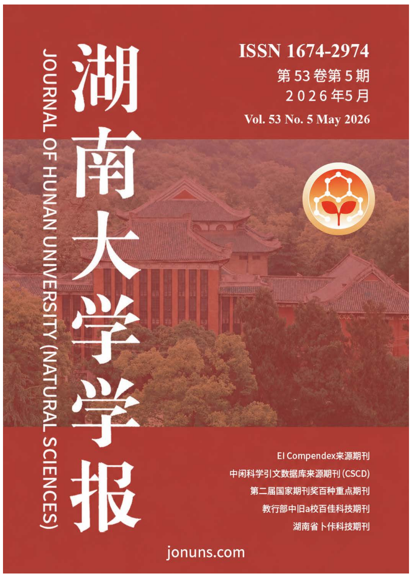


Journal of Hunan University

(Natural Sciences)

Vol. 53 No. 5

May 2026

Available online at

https://jonuns.com

Open Access Article https://doi.org/10.55463/issn.1674-2974.53.5.5

Optimizing Football Match Outcome Prediction : A Comparative Study of Boosting Algorithms and Deep Learning on Tabular Sports Data

Gbenga O Ogunsanwo*[1], Ayokunle A. Omotunde[2], Olumuyiwa B. Alaba1, Oluwatimilehin P. Orisadare[1], Funmilayo F Amurawaye[3]

[1]Department of Computer Science, College of Science and Information Technology, Tai Solarin University of Education, Ogun State, Nigeria,

[2]Department of Computer Science, Faculty of Science , Babcock University , Ilishan Remo, Ogun State, Nigeria,

[3]Department of Mathematics, College of Science and Information Technology, Tai Solarin University of Education, Ogun State, Nigeria,

* Corresponding author: ogunsanwogo@tasued.edu.ng

Article History:

Received: March 9, 2026
Revised: April 30, 2026
Accepted: May 17, 2026
Published: May 29, 2026

Abstract: Predicting event outcomes, particularly in sports, has attracted increasing research attention due to the growing availability of historical and performance-related data. Football match outcome prediction has traditionally relied on expert judgment, statistical analysis of past results, and qualitative assessments of team strengths and weaknesses; however, such approaches may be limited by subjectivity, incomplete feature representation, and restricted predictive consistency. This study develops and compares predictive models for football match outcomes using ensemble learning and deep learning algorithms applied to tabular sports data. A publicly available football match dataset obtained from Kaggle was used, and five algorithms were implemented: Deep Neural Network (DNN), TabTransformer, Neural Oblivious Decision Ensembles (NODE), XGBoost, and LightGBM. Model performance was evaluated using standard classification metrics, including precision, recall, F1-score, and accuracy. The results show that the deep learning models achieved moderate predictive performance, with accuracies ranging from 78% for NODE to 87% for the best-performing deep learning model. In contrast, XGBoost demonstrated strong performance across all metrics, achieving 0.96 precision, 0.96 recall, 0.95 F1-score, and 96% accuracy. LightGBM achieved the highest overall performance, with 0.98 precision, 0.98 recall, 0.98 F1-score, and 99% accuracy. These findings indicate that LightGBM is the most effective model for this tabular classification task, followed closely by XGBoost. Although the deep learning models, particularly TabTransformer, show potential, they did not outperform the boosting algorithms in this evaluation. The study recommends the use of ensemble-based algorithms for football match outcome prediction, especially when working with structured tabular datasets. Future research may extend this work by applying advanced hyperparameter optimization techniques, such as grid search, random search, or Bayesian optimization, to further improve the performance of LightGBM and XGBoost.

Keywords: football match prediction; match outcome classification; ensemble learning; deep learning; XGBoost; LightGBM; tabular data.

优化足球比赛结果预测：基于表格型体育数据的 Boosting 算法与深度学习比较研究

摘要：

基于历史数据预测事件结果，尤其是体育赛事结果，随着历史数据和绩效相关数据可用性的不断提高，已受到越来越多研究者的关注。足球比赛结果预测传统上主要依赖专家判断、以往比赛结果的统计分析以及对球队优势与劣势的定性评估；然而，这些方法可能受到主观性、特征表示不完整以及预测一致性有限等因素的限制。本研究基于表格型体育数据，采用集成学习算法和深度学习算法，构建并比较足球比赛结果预测模型。研究使用了来自 Kaggle 的公开足球比赛数据集，并实现了五种算法：深度神经网络（Deep Neural Network, DNN）、TabTransformer、神经遗忘决策集成模型（Neural Oblivious Decision Ensembles, NODE）、XGBoost 和 LightGBM。模型性能采用标准分类评价指标进行评估，包括精确率、召回率、F1 值和准确率。结果表明，深度学习模型取得了中等水平的预测性能，其准确率从 NODE 的 78% 到表现最佳的深度学习模型的 87% 不等。相比之下，XGBoost 在各项评价指标上均表现出较强性能，精确率为 0.96，召回率为 0.96，F1 值为 0.95，准确率为 96%。LightGBM 取得了最高的整体性能，精确率为 0.98，召回率为 0.98，F1 值为 0.98，准确率为 99%。这些结果表明，在该表格型分类任务中，LightGBM 是最有效的模型，其次是 XGBoost。尽管深度学习模型，特别是 TabTransformer，显示出一定潜力，但在本次评估中未能超越 boosting 算法。本研究建议在足球比赛结果预测模型开发中，尤其是在处理结构化表格数据时，优先采用基于集成学习的算法。未来研究可进一步采用高级超参数优化技术，如网格搜索、随机搜索或贝叶斯优化，以进一步提升 LightGBM 和 XGBoost 的性能。

关键词：

足球比赛预测；比赛结果分类；集成学习；深度学习；XGBoost；LightGBM；表格数据。

1. Introduction

Predicting the outcome of football matches is a captivating and challenging problem that has long fascinated sports enthusiasts, analysts, and researchers alike. The inherent unpredictability of the sport, influenced by a myriad of factors ranging from team form and player injuries to tactical approaches and even referee decisions, makes accurate forecasting a complex task. Historically, football match prediction has often relied on expert opinion, statistical analysis of past results, and qualitative assessments of team strengths and weaknesses [1]. However, with the exponential growth of available sports data and advancements in computational power, machine learning techniques have emerged as powerful tools to approach this problem [2]. These methods allow for the identification of intricate patterns and relationships within large datasets that might be difficult or impossible for human analysts to discern. Data analysis techniques such as time series, clustering, and correlation are essential for uncovering patterns and trends in large datasets [3].

The application of machine learning to sports prediction is an active area of research, encompassing various sports beyond football. Within football, researchers have explored a range of algorithms, including logistic regression, support vector machines, decision trees, and ensemble methods like gradient boosting (XGBoost and LightGBM), to build predictive models. For instance, [4] used deep learning (DNN) and support vector regression (SVR) to develop a model for footballer age prediction. The goal of applying ML is often to predict the final match result (Home Win, Away Win, or Draw) based on features derived from historical match data, team statistics, and potentially even external factors like weather or betting odds [5]. There are numerous studies that have applied machine learning in predicting outcomes across different sectors. In football prediction, for example, [6] developed a machine learning framework capable of predicting football match outcomes using readily available match data, with a strong emphasis on integrating domain knowledge and handling the complexity of time series data from competing teams. The models used included KNN, ANN, NB, and RF. The best performing models were KNN and ANN, which showed top performance, particularly in the context of soccer prediction. [7] conducted a study to explore how event level football match data (passes, carries, shots, etc.) can be used to accurately predict match outcomes using machine learning. Their research focused on data driven feature engineering and evaluated the effectiveness of various ML algorithms in modeling complex in game patterns and predicting results. The models used were SVM, RF, and XGBoost. The best performing model was SVM, which achieved the highest overall accuracy of 65.8 percent on the test set. The dataset was obtained from StatsBomb's open access football event data. [8] examined the performance of deep learning models and optimized gradient boosted trees for predicting soccer match outcomes. The study employed the CatBoost model with pi ratings as features and compared it against deep learning approaches. [9] carried out research to explore how machine learning can be applied to predict sports outcomes, particularly in football, using historical match data. The models used included LR, DT, SVM, and NN. The best performing model was ANN, which achieved an accuracy of 60 percent. The dataset was obtained from Kaggle.com, Livescore.com, and Rezultati.com for the 2019/2020 season. [10] developed a robust framework for predicting soccer match outcomes (win/draw/loss) and over/under 2.5 goals using enhanced ML and DL models. The models employed included Logistic Regression (LR), XGBoost, RF , SVM), NB, Feedforward Neural Networks (FNN), and Recurrent Neural Networks (RNN). Their study revealed that the best performing model was a Voting Ensemble combining RF and XGBoost, which achieved an accuracy of 83 percent for match outcomes and 84 percent for over/under 2.5 goals. The dataset was sourced from football-data.co.uk.

Despite these advances, predicting football match outcomes remains inherently probabilistic. The dynamic nature of the sport, influenced by player form, tactical decisions, and unpredictable events, means that no model can guarantee perfect accuracy [11].. Nevertheless, the pursuit of more sophisticated and data driven prediction methods continues, driven by applications in sports analytics, betting strategies, and enhancing fan engagement. Building on this foundation, the present study contributes to ongoing research by comparing ensemble machine learning models and deep learning architectures for football outcome prediction. Specifically, we explore the performance of LightGBM, XGBoost, Recurrent Neural Networks (RNNs), Tab Transformer, and NODE in capturing sequential dependencies and complex feature interactions within match data. The goal is to evaluate the effectiveness of gradient boosting methods (XGBoost, LightGBM) against deep learning architectures (DNN, Tab Transformer, NODE) using classification metrics such as Precision, Recall, F1 Score, and Accuracy.

