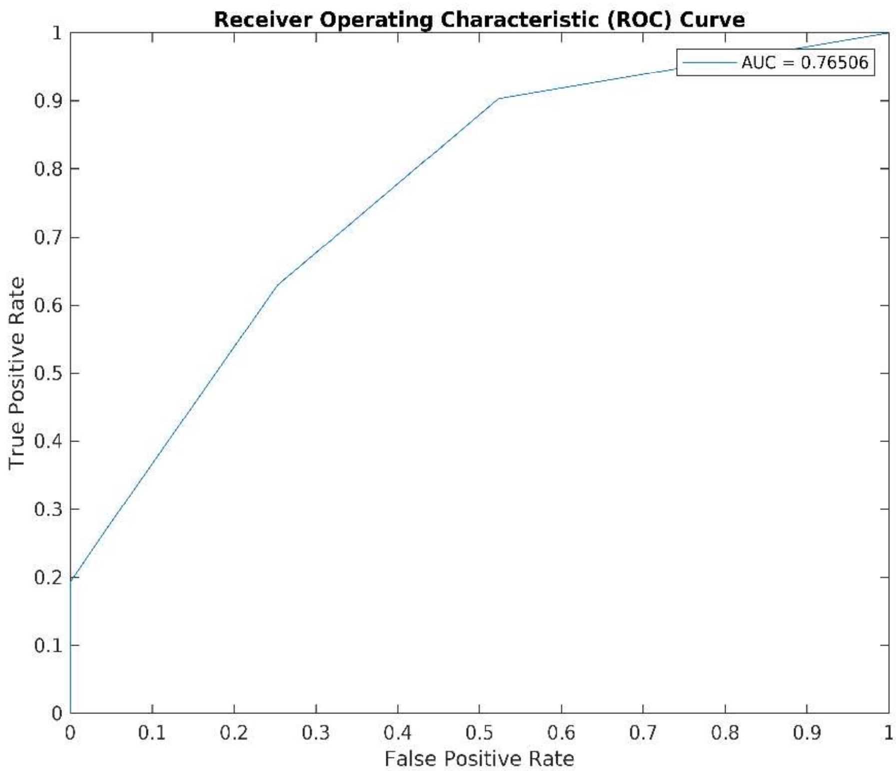

Figure 4.1. Nearest Neighbour Attention Model Receiver Operating Characteristic Curve

Similarly, the second model, the KNN memory model, was developed to predict classes of perfect performance and other performance groups in the memory task. The optimal configuration for this model also involved setting the number of neighbours to 4. The model was evaluated using leave-one-out cross-validation technique and the results of this evaluation are given in the Table 4.4 and Figure 4.2.

Table 4.4. Nearest Neighbour Memory Model Evaluation Metrics

Evaluation Metrics	**Value**
Sensitivity	87.57%
Specificity	44.78%
Precision	80%
Accuracy	75.42%
F1 score	0.84
Area Under the ROC Curve (AUC)	0.71