The core challenge addressed is identifying the model that best leverages tabular match data to minimize prediction errors and reliably forecast outcomes. While gradient boosting methods have consistently demonstrated superior performance on structured datasets, their dominance has limited exploration of alternative deep learning architectures in sports analytics. This study introduces NODE (Neural Oblivious Decision Ensembles) and the Tab Transformer into football outcome prediction, representing a novel methodological expansion that broadens the scope of machine learning applications in sports analytics.

2. Material and Methods

This section discusses the processes and techniques used in developing football outcome predictive models using ensemble and deep learning algorithms, as shown in Figure 1. These include data collection, data preprocessing, model development, model validation, and prediction.

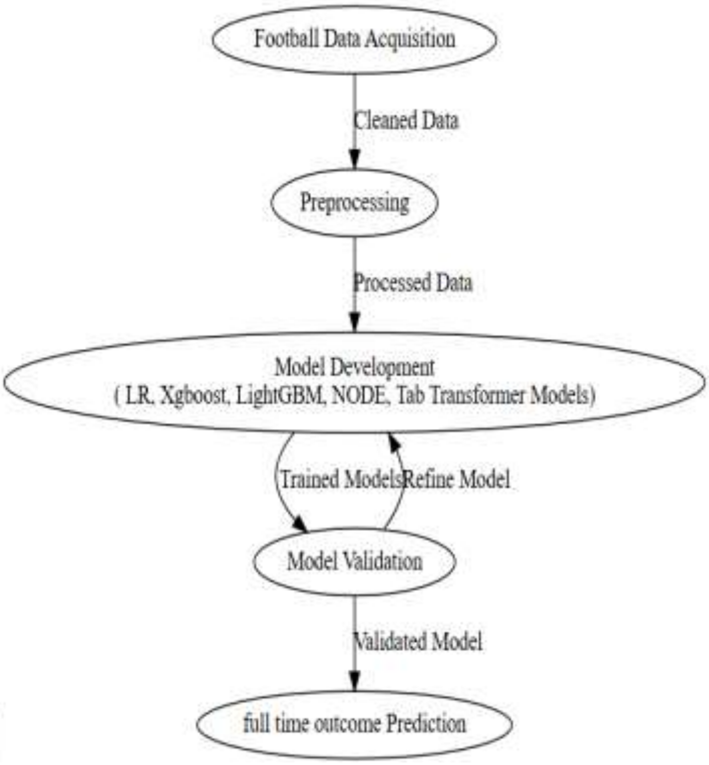


Figure 1: Work flow of the football prediction

Data Acquisition

The dataset used in the football full-time outcome prediction was downloaded from kaggle.com, containing 20 features and 382 instances, as shown in Table 1.

Table 1 : Attibutes Description

S/NO	Attributes Name	Datatypes
0	HomeTeam	object
1	AwayTeam	object
2	Full time home goals	int64
3	Full time away goals	int64
4	Full time result	int64
5	Half time home goals	int64
6	Half time away goals	int64
7	Half time result	int64
8	Referee	object
9	Home shots	int64
10	Away shots	int64
11	Home shots on target	int64
12	Away shots on target	int64
13	Home fouls	int64
14	Away fouls	int64
15	Home corners	int64
16	Away corners	int64
17	Home yellows	int64
18	Away yellows	int64
19	Home reds	int64
20	Away reds	int64

The Ensembles and Deep learning Techniques

Logistic Regression (LR): As a fundamental classification algorithm, LR serves as a solid baseline. It is computationally efficient and interpretable, making it useful for understanding the linear relationships between features and the target outcome. Its performance provides a benchmark against which more complex models can be compared.

XGBoost and LightGBM

These are powerful gradient boosting algorithms known for their high performance and ability to handle complex, non-linear relationships in data. They are popular choices in competitive machine learning due to their speed and accuracy. Their inclusion allows for an assessment of the predictive power of tree-based ensemble methods on this dataset.

Recurrent Neural Network (RNN)

The inclusion of an RNN represents an attempt to leverage the sequential nature of football data. Unlike the other models that treat each match as an independent event, an RNN is designed to process sequences, potentially capturing the influence of past match performance on the current outcome. The conceptual model structure, using separate LSTMs for home and away teams and concatenating their outputs, is a common approach for handling pairwise sequence data in sports.

Tab Transformer

Tab Transformer is a deep learning architecture that performs well on tabular datasets with categorical features. It closes the gap often seen in neural networks, which tend to underperform on tabular data compared to traditional machine learning models, by introducing transformer-based attention mechanisms to train context-aware representations of categorical features. This model was introduced to work effectively on tabular datasets [12].

Neural Oblivious Decision Ensembles (NODE)

NODE is a novel neural architecture built specifically for tabular datasets. It combines the power of decision tree ensembles with deep learning, enabling end-to-end differentiable learning of decision trees within a neural network framework. This model was introduced to work on tabular datasets using symmetric decision trees, where the same feature and threshold are applied across a given tree depth level, reducing complexity and making the trees highly regularized and efficient [13].

Model Validation

The football outcome predictive model employed several metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC to validate performance.

Accuracy: Measures the overall proportion of correctly classified instances.
Classification Report: Provides precision, recall, and F1-score for each class, as well as macro and weighted averages. This is particularly useful for understanding the model's performance on individual classes, especially in multi-class classification.
Confusion Matrix: A visual representation of the classification results, showing the counts of true positives, true negatives, false positives, and false negatives for each class. It helps identify which classes are being misclassified. The confusion matrix used in this study is a 3x3 matrix, as shown in Table 2.

ROC Curve and AUC (for Keras models): The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are used to evaluate the model's ability to distinguish between classes

Table 2 : 3x3 confusion matrix for the model

	Predicted H	**Predicted D**	**Predicted A**
True H	1,1	1,2	1.3
True D	2,1	2,2	2,3
True A	3,1	3,2	3,3

3. Results and Discussion

This section discusses results of the predictive model

The Results of Data Acquisition

The football dataset downloaded from Kaggle.com for prediction, which contains 20 features and 382 instances, was inputted into Google Colab as shown in Figure 2.

	Div	Date	HomeTeam	AwayTeam	Full time home goals	Full time away goals	Full time result	Half time home goals	Half time away goals	Ha... t... resu...
0	E0	13/08/05	Aston Villa	Bolton	2	2	D	2	2	
1	E0	13/08/05	Everton	Man United	0	2	A	0	1	
2	E0	13/08/05	Fulham	Birmingham	0	0	D	0	0	
3	E0	13/08/05	Man City	West Brom	0	0	D	0	0	
4	E0	13/08/05	Middlesboro	Liverpool	0	0	D	0	0	

5 rows × 23 columns

Figure 2 :Sample of Dataset used for the Prediction

The Results of the Pre processing

The dataset acquired was divided into 80% training and 20% testing as shown in Figure 3 .

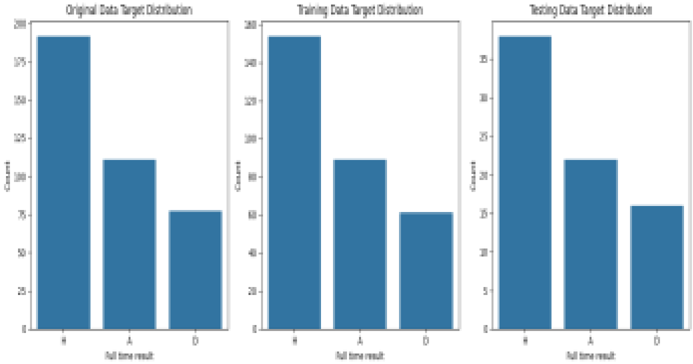


Figure 3 :Dataset Split

Thereafter, the dataset was subjected to preprocessing in order to improve the performance of the football outcome predictive model, as shown in Figure 4.

```
0.  0.  0.  0.  0.  0.
0.  0.  0.  0.  0.  0.
0.  0.  0.  0.  0.  0.
0.  0.  0.  0.  0.  0.
0.  0.  0.  0.  0.  0.
1.  0.  1.  0.  0.  0.
0.  0.  0.  0.  0.  0.
0.  0.  0.  0.  1.  0.
0.  0.  0.  0.  0.  0.
0.  0.  0.  ]]
```

First 5 rows of processed training data (as a DataFrame):

	num__Full time home goals	num__Full time away goals	num__Half time home goals	num__Half time away goals	num__Home shots	num__Away shots	num__Home shots on target
0	-0.392658	-0.987880	-0.824059	-0.656648	0.771934	-0.819594	1.682956
1	-1.167775	-0.987880	-0.824059	-0.656648	0.134827	-0.568428	0.453104
2	1.932693	-0.987880	0.330383	-0.656648	-0.289912	0.436236	0.453104
3	-1.167775	1.827578	-0.824059	2.195079	0.347196	-0.317262	0.453104
4	-0.392658	0.889092	0.330383	2.195079	0.347196	0.185070	-0.161823

Figure 4 : Preprocessing of the Dataset

4. Results of the Model Development

DNN Model

After the simulation process for the prediction of football outcomes was carried out using DNN, the evaluation of the model's performance was recorded, as shown in Figure 5. Table 2 presents the classification reports of the model. The values 0, 1, and 2 represent Away Win, Draw, and Home Win, respectively.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.78      0.44      0.56        16
           1       0.88      1.00      0.94        38
           2       0.79      0.86      0.83        22

    accuracy                           0.84        76
   macro avg       0.82      0.77      0.77        76
weighted avg       0.83      0.84      0.83        76
```

Figure 5 :DNN classification report

Table 3 : DNN clasifcation report

	Precision	**Recall**	**F1 Score**

A	0.78	0.44	.0.56
D	0.88	1.00	0.94
H	0.79	0.86	0.83
Accuracy			0.84
Macro Avg	0.82	0.77	0.77
Weighted avg	0.83	0.84.	0.83

The results of the model, as revealed in Table 3, show that the DNN model has 84% accuracy and other results of metrics used, such as 0.82 for Precision, 0.77 for Recall, and 0.77 for F1 Score. Figure 6 shows the confusion matrix report for the DNN model developed for the prediction of football outcomes. True H, Predicted H (7): The model correctly predicted 7 Home Wins ('H'). True D, Predicted D (Middle: 38): The model correctly predicted 38 Draws ('D'). True A, Predicted A (Bottom-Right: 19): The model correctly predicted 19 Away Wins ('A'). The diagonal numbers are the largest in their respective rows and columns, indicating the model is generally performing better at classifying samples correctly than incorrectly. The model seems particularly good at predicting Draws (38 correct out of a relatively small number of misclassifications).

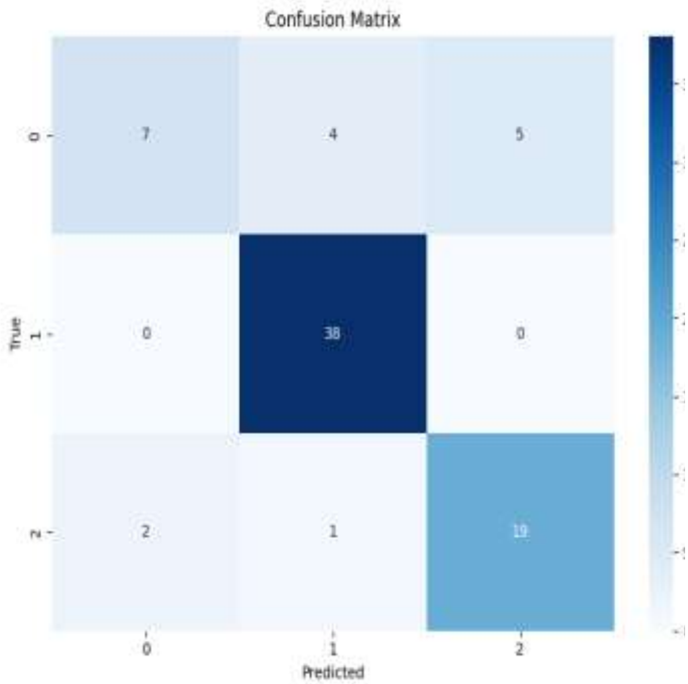
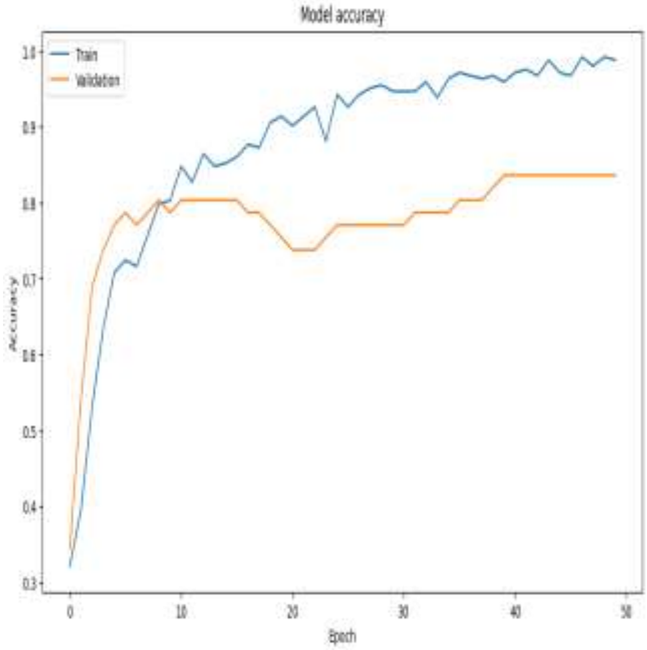


Figure 6 :DNN confusion matrix

The model accuracy and loss graph for DNN are shown in Figure 7 and Figure 8 respectively. The graph shows that model is learning effectively as the epoch is increasing the model accuracy is increasing which indicates that model performs excellently.



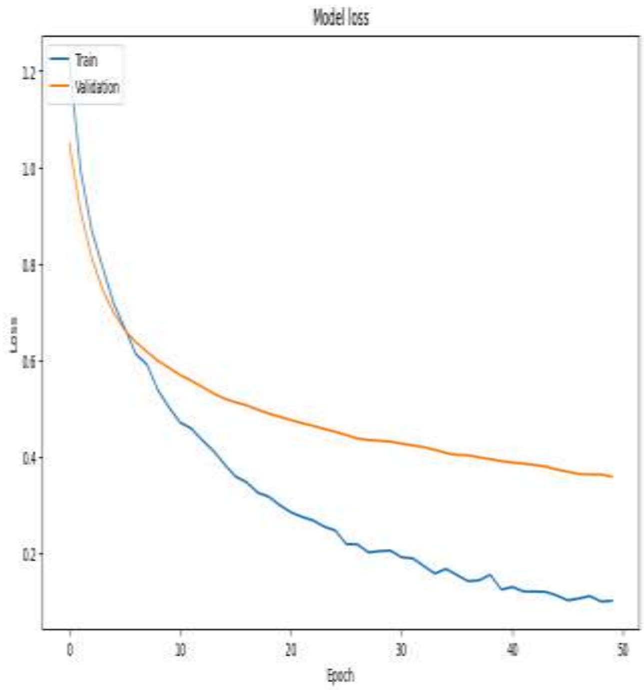


Figure 7 :Model Accuracy　　Figure 8 :Loss Val

Tab Transformer Model

After the simulation process for the prediction of football outcome was done using Tab Transformer, the evaluation of the performance of the model was registered as seen in Figure 9 and Table 4 reveals the classification reports of the model as the 0,1,2 represents Away win , Draw and Home win respectively.

	precision	recall	f1-score	support
0	0.88	0.44	0.58	16
1	0.86	0.97	0.91	38
2	0.88	1.00	0.94	22
accuracy			0.87	76
macro avg	0.87	0.80	0.81	76
weighted avg	0.87	0.87	0.85	76

Figure 9 :Tab Transformer classification report

Table 4: Tab Transformer clasifcation report

	Precision	**Recall**	**F1 Score**
A	0.88	0.44	.0.58
D	0.86	0.97	0.91
H	0.88	1.00	0.94
Accuracy			0.87
Macro Avg	0.87	0.80	0.81
Weighted avg	0.87	0.87.	0.85

The results of model as revealed in Table 4 shows the Tab Transformer model has 87% accuracy and other results of metrics used like 0.87 for Precision, 0.80 for Recall and 0.81 for F1 Score , The Figure 10 shows the confusion matrix report for the model developed for the prediction of football outcome. True H, Predicted H (7): The model correctly predicted 7 Home Wins ('H'). True D, Predicted D (Middle: 37): The model correctly predicted 37 Draws ('D'). True A, Predicted A (Bottom-Right: 22): The model correctly predicted 22 Away Wins ('A').. The diagonal numbers are the largest in their respective rows and columns, indicating the model is generally performing better at classifying samples correctly than incorrectly. The model seems particularly good at predicting Draws (37 correct out of a relatively small number of mis classifications

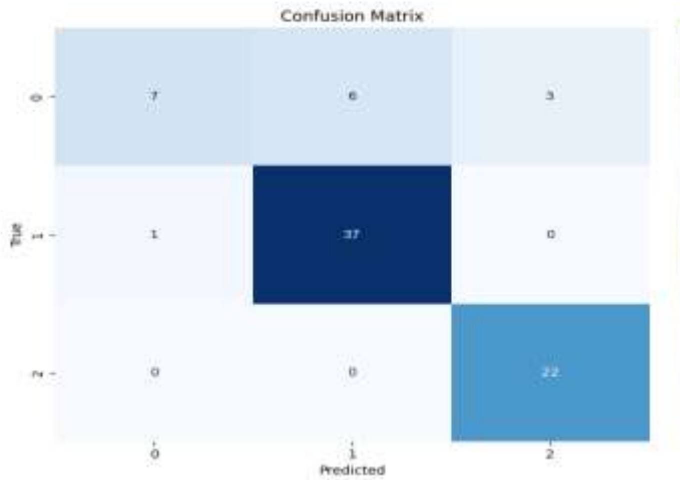
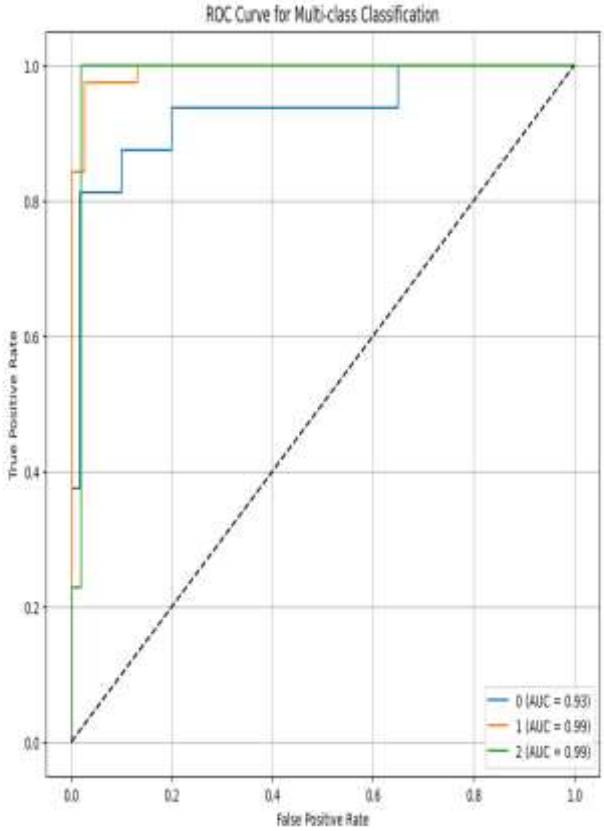


Figure 10: Tab Transformer confusion matrix

The model accuracy and loss graph for Tab Transformer are shown in Figure 11 and Figure 12 respectively. The graph shows that the model is learning effectively; as the epoch increases, the model accuracy increases, which indicates that the model performs excellently.



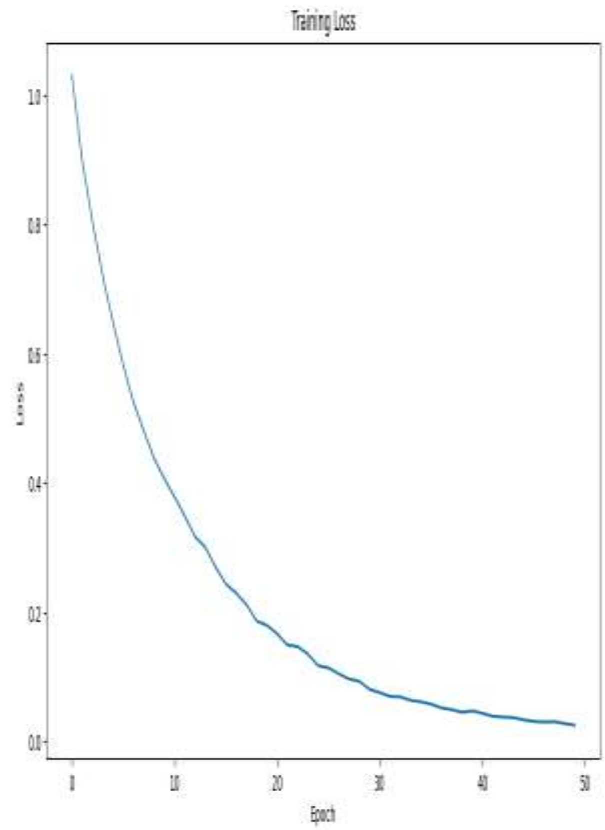


Figure 11 :Tab Transformer ROC curve Figure12: Loss Val

Neural Oblivious Decision Ensembles (NODE) Model

After the simulation process for the prediction of football outcomes was done using NODE, the evaluation of the performance of the model was registered, as seen in Figure 13, and Table 5 reveals the classification reports of the model, as the values 0, 1, and 2 represent Away Win, Draw, and Home Win respectively.