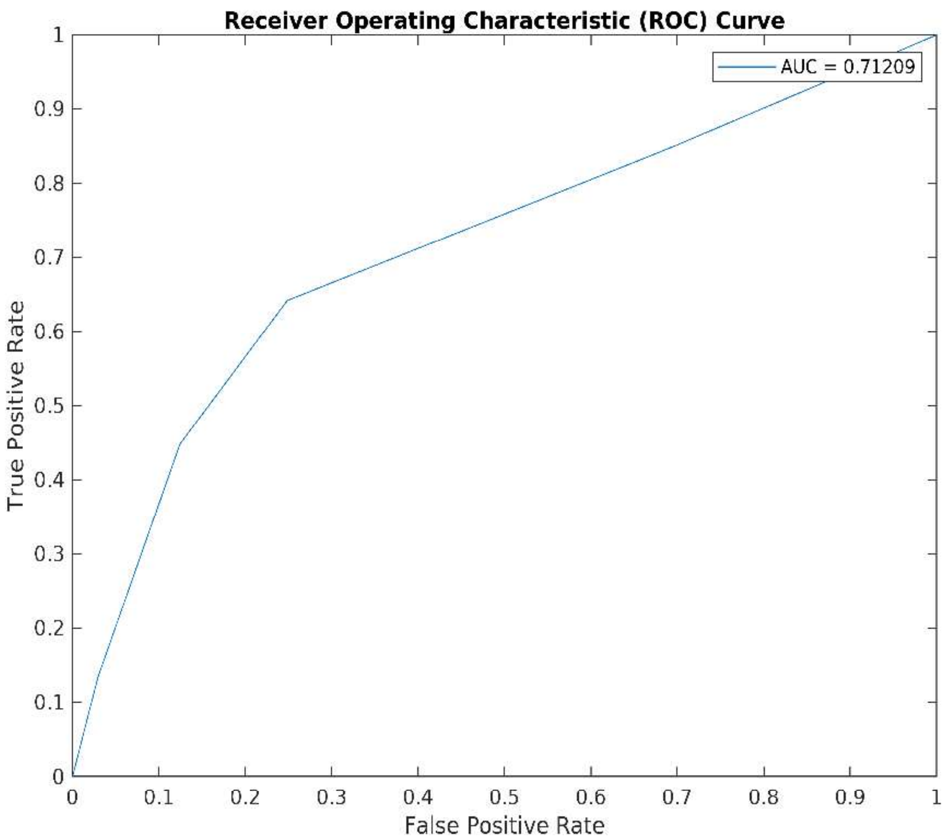

Figure 4.2. Nearest Neighbour Memory Model Receiver Operating Characteristic Curve

4.5.2. Decision Trees Model:

In addition to the KNN models, decision tree models were also developed to predict performance in attention and memory tasks. The decision tree attention model was trained using the decision tree algorithm and pruned to level 4. Pruning a decision tree machine learning model helps prevent overfitting by reducing the complexity of the tree and removing unnecessary branches or nodes. It aims to strike a balance between model accuracy on the training data and its ability to generalise well to unseen data, resulting in a more robust and less overfitted model. Finally, the model was evaluated using leave-one-out cross-validation technique and the results of this evaluation are given in the Table 4.5 and Figure 4.3.

Table 4.5. Decision Tree Attention Model Evaluation Metrics

Evaluation Metrics	**Value**
Sensitivity	94.25%
Specificity	53.23%
Precision	84.97%
Accuracy	83.47%
F1 score	0.89
Area Under the ROC Curve (AUC)	0.83