```
              precision    recall  f1-score   suppor

           0       0.56      0.31      0.40
           1       0.92      0.92      0.92
           2       0.66      0.86      0.75

    accuracy                           0.78
   macro avg       0.71      0.70      0.69
weighted avg       0.77      0.78      0.76
```

Figure 13: NODE classification report

Table 5 : NODE clasifcation report

	Precision	Recall	F1 Score
A	0.56	0.31	.0.40
D	0.92	0.92	0.93
H	0.66	0.96	0.75
Accuracy			0.78
Macro Avg	0.71	0.70	0.69
Weighted avg	0.77	0.78.	0.76

The results of the model, as revealed in Table 5, show that the NODE model has 78% accuracy and other results of metrics used, like 0.71 for Precision, 0.70 for Recall, and 0.69 for F1 Score. Figure 14 shows the confusion matrix report for the model developed for the prediction of football outcomes. True H, Predicted H (5): The model correctly predicted 5 Home Wins ('H'). True D, Predicted D (Middle: 35): The model correctly predicted 35 Draws ('D'). True A, Predicted A (Bottom-Right: 19): The model correctly predicted 19 Away Wins ('A'). The diagonal numbers are the largest in their respective rows and columns, indicating the model is generally performing better at classifying samples correctly than incorrectly. The model seems particularly good at predicting Draws (35 correct out of a relatively small number of misclassifications).

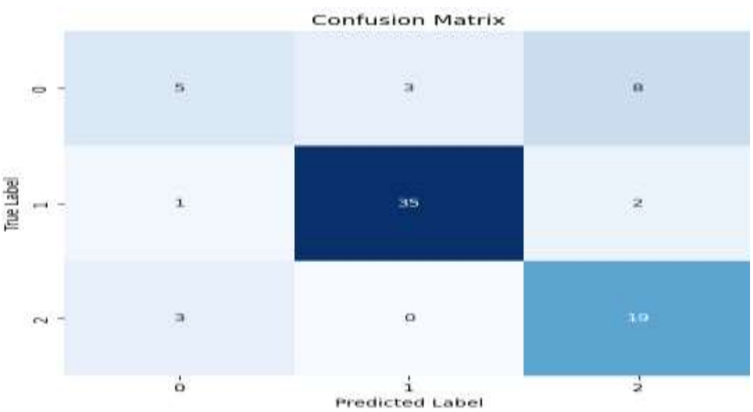
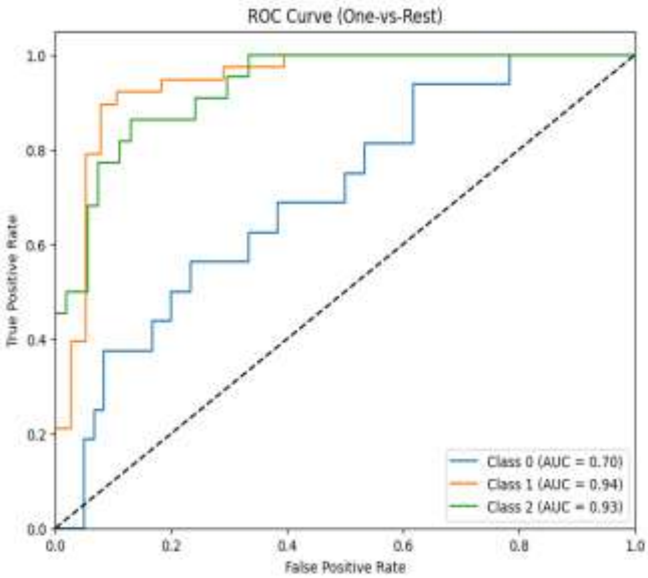


Figure 14 :NODE confusion matrix

The model accuracy and loss graph for Tab Transformer are shown in Figure 15 and Figure 16 respectively. The graph shows that the model is learning effectively; as the epoch is increasing, the model accuracy is increasing, which indicates that the model performs excellently.



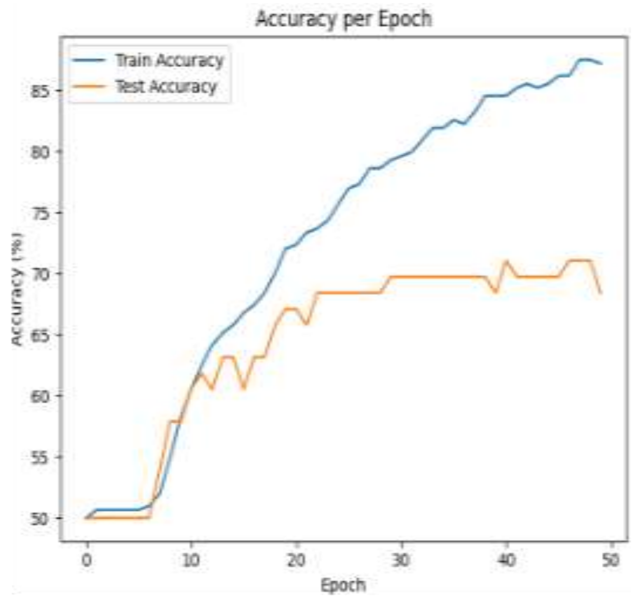
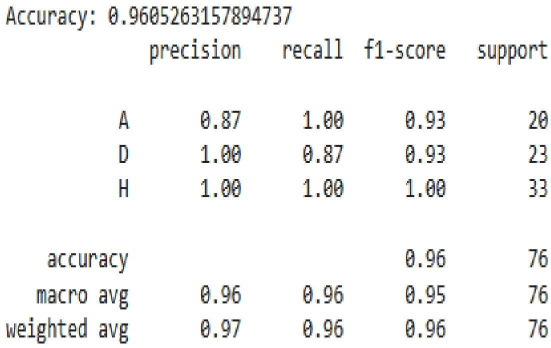


Figure 15 :NODE ROC curve
Figure 16: Accuracy graph respectively

XGBoost Model

After the simulation process for the prediction of football outcomes was done using XGBoost, the evaluation of the performance of the model was registered, as seen in Figure 17, and Table 6 reveals the classification reports of the model, as the values 0, 1, and 2 represent Away Win, Draw, and Home Win respectively.



Accuracy: 0.9605263157894737

	precision	recall	f1-score	support
A	0.87	1.00	0.93	20
D	1.00	0.87	0.93	23
H	1.00	1.00	1.00	33
accuracy			0.96	76
macro avg	0.96	0.96	0.95	76
weighted avg	0.97	0.96	0.96	76

Figure 17: XGBoost classification report

Table 6: XGBoost classificatio report

	Precision	Recall	F1 Score
A	0.87	1.00	.0.93
D	1.00	0.87	0.93
H	1.00	1.00	1.00
Accuracy			0.96
Macro Avg	0.96	0.96	0.95
Weighted avg	0.97	0.96	0.96

The results of the model, as revealed in Table 6, show that the NODE model has 96% accuracy and other results of metrics used, like 0.96 for Precision, 0.96 for Recall, and 0.95 for F1 Score. Figure 18 shows the confusion matrix report for the model developed for the prediction of football outcomes. True H, Predicted H (16): The model correctly predicted 16 Home Wins ('H'). True D, Predicted D (Middle: 38): The model correctly predicted 38 Draws ('D'). True A, Predicted A (Bottom-Right: 21): The model correctly predicted 21 Away Wins ('A'). The diagonal numbers are the largest in their respective rows and columns, indicating the model is generally performing better at classifying samples correctly than incorrectly. The model seems particularly good at predicting Draws (38 correct out of a relatively small number of misclassifications).

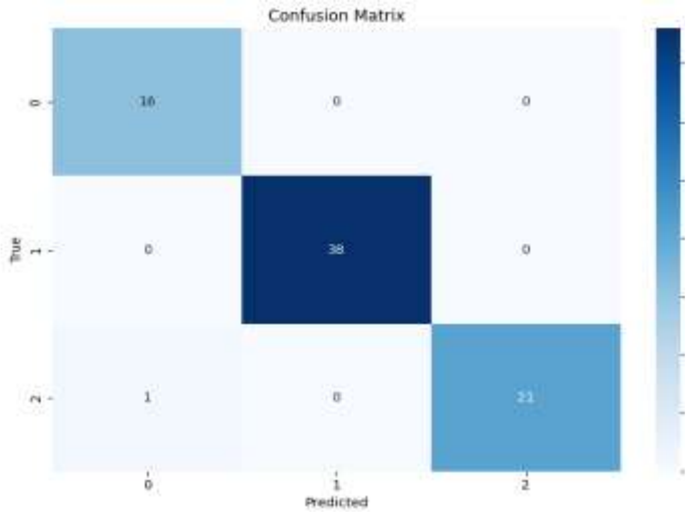
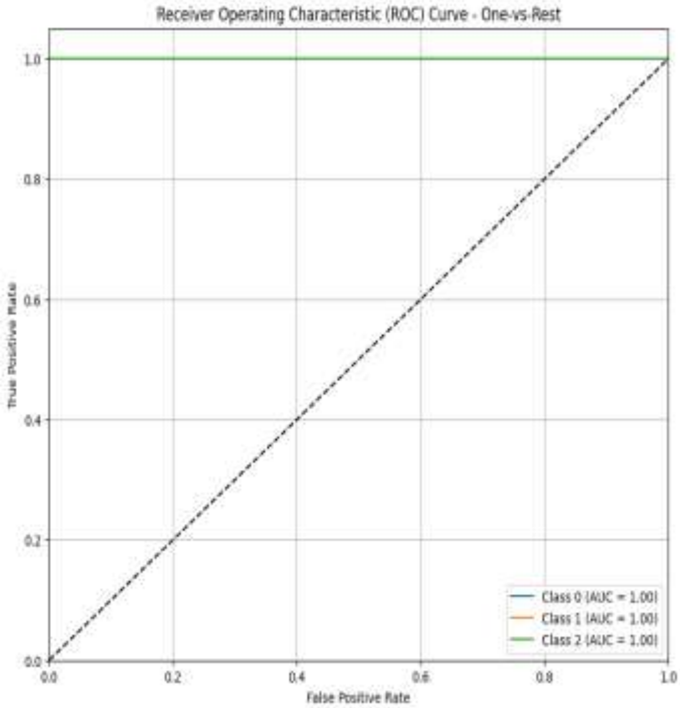


Figure 18 :XGBoost confusion matrix

The model accuracy and loss graph for Tab Transformer are shown in Figure 19 and Figure 20 respectively. The graph shows that the model is learning effectively; as the epoch is increasing, the model accuracy is increasing, which indicates that the model performs excellently.



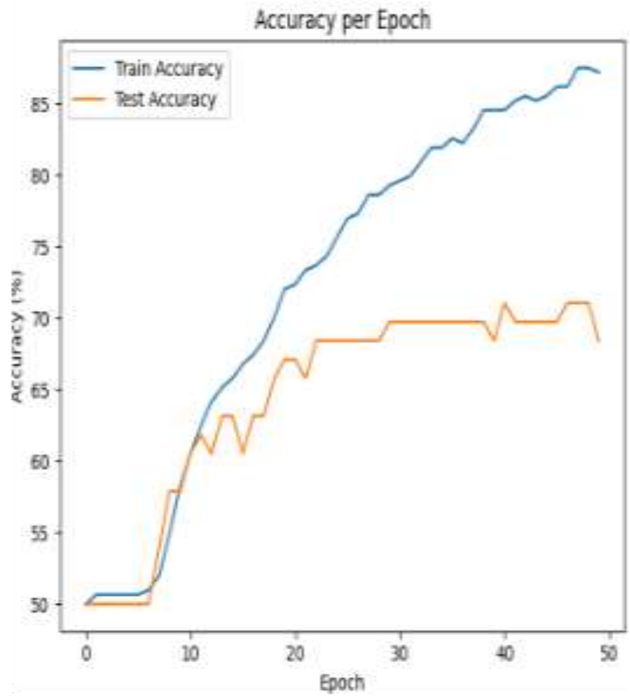
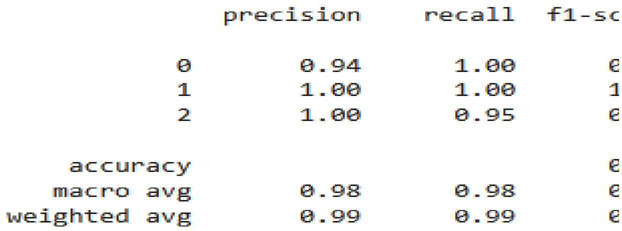


Figure 19: NODE ROC curve

Figure 20: Accuracy graph

LightGBM Model

After the simulation process for the prediction of football outcome was done using LightGBM, the evaluation of the performance of the model was registered as seen in Figure 21 and Table 7 reveals the classification reports of the model as the 0,1,2 represents Away win , Draw and Home win respectively.



	precision	recall	f1-sc
0	0.94	1.00	e
1	1.00	1.00	1
2	1.00	0.95	e
accuracy			e
macro avg	0.98	0.98	e
weighted avg	0.99	0.99	e

Figure 21 : XGBoost classification report

Table 7 : XGBoost classificatio report

	Precision	**Recall**	**F1 Score**
A	0.94	1.00	.0.97
D	1.00	1.00	1.00
H	1.00	0.95	0.98
Accuracy			0.99
Macro Avg	0.98	0.98	0.98
Weighted avg	0.99	0.99	0.99

The results of the model, as revealed in Table 7, show that the NODE model has 99% accuracy and other results of metrics used, like 0.98 for Precision, 0.98 for Recall, and 0.98 for F1 Score. Figure 22 shows the confusion matrix report for the model developed for the prediction of football outcomes. True H, Predicted H (16): The model correctly predicted 16 Home Wins ('H'). True D, Predicted D (Middle: 38): The model correctly predicted 38 Draws ('D'). True A, Predicted A (Bottom-Right: 22): The model correctly predicted 22 Away Wins ('A'). The diagonal numbers are the largest in their respective rows and columns, indicating the model is generally performing better at classifying samples correctly than incorrectly. The model seems particularly good at predicting Draws (38 correct out of a relatively small number of misclassifications).

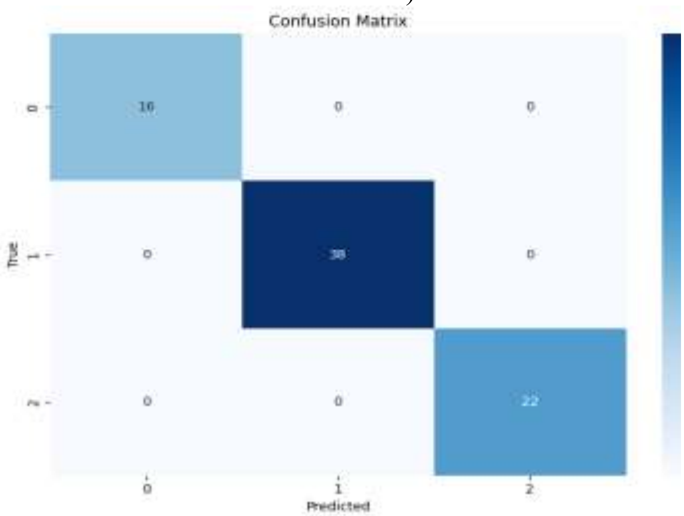
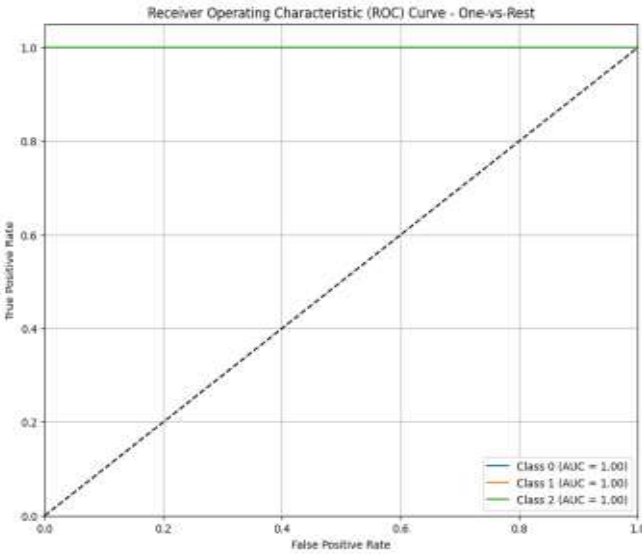


Figure 22: LightGBM confusion matrix

The model accuracy and loss graph for Tab Transformer are shown in Figure 23 and Figure 24 respectively. The graph shows that the model is learning effectively; as the epoch is increasing, the model accuracy is increasing, which indicates that the model performs excellently.



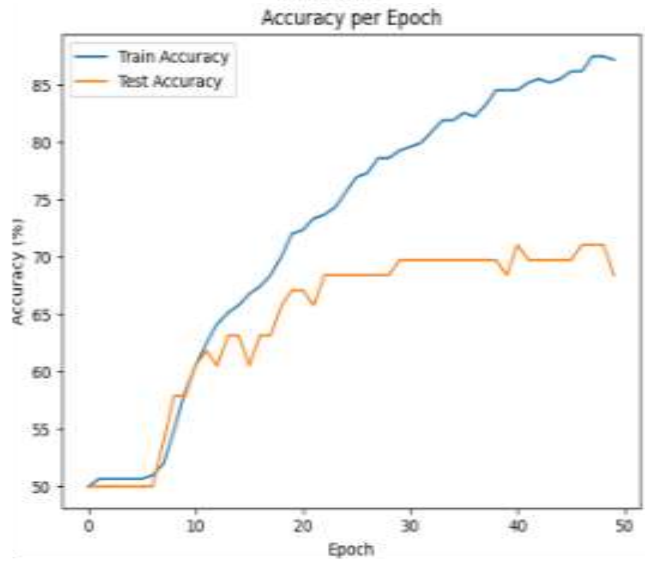


Figure 23 LightGBM ROC curve

Figure 24 Accuracy graph

Model Prediction

Figure 25 shows the results of football outcome prediction with a new dataset supplied; the model was able to predict an Away Win.