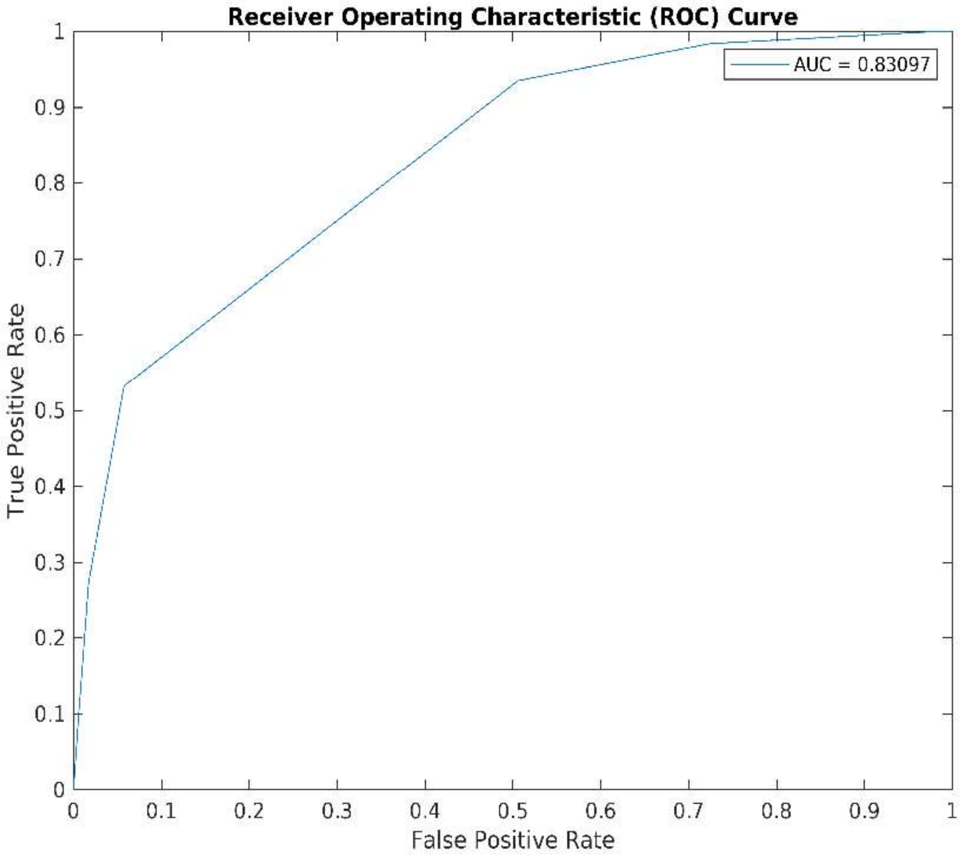

Figure 4.3. Decision Tree Attention Model Receiver Operating Characteristic Curve

Similarly, the decision tree memory model was trained using the decision tree algorithm and pruned to level 3. Finally, the model was evaluated using leave-one-out cross-validation technique and the results of this evaluation are given in the Table 4. 6 and Figure 4.4.

Table 4.6. Decision Tree Memory Model Evaluation Metrics

Evaluation Metrics	**Value**
Sensitivity	94.25%
Specificity	53.23%
Precision	84.97%
Accuracy	83.47%
F1 score	0.89
Area Under the ROC Curve (AUC)	0.83