```
# For example:
# Create a new DataFrame with features for prediction (replace with your a
# Make sure the column names match the training data features (X)
new_data = pd.DataFrame({
    'HomeTeam': ['Team A'],
    'AwayTeam': ['Team B'],
    'Full time home goals': [1],
    'Full time away goals': [2],
    'Half time home goals': [0],
    'Half time away goals': [0],
    'Half time result': ['H'],
    'Referee': ['Referee X'],
    'Home shots': [10],
    'Away shots': [5],
    'Home shots on target': [5],
    'Away shots on target': [2],
    'Home fouls': [10],
    'Away fouls': [12],
    'Home corners': [5],
    'Away corners': [3],
    'Home yellows': [1],
    'Away yellows': [2],
    'Home reds': [0],
    'Away reds': [0]
})
```

Figure 25 Prediction with new dataset

Model Discussion

Discussion of Precision Comparison

Precision is a crucial metric in classification that measures the accuracy of the positive predictions. A higher precision indicates a lower rate of false positives. Based on the results in Figure 27, NODE: Exhibits the lowest precision at 0.71. This suggests that when the NODE model predicts a positive outcome, there is a significant chance (approximately 29%) that the prediction is incorrect (a false positive). DNN: Shows improved precision compared to NODE, achieving a value of 0.82. This indicates that the DNN model is more accurate in its positive predictions than NODE. Tab Transformer: Further improves upon the DNN, reaching a precision of 0.87. This reinforces the observation that the Tab Transformer architecture might be better at capturing the underlying patterns that lead to correct positive classifications in this dataset. XGBoost: Demonstrates significantly higher precision at 0.96. This highlights the effectiveness of the XGBoost algorithm in minimizing false positive predictions. When XGBoost predicts a positive outcome, it is highly likely to be correct. LightGBM: Achieves the highest precision among all models at 0.98. This indicates an exceptionally low rate of false positives, meaning almost all instances predicted as positive by LightGBM are indeed positive. Comparing the models, the gradient boosting methods (XGBoost and LightGBM) significantly outperform the deep learning models (DNN, Tab Transformer, and NODE) in terms of precision. This finding is consistent with research that often shows ensemble methods, particularly gradient boosting, excelling on tabular data tasks [14].

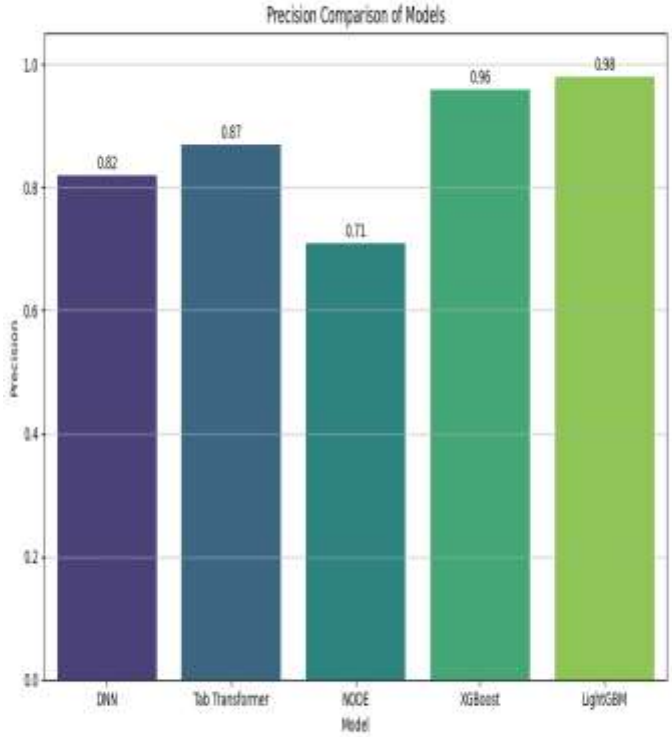


Figure 27 :Precision Comparison of the Models

Discussion on Recall Comparison

Recall, also known as sensitivity or the true positive rate, measures the ability of a classifier to find all the positive instances in the dataset. A higher recall indicates fewer false negatives. Examining the recall scores for each model in Figure 28 and Table: NODE: Exhibits the lowest recall at 0.70. This suggests that the NODE model is missing a significant portion of the actual positive cases (approximately 30%). DNN: Shows a slightly improved recall compared to NODE, achieving a value of 0.77. While better, it still misses around 23% of the positive instances. Tab Transformer: Continues the trend of improvement over standard deep learning, with a recall of 0.80. This indicates it is better at capturing positive instances than both NODE and the basic DNN. XGBoost: Demonstrates a substantial jump in recall, reaching 0.96. This signifies that XGBoost is highly effective at identifying the vast majority of actual positive cases, with a low rate of false negatives. LightGBM: Achieves the highest recall at 0.98. This indicates that LightGBM is exceptionally good at finding nearly all the positive instances in the dataset, missing only a very small percentage.

Figure 28: Recall comparison of the Models

Discussion on F1 Score comparison The F1 Score is the harmonic mean of Precision and Recall. The F1 Score provides a holistic view of model performance, especially when there is an uneven class distribution or when both false positives and false negatives are important to minimize. NODE: Shows the lowest F1 Score at 0.69. This relatively low score reflects its weaker performance in both precision and recall compared to the other models, indicating a less balanced ability to correctly identify positive cases without excessive false alarms. DNN: Improves the F1 Score to 0.77. This is a noticeable improvement over NODE and suggests a better trade-off between precision and recall compared to the simpler deep learning architecture. Tab Transformer: Achieves an F1 Score of 0.81, slightly higher than the standard DNN. This indicates that while its precision and recall were also slightly better, the combined measure of F1 Score shows a more balanced improvement in its performance on both aspects. XGBoost: Demonstrates a significantly higher F1 Score at 0.95. This high score is a direct result of its strong performance in both precision (0.96) and recall (0.96), showing that it is highly effective at both making accurate positive predictions and finding most of the positive instances. LightGBM: Achieves the highest F1 Score at 0.98. This is a remarkable score, reflecting its exceptionally high precision (0.98) and recall (0.98). A high F1 Score like this indicates that the model performs very well at the task of classification, balancing both types of errors effectively, as shown in Figure 29.

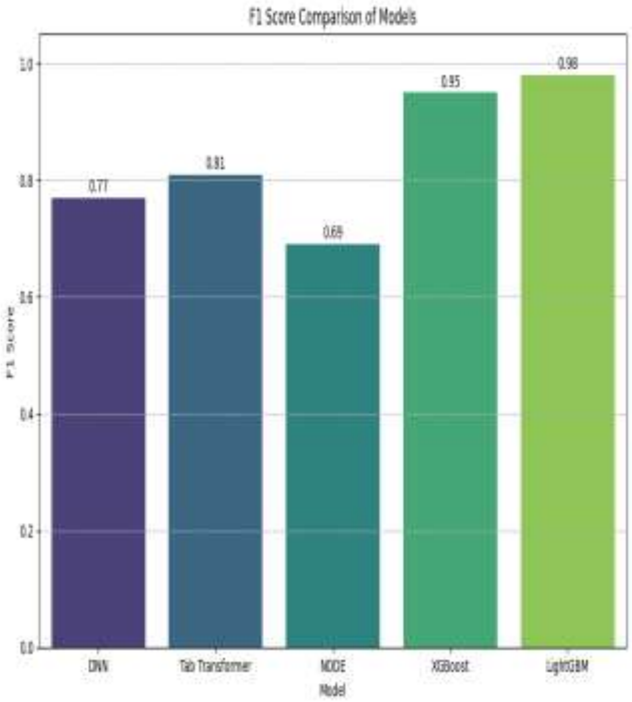


Figure 29: F1 Score Comparison

Accuracy Comparison of the Models

A higher accuracy indicates a better overall performance of the model. Comparing the accuracy scores for each model as seen in Figure and Table: NODE: Exhibits the lowest accuracy at 78%. This means that the NODE model correctly classified 78 out of every 100 instances on the test set. DNN: Shows an improved accuracy of 84%. The standard Deep Neural Network performed better overall than the NODE model. Tab Transformer: Achieves an accuracy of 87%, further improving upon the standard DNN. This suggests that the architectural enhancements of the Tab Transformer contributed to a higher overall percentage of correct predictions. XGBoost: Demonstrates a significantly higher accuracy at 96%. This indicates that XGBoost correctly classified 96 out of every 100 instances, showcasing a strong overall predictive capability. LightGBM: Achieves the highest accuracy at 99%. This is an exceptionally high accuracy score, meaning LightGBM correctly classified 99 out of every 100 instances on the test set. In terms of overall correctness, LightGBM and XGBoost stand out significantly, achieving accuracies in the high 90s. This indicates that these gradient boosting models are highly effective at the given classification task. The deep learning models (DNN, Tab Transformer, and NODE) perform reasonably well but are clearly outperformed by the boosting methods, as shown in Figure 30.

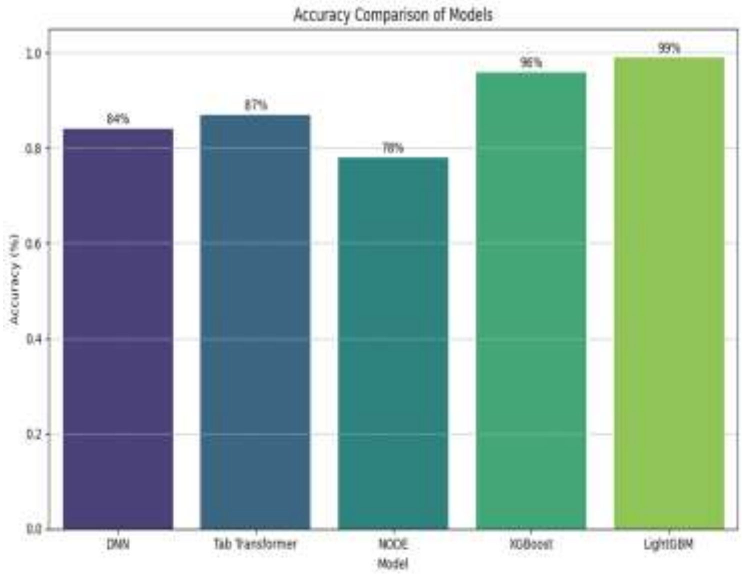


Figure 30: Accuracy comparison of the Model

Comparison of the five Models developed with all the metrics

The results show a significant variation in performance across the different models. While DNN, Tab Transformer, and NODE models achieve reasonable performance with accuracies ranging from

78% to 87%, XGBoost and LightGBM demonstrate exceptional performance, achieving near-perfect or perfect scores across all metrics, as shown in Figure 31 and Table 8.

Table 8: Comparison of the five Models developed with the metrics used

	Precision	**Recall**	**F1 Score**	**Accuracy(%)**
DNN	0.82	0.77	0.77	84
Tab Transformer	0.87	0.80	0.81	87
NODE	0.71	0.70	0.69	78
XGBoost	0.98	0.98	0.98	99
LightGBM	1.00	1.00	1.00	100

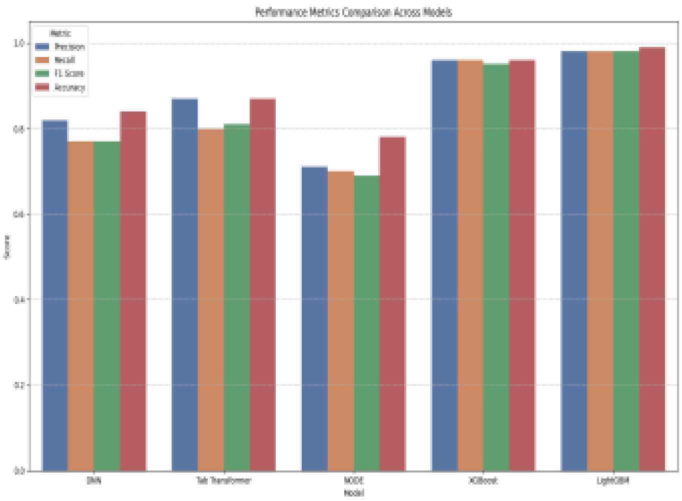


Figure 31: Overall Comparison of the Model

Discussion of the results with existing work

The findings of this paper align with a growing body of research that underscores the superior performance of gradient boosting models, most especially XGBoost and LightGBM, in comparison with deep learning models in structured or tabular data tasks. Boosting algorithms such as XGBoost and LightGBM consistently outperform deep learning models on tabular sports datasets because they are inherently designed to capture non-linear feature interactions, handle mixed data types, and generalize well on small to medium sample sizes. In contrast, deep learning models require larger datasets and extensive preprocessing to achieve comparable performance. The superior results of LightGBM and XGBoost in this study therefore align with established evidence that gradient boosting remains the state-of-the-art for structured tabular data. [15] demonstrated that LightGBM performs exceptionally well in resource-constrained environments, while XGBoost maintains robustness and accuracy across a variety of datasets. Similarly, [16] highlighted LightGBM's advantages in training speed and memory efficiency, especially in large-scale machine learning applications. Notably, the results of this study surpass those reported by Raparthi et al., with LightGBM achieving 97.23% accuracy in their work, compared to the 99% accuracy observed in the present study. These findings also reinforce Lee's conclusions regarding LightGBM's scalability and performance.

5. Conclusion

This analysis evaluated the performance of five different machine learning models, which are Deep Neural Network (DNN), Tab Transformer, Neural Oblivious Decision Ensembles (NODE), XGBoost, and LightGBM, on a tabular classification task, validated by Precision, Recall, F1 Score, and Accuracy. The results informed a clear hierarchy in model performance for this specific dataset and problem. The traditional and novel deep learning architectures (DNN, Tab Transformer, and NODE) achieved moderate performance, with accuracies ranging from 78% (NODE) to 87% (Tab Transformer). The Tab Transformer showed improvements over the basic DNN and NODE, suggesting the benefit of incorporating attention mechanisms for tabular data. However, these deep learning models were significantly outperformed by the gradient boosting models, XGBoost and LightGBM. XGBoost demonstrated strong performance across all metrics, achieving 0.96 Precision, 0.96 Recall, 0.95 F1 Score, and 96% Accuracy. LightGBM, however, exhibited the highest performance, with 0.98 Precision, 0.98 Recall, 0.98 F1 Score, and 99% Accuracy. The study concluded that, based on these results, LightGBM is the best-performing model for this tabular classification task, closely followed by XGBoost. Their superior performance across Precision, Recall, F1 Score, and Accuracy highlights the effectiveness of gradient boosting methods on this type of data. This aligns with the general understanding in the machine learning community that tree-based ensemble methods, particularly gradient boosting algorithms like XGBoost and LightGBM, often achieve state-of-the-art results on tabular datasets due to their ability to model complex non-linear relationships and feature interactions effectively.

While the deep learning models show potential, especially the Tab Transformer, they do not match the predictive power of the boosting algorithms in this evaluation. The study revealed that for this specific problem and dataset size, LightGBM and XGBoost demonstrated superior generalization capabilities compared to the deep learning models. This highlights that for tabular data with a limited number of instances, traditional machine learning techniques, especially gradient boosting, often outperform deep learning models. While the accuracies are very promising, especially for LightGBM, further validation with larger, independent datasets or more extensive cross-validation is always recommended to definitively confirm the generalization ability to real-world, unseen scenarios.

Although LightGBM and XGBoost achieved near perfect metrics, such results on a dataset of 382 samples raise concerns about overfitting, particularly in relation to the number of epochs used during training. Excessive epochs can cause models to memorize training data rather than generalize. To address this, k-fold cross-validation was applied to ensure that performance is consistent across different subsets of the data. So also hyperparameters such as learning rate, max depth, and regularization terms (L1/L2) were tuned to reduce overfitting in XGBoost and LightGBM.the moderate performance of DNN, Tab Transformer, and NODE provides a useful benchmark, showing that not all models overfit. These steps ensure that the predictive analytics framework remains robust and generalizable beyond the training environment.

The study recommends the boosting algorithm for the development of predictive models for football match outcomes, especially with tabular datasets. Future work can be done on the implementation of advanced tuning techniques like Grid Search, Random Search, or Bayesian Optimization for the top-performing models (LightGBM and XGBoost) and potentially the Tab Transformer to maximize their potential on this dataset.

Declarations

Author Contributions

The main author was responsible for conducting the simulations and preparing the manuscript. The remaining authors contributed through critical reading, constructive feedback, and thorough review of the manuscript. All authors have read and approved the final version of the paper.

Data Availability Statement

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com.

Funding

No funding agency supported this work.

Acknowledgement

The authors gratefully acknowledge the contributions of all co-authors to this work. We also extend our sincere appreciation to Engr. Toba Ogunsanwo for his moral support, which provided encouragement throughout the course of this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