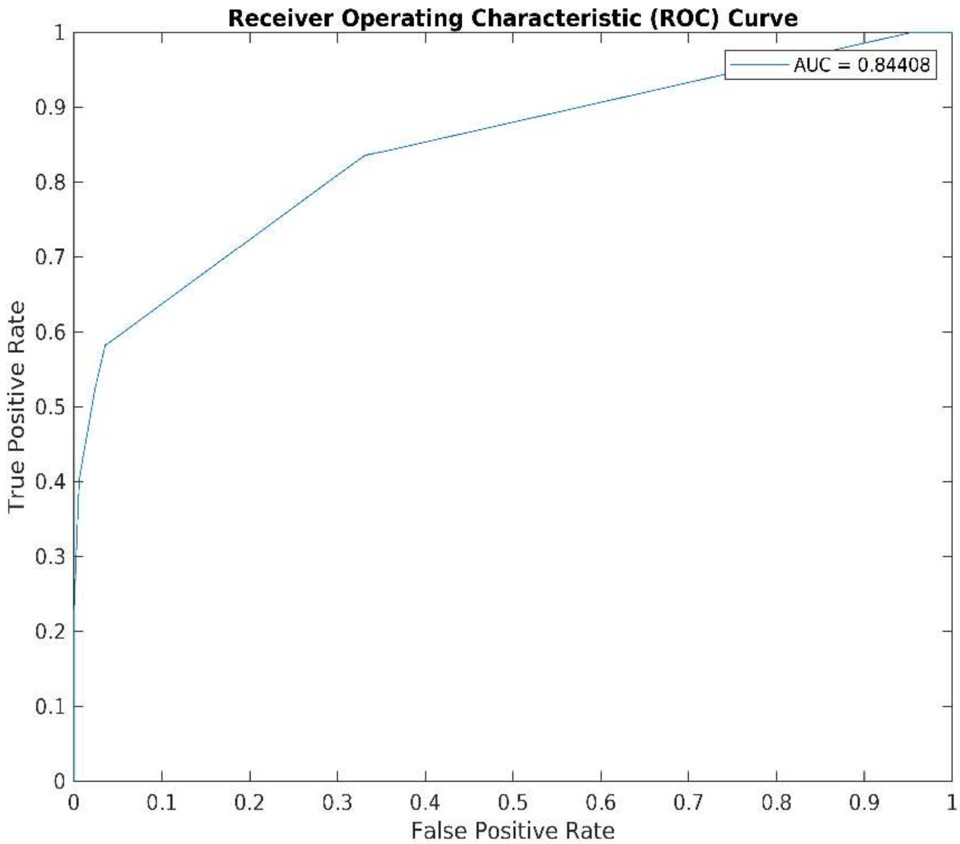

Figure 4.4. Decision Tree Memory Model Receiver Operating Characteristic Curve

4.5.3. Support Vector Machines Model:

In addition to the KNN and decision tree models, support vector machine (SVM) models were also developed to predict performance in attention and memory tasks. The SVM attention model was trained using the SVM algorithm and the "KernelFunction" parameter was set to the value "polynomial". A polynomial kernel calculates the similarity between data points by transforming them into a higher-dimensional feature space using polynomial functions, enabling the SVM model to capture nonlinear relationships in the data. Finally, the model was evaluated using leave-one-out cross-validation technique and the results of this evaluation are given in the Table 4.7 and Figure 4.5.

Table 4.7. Support Vector Machines Attention Model Evaluation Metrics

Evaluation Metrics	Value
Sensitivity	98.28%
Specificity	16.13%
Precision	76.68%
Accuracy	76.69%
F1 score	0.86
Area Under the ROC Curve (AUC)	0.72

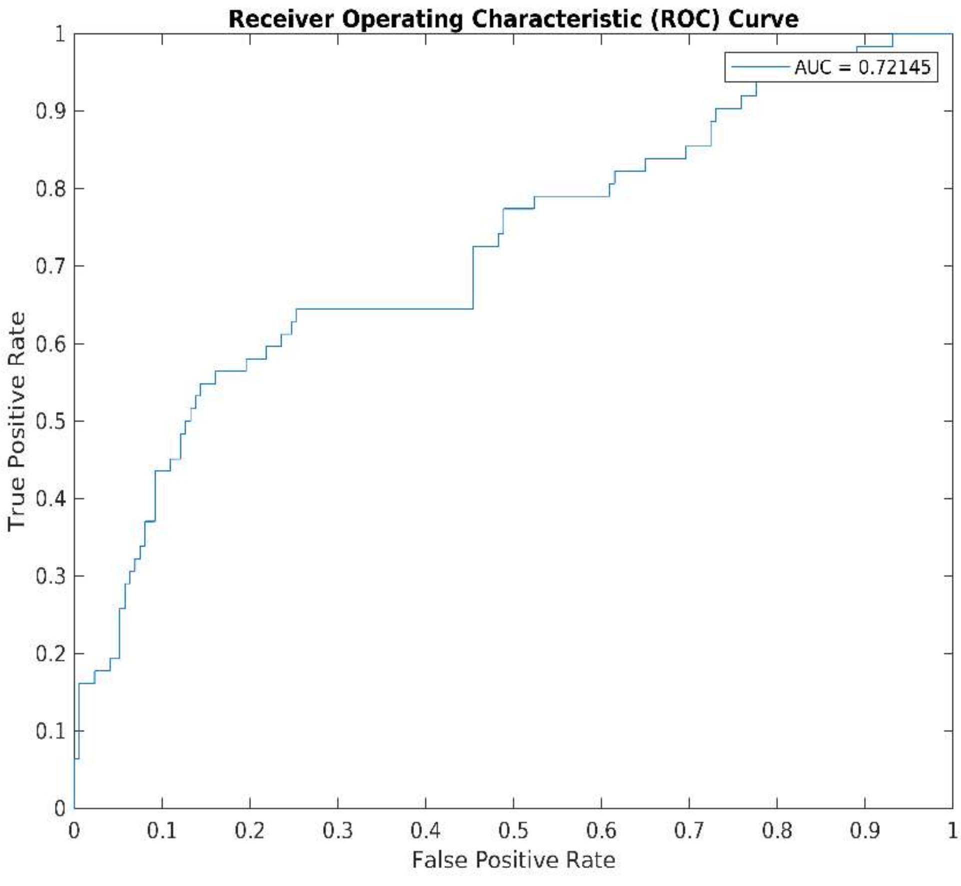

Figure 4.5. Support Vector Machines Attention Model Receiver Operating Characteristic Curve

Similarly, the SVM memory model was trained using the same settings and the model was evaluated using leave-one-out cross-validation technique. Results of the evaluation are given in the Table 4.8 and Figure 4.6.

Table 4.8. Support Vector Machines Memory Model Evaluation Metrics

Evaluation Metrics	Value
Sensitivity	97.04%
Specificity	46.27%
Precision	82.00%
Accuracy	82.63%
F1 score	0.89
Area Under the ROC Curve (AUC)	0.84

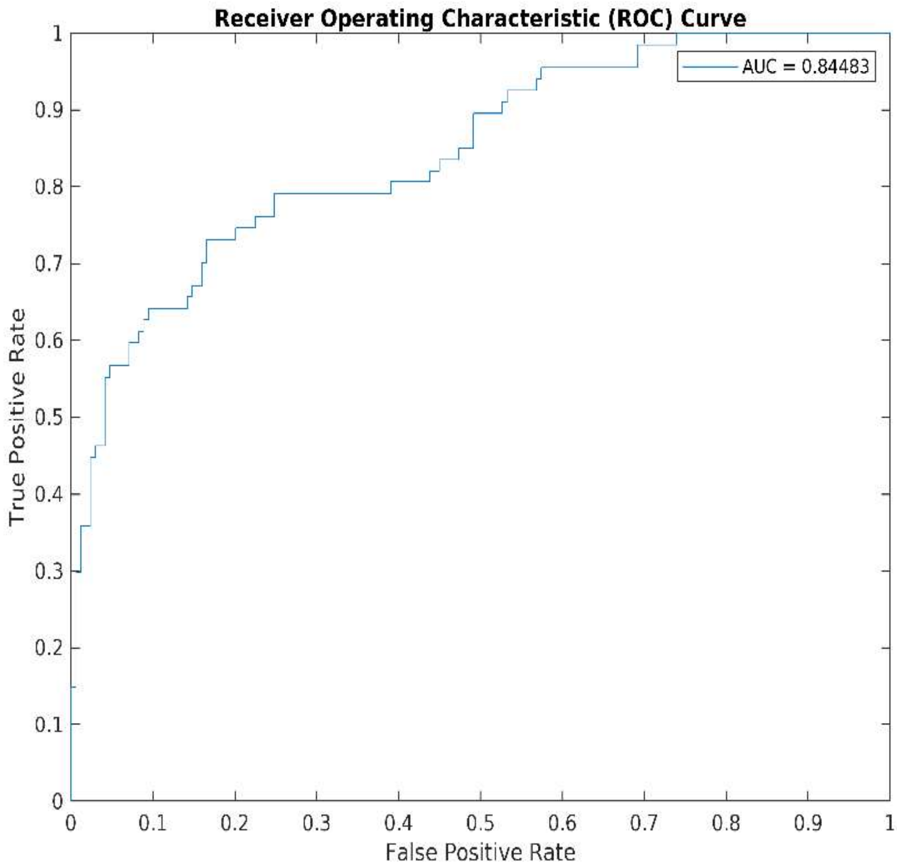

Figure 4.6. Table 4.9 Support Vector Machines Memory Model Evaluation Metrics Receiver Operating Characteristic curve

4.5.4. Artificial Neural Network Model:

Moreover, artificial neural network (ANN) models were developed to predict performance in attention and memory tasks. The neural network attention model was trained using a neural network with a hidden layer of 5 neurons. Finally, the model was evaluated using leave-one-out cross-validation technique and the results of this evaluation are given in the Table 4.9 and Figure 4.7.