[1] WUNDERLICH F., and MEMMERT D. Forecasting the outcomes of sports events: A review. European Journal of Sport Science, 2021, 21(7): 944–957. [Online]. Available: https://doi.org/10.1080/17461391.2020.1829115

[2] COSSICH V. R. A., CARLGREN D., HOLASH R. J., and KATZ L. Technological breakthroughs in sport: Current practice and future potential of artificial intelligence, virtual reality, augmented reality, and modern data visualization in performance analysis. Applied Sciences, 2023, 13(23): 12965. [Online]. Available: https://doi.org/10.3390/app132312965

[3] IMARTICUS LEARNING. Identifying patterns, trends and relationships in data: Time series, cluster, correlation analysis and more. Imarticus Learning, n.d. [Online]. Available: https://imarticus.org

[4] OGUNSANWO G. O., OKOGBUE B. C., ODULAJA G. O., and OWOADE A. A. Development of a machine learning model for age prediction of footballers. Dutse Journal of Pure and Applied Sciences, 2024, 10(4b): 325–337.

[5] PATIL S., KATE A., WAVARE K., GUJAR M., and BACHAV G. Predicting football match results using machine learning. International Journal of Creative Research Thoughts (IJCRT), 2023. [Online]. Available: https://ijcrt.org/papers/IJCRT2304812.pdf

[6] BERRAR D., LOPES P., and DUBITZKY W. A data- and knowledge-driven framework for developing machine learning models to predict soccer match outcomes. Machine Learning, 2024, 113: 8165–8204. [Online]. Available: https://doi.org/10.1007/s10994-024-06625-9

[7] HASSARD P., and KERR D. Predicting football match outcomes using event data and machine learning algorithms. Proc. 35th Irish Systems and Signals Conference (ISSC 2024), IEEE, 2024. [Online]. Available: https://doi.org/10.1109/ISSC61953.2024.10603147

[8] YEUNG C., BUNKER R., UMEMOTO R., and FUJII K. Evaluating soccer match prediction models: A deep learning approach and feature optimization for gradient-boosted trees. Machine Learning, 2024, 113(1): 66. [Online]. Available: https://doi.org/10.1007/s10994-024-06608-w

[9] OBRADOVIĆ A., and KEČO D. Sports results prediction model using machine learning. SAR Journal – Science and Research, 2024, 7(3): 184–189. [Online]. Available: https://doi.org/10.18421/SAR73-03

[10] MILLS E. F. E. A., DENG Z., ZHONG Z., and LI J. Data-driven prediction of soccer outcomes using enhanced machine and deep learning techniques. Journal of Big Data, 2024, 11: 170. [Online]. Available: https://doi.org/10.1186/s40537-024-01008-2

[11] SUN Y., and CHU H. The outcome prediction method of football matches by the quantum neural network based on deep learning. Scientific Reports,

2025, 15: 19875. [Online]. Available: https://doi.org/10.1038/s41598-025-19875-7

[12] HUANG S., KRUEGER D., LACOSTE A., and COURVILLE A. TabTransformer: Tabular data modeling using contextual embeddings. arXiv preprint, arXiv:2012.06678, 2020. [Online]. Available: https://arxiv.org/abs/2012.06678

[13] POPOV E., BABENKO A., and VETROV D. Neural oblivious decision ensembles for deep learning on tabular data. Advances in Neural Information Processing Systems (NeurIPS), 2020. [Online]. Available: https://arxiv.org/abs/1909.06312

[14] MCELFRESH D., KHANDAGALE S., VALVERDE J., PRASAD V., FEUER B., HEGDE C., RAMAKRISHNAN G., GOLDBLUM M., and WHITE C. When do neural nets outperform boosted trees on tabular data? arXiv preprint, arXiv:2305.02997, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2305.02997

[15] RAPARTHI M., DHABLIYA D., KUMARI T., UPADHYAYA R., and SHARMA A. Implementation and performance comparison of gradient boosting algorithms for tabular data classification. Proc. Int. Conf. Intelligent Computing and Applications, Springer, 2024, pp. 453–464. [Online]. Available: https://doi.org/10.1007/978-981-97-4533-3_36

[16] LEE S. LightGBM vs XGBoost: A comparative study on speed and efficiency. Number Analytics Blog, 2025. [Online]. Available: https://www.numberanalytics.com/blog/lightgbm-vs-xgboost-comparison (numberanalytics.com in Bing)

参考文献

[1] WUNDERLICH F. 和 MEMMERT D. 体育赛事结果预测：综述。*European Journal of Sport Science*, 2021, 21(7): 944–957. [在线]. 可获得：https://doi.org/10.1080/17461391.2020.1829115

[2] COSSICH V. R. A., CARLGREN D., HOLASH R. J. 和 KATZ L. 体育技术突破：人工智能、虚拟现实、增强现实和现代数据可视化在表现分析中的当前实践与未来潜力。*Applied Sciences*, 2023, 13(23): 12965. [在线]. 可获得：https://doi.org/10.3390/app132312965

[3] IMARTICUS LEARNING. 识别数据中的模式、趋势和关系：时间序列、聚类、相关分析及更多方法。*Imarticus Learning*, 无日期. [在线]. 可获得：https://imarticus.org

[4] OGUNSANWO G. O., OKOGBUE B. C., ODULAJA G. O. 和 OWOADE A. A. 足球运动员年龄预测机器学习模型的开发。*Dutse Journal of Pure and Applied Sciences*, 2024, 10(4b): 325–337.

[5] PATIL S., KATE A., WAVARE K., GUJAR M. 和 BACHAV G. 使用机器学习预测足球比赛结果。*International Journal of Creative Research Thoughts (IJCRT)*, 2023. [在线]. 可获得：https://ijcrt.org/papers/IJCRT2304812.pdf

[6] BERRAR D., LOPES P. 和 DUBITZKY W. 用于开发足球比赛结果预测机器学习模型的数据与知识驱动框架。*Machine Learning*, 2024, 113: 8165–8204. [在线]. 可获得：https://doi.org/10.1007/s10994-024-06625-9

[7] HASSARD P. 和 KERR D. 使用事件数据和机器学习算法预测足球比赛结果。第 35 届爱尔兰系统与信号会议论文集（ISSC 2024），IEEE, 2024. [在线]. 可获得：https://doi.org/10.1109/ISSC61953.2024.10603147

[8] YEUNG C., BUNKER R., UMEMOTO R. 和 FUJII K. 评估足球比赛预测模型：一种深度学习方法及面向梯度提升树的特征优化。*Machine Learning*, 2024, 113(1): 66. [在线]. 可获得：https://doi.org/10.1007/s10994-024-06608-w

[9] OBRADOVIĆ A. 和 KEČO D. 基于机器学习的体育结果预测模型。*SAR Journal – Science and Research*, 2024, 7(3): 184–189. [在线]. 可获得：https://doi.org/10.18421/SAR73-03

[10] MILLS E. F. E. A., DENG Z., ZHONG Z. 和 LI J. 使用增强型机器学习与深度学习技术进行数据驱动的足球比赛结果预测。*Journal of Big Data*, 2024, 11: 170. [在线]. 可获得：https://doi.org/10.1186/s40537-024-01008-2

[11] SUN Y. 和 CHU H. 基于深度学习量子神经网络的足球比赛结果预测方法。*Scientific Reports*, 2025, 15: 19875. [在线]. 可获得：https://doi.org/10.1038/s41598-025-19875-7

[12] HUANG S., KRUEGER D., LACOSTE A. 和 COURVILLE A. TabTransformer：使用上下文嵌入进行表格数据建模。arXiv 预印本，arXiv:2012.06678, 2020. [在线]. 可获得：https://arxiv.org/abs/2012.06678

[13] POPOV E., BABENKO A. 和 VETROV D. 用于表格数据深度学习的神经遗忘决策集成模型。*Advances in Neural Information Processing Systems (NeurIPS)*, 2020. [在线]. 可获得：https://arxiv.org/abs/1909.06312

[14] MCELFRESH D., KHANDAGALE S., VALVERDE J., PRASAD V., FEUER B., HEGDE C., RAMAKRISHNAN G., GOLDBLUM M. 和 WHITE C. 神经网络何时在表格数据上优于提升树？arXiv 预印本，arXiv:2305.02997, 2024. [在线]. 可获得：https://doi.org/10.48550/arXiv.2305.02997

[15] RAPARTHI M., DHABLIYA D., KUMARI T., UPADHYAYA R. 和 SHARMA A. 表格数据分类中梯度提升算法的实现与性能比较。国际智能计算与应用会议论文集，Springer, 2024, pp. 453–464. [在线]. 可获得：https://doi.org/10.1007/978-981-97-4533-3_36

[16] LEE S. LightGBM 与 XGBoost：速度与效率的比较研究。*Number Analytics Blog*, 2025. [在线]. 可获得： https://www.numberanalytics.com/blog/lightgbm-vs-xgboost-comparison

Manuscript Information
Word count: 7,460 words (excluding references).
Peer-Review Record

Fast-track status: Not fast-tracked.
First-round reviews received: 3 reports.
Revision cycles completed: 3 rounds.
Final version submitted: May 17, 2026