Table 4.10. Artificial Neural Network Attention Model Evaluation Metrics

Evaluation Metrics	Value
Sensitivity	92.53%
Specificity	33.87%
Precision	79.70%
Accuracy	77.12%
F1 score	0.86
Area Under the ROC Curve (AUC)	0.80

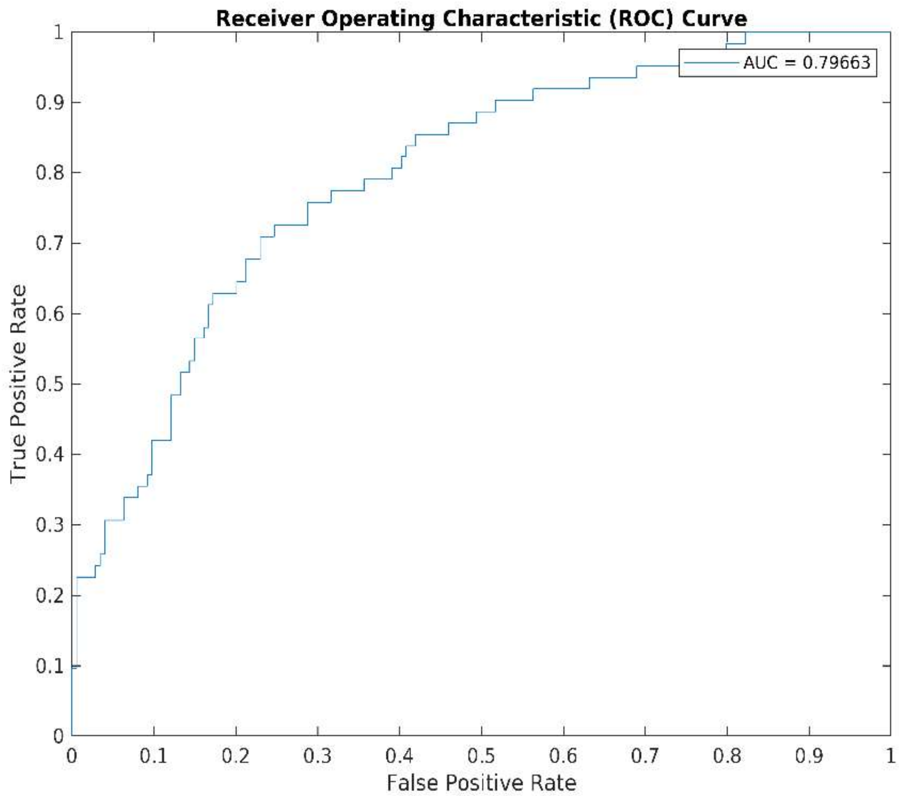

Figure 4.7. Artificial Neural Network Attention Model Receiver Operating Characteristic curve

Similarly, the neural network memory model was trained using the same settings, including a hidden layer of 5 neurons. The model was evaluated using leave-one-out cross-validation technique and the results of this evaluation are given in the Table 4.10 and Figure 4.8.

Table 4.11. Artificial Neural Network Memory Model Evaluation Metrics

Evaluation Metrics	**Value**
Sensitivity	94.67%
Specificity	56.72%
Precision	84.66%
Accuracy	83.90%
F1 score	0.89
Area Under the ROC Curve (AUC)	0.89

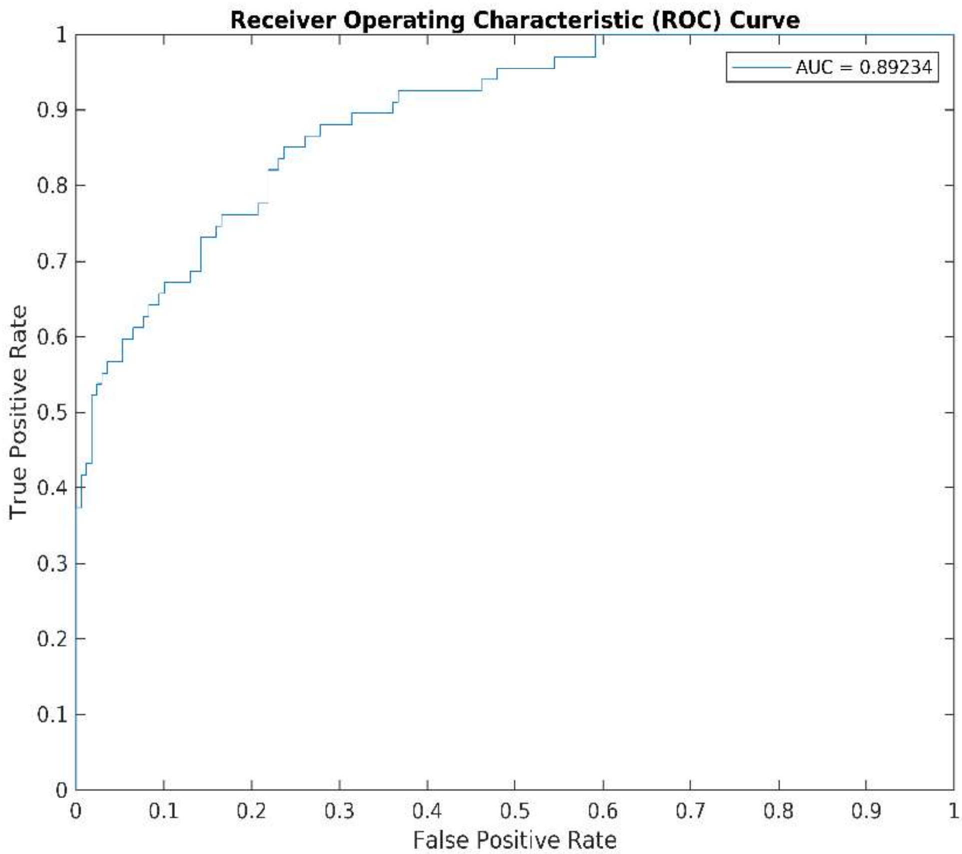

Figure 4.8. Artificial Neural Network Memory Model Receiver Operating Characteristic Curve

4.6. Discussion

The findings of this study have several significant implications in the field of cognitive performance prediction. Motivated from the previous works in cognitive science, this work focused on attentional and working memory status as important indicators of cognitive performance and utilised EEG recordings from the prefrontal cortex to assess these indicators [88]. As suggested by renewed researchers in the field, EEG was chosen among various physiological indicators thanks to its high temporal resolution and established connection with cognitive processes [89] [90].

This study also extracted eye blink features from the EEG data to capture and represent the attentional status accurately [91]. Eye blinks have been linked to cognitive processes and attentional engagement, making them relevant features for assessing cognitive performance. Moreover, the study also considered the power values in different frequency bands, further enriching the feature set for cognitive performance prediction [92].

To effectively utilise the extracted features, motivated by the previous studies, this work employed machine learning models [93] [94]. The use of machine learning algorithms allows for the identification of relationships in the data that can predict cognitive performance. The trial-and-error approach in developing the models helped refine the accuracy of the predictions (Machine Learning with MATLAB online course).

The developed machine learning models, including KNN, decision trees, support vector machines and neural networks, demonstrated the ability to predict performance in attention and memory tasks with reasonable accuracies ranging from 75.42% to 83.90%. A comparative study on the performance of the different ML models is given in the Table 4.11 and Table 4.12. Among the different machine learning techniques evaluated in this research, the decision tree model emerged as the best performing model based on the provided results. It achieved a sensitivity of 94.25%, indicating its ability to accurately identify individuals with positive attentional performance. The specificity of 53.23% suggests that it is capable of correctly classifying individuals with negative attentional performance. The precision of 84.97% highlights the model's ability to correctly identify true positive cases, minimising the occurrence of false positives.

A comparative study on the performance of the different ML attention models:

Table 4.12. A comparative study on the performance of the different ML attention models

Machine Learning Technique	Sensitivity	Specificity	Precision	Accuracy	F1 score	AUC
k-Nearest Neighbour	95.40%	27.42%	78.67%	77.54%	0.86	0.77
Decision Trees	94.25%	53.23%	84.97%	83.47%	0.89	0.83
Support Vector Machines	98.28%	16.13%	76.68%	76.69%	0.86	0.72
Artificial Neural Network	92.53%	33.87%	79.70%	77.12%	0.86	0.80

A comparative study on the performance of the different ML memory models:

Table 4.13. A comparative study on the performance of the different ML memory models

Machine Learning Technique	Sensitivity	Specificity	Precision	Accuracy	F1 score	AUC
k-Nearest Neighbour	87.57%	44.78%	80%	75.42%	0.84	0.71
Decision Trees	94.25%	53.23%	84.97%	83.47%	0.89	0.83
Support Vector Machines	97.04%	46.27%	82.00%	82.63%	0.89	0.84
Artificial Neural Network	94.67%	56.72%	84.66%	83.90%	0.89	0.89

In terms of overall accuracy, the decision tree model achieved an impressive accuracy of 83.47%, indicating that it can effectively predict attentional states with a high degree of accuracy. The F1 score of 0.89 demonstrates a good balance between precision and recall, indicating the model's robustness in classifying attentional performance accurately. Furthermore, the area under the curve (AUC) value of 0.83 indicates a strong discriminatory power of the model in distinguishing between different attentional states.

Among the machine learning techniques evaluated in this study, the artificial neural network (ANN) model stands out as the best performer based on the provided results in memory models. With a sensitivity of 94.67%, the ANN model demonstrated a high ability to correctly identify individuals with positive working memory performance. The specificity of 56.72% suggests its capability to accurately classify individuals with negative working memory performance. The precision of 84.66% indicates that the model has the capability to correctly identify true positive cases while minimising false positives.

In terms of overall accuracy, the ANN model achieved an impressive accuracy of 83.90%, indicating its effectiveness in predicting working memory performance states with a high degree of accuracy. The F1 score of 0.89 demonstrates a good balance between precision and recall, highlighting the model's robustness in classifying working memory performance accurately. Furthermore, the area under the curve (AUC) value of 0.89 indicates a strong discriminatory power of the model in distinguishing between different working memory states.

These results are in line with other studies, which got similar classification accuracy of 83% using pupillary response features and the AdaBoost algorithm [95]. In another study which used blink number and GSR frequency power as features, SVM and Naïve Bayes machine learning algorithms got an accuracy of 71% and 75% respectively, indicating that the chosen combination of features from EEG and eye blink data in this study, along with the machine learning approach, yielded promising results in predicting attentional performance.

The significance of these findings lies in their potential application in various domains. Accurately predicting cognitive performance can have practical implications in fields such as education, workforce performance optimisation and human-machine interactions.

By understanding an individual's attentional and working memory status through non-invasive EEG recordings, interventions and strategies can be tailored to enhance cognitive performance in challenging tasks or situations.

It is important to note that the findings of this study contribute to the existing body of research that supports the use of EEG and eye blink features for assessing cognitive performance. The high accuracy achieved by the decision tree and neural network model reinforces the potential of these indicators and the machine learning approach for predicting cognitive performance.

In summary, the findings of this study highlight the significance of attentional and working memory indicators, EEG recordings and eye blink features in predicting cognitive performance. The accurate machine learning models provide valuable insights and pave the way for future research and practical applications in optimising cognitive performance in various domains.