International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

# Intelligent Scheme for Footballer Performance Evaluation Using Deep-Learning Models

**Baydaa M. Merzah[1,2], Muayad S. Croock[3], Ahmed N. Rashid[4]**

*[1] Department of Computer Science, Collage of Computer Engineering and Information Technology, Anbar, Iraq*
*[2]Al-Nahrain University, Baghdad, Iraq*

*[3]Department of Control and Systems, University of Technology, Baghdad, Iraq*

*[4]Department of Computer Networks Systems, College of Computer Science and Information Technology, University of Anbar, Iraq*

*E-mail address: bai21c1007@uoanbar.edu.iq, muayad.s.croock@uotechnology.edu.iq, rashidisgr@uoanbar.edu.iq*

*Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##*

**Abstract:** The notable effectiveness of Deep Learning (DL) algorithms has led to a significant increase in their application across various academic domains and diverse sports fields. Football is renowned for the extensive data gathered for each player, team, match, and season. Consequently, football provides an ideal context for exploring various data analysis techniques to extract valuable insights. In this research, two datasets are employed to investigate the performance of football players at training and match sessions. The focus is on evaluating players' physical performance metrics during these sessions and providing suggestions for enhancing future training loads or decision-making by the coach during the match. Feedforward Neural Networks (FNN) are used to train the models with different architectures to the employed datasets. The performance of the models is optimal, as reflected by an accuracy of 100% for the match dataset and 99.29% for the training session data. The precision, recall, and F1-score are registered as 1.00 for the first dataset, while 0.9928, 0.9981, and 0.9954 for the second dataset. The test time, another factor used in assessing the applicability of the models for online applications, also shows promising results. Since the datasets are new, the results are validated using machine learning (ML) algorithms and 5-fold cross-validation. Our conclusive findings, obtained through the analysis of players' performance classification, underscore that the deep neural network models outperformed machine learning models in both time and accuracy.

**Keywords:** Deep Learning, football, player performance, feed-forward neural network

## 1. INTRODUCTION

DL, a subset of ML, seeks patterns within extensive datasets and can forecast forthcoming decisions, with applications across diverse scientific domains. In their research, [1] introduced an auto-stop car system driven by deep learning technology. For target tracking, another study [2] used the Long Short-Term Memory (LSTM) networks for predictive tracking. This system specifically incorporated two Convolutional Neural Networks (CNNs) for the tasks of face recognition and detecting travel drowsiness. In another study [3], a sophisticated multiclass classification model was developed and employed a deep convolutional neural network (CNN) to accurately classify license plates from three distinct countries. A Deep neural network was used in the Internet of Medical Things field to enhance security and cost with the aim of bolstering vehicle security and enhancing driver safety by developing two DL models [4].

In recent decades, the sports field has employed artificial intelligence in different applications. However, some challenges and limitations exist such as data availability, quality, and privacy; ethical and social implications; and the explainability and interpretability of the models. The application of DL extends to providing teams with enhanced insights for selecting players [5] and the best shooter [6], forecasting match outcomes [7], and identifying the winner [8], Additionally, it aids in devising player training strategies [9], predicting [10] and assessing players injuries [11], evaluating [12], and analyzing players performance [13]. Fig. 1 presents the major fields that have employed DL in football, as documented in various studies [14], [15] and [16].

Football player evaluation is a process of assessing the performance and potential of football players based on various criteria, such as technical skills [17], tactical knowledge [18], physical fitness and psychological attributes [12] as seen in Fig. 2. It is a vital step in the development of football players' processes and can improve the quality and competitiveness of football players, teams, and clubs [19]. It can also enhance the enjoyment and satisfaction of players and coaches. Accordingly, it is important to conduct football players' evaluations regularly and effectively.

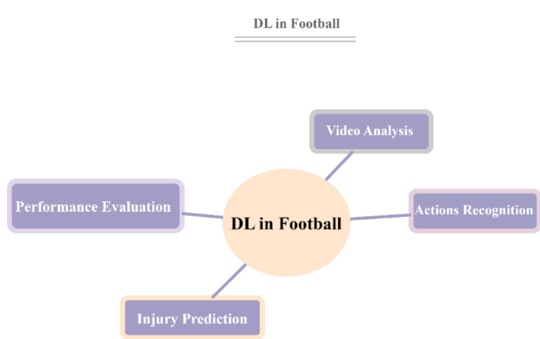


Figure 1. DL in football

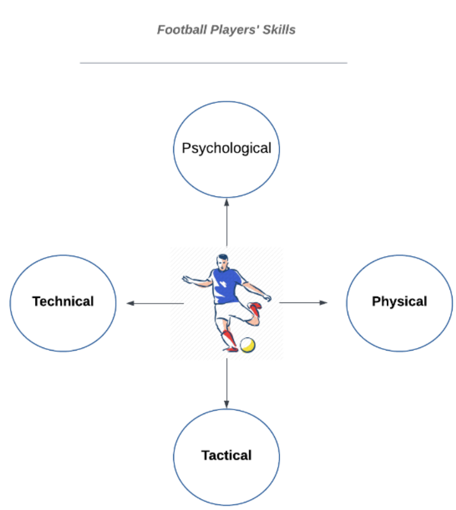


Figure 2. Football players' skills

Accordingly, utilizing DL in player performance assessments is an active and evolving research area. Evaluating football players is important for several reasons:

- It helps coaches and scouts select the most suitable players for their teams and identify the areas for improvement [20].
- It helps coaches and scouts select the most suitable players for their teams and identify the areas for improvement [21].
- It helps players understand their strengths and weaknesses and set realistic goals for their development [22].
- It helps clubs and organizations track players' progress, ensure they are on the appropriate level, and provide feedback and support [23].

This work aims to incorporate DL into performance assessments of football players. While previous research has predominantly focused on predicting game outcomes or leveraging DL alongside advanced technical tools for the collection and analysis of sports players' data, our approach centers on evaluating players' metrics during training sessions and matches. This involves tracking players' physical performance without requiring intricate technical tools. DL is employed to extract valuable information and assess the performance of various DL models using football players' datasets. The proposed evaluation models prove effective in performance classification utilizing accuracy, precision, recall, F1 score, and test time. The remainder of this paper is organized as follows. Section 2 is the literature review, Section 3 outlines the methodology utilized, and Section 4 presents the findings of the proposed work along with the discussion.

## 2. LITERATURE REVIEW

DL is a branch of artificial intelligence that uses multiple layers of neural networks to learn from large amounts of data. One of the emerging applications of DL is football, the most popular sport in the world. Football is a complex and dynamic game involving multiple factors, such as player skills, team tactics, game events, and environmental conditions. Analyzing and understanding football data can provide valuable insights for players, coaches, fans, and broadcasters. However, traditional methods of football analysis, such as manual annotation, rule-based systems, and classic ML, are limited in their abilities to handle the high-dimensional, noisy, and heterogeneous data associated with football. Hence, DL presents a promising alternative for tackling the challenges inherent in football analysis and the assessment of player performance. The scrutiny of player performance has emerged as a prominent subject, garnering increased attention from researchers in recent years.

In his work, [24] utilized data from 572 players input into ML and DL models to analyze their performance using historical data. In addition, he introduced a mathematical model leveraging game-related factors, including historical statistics from previous games and the playing position and skills of each player. The optimal model for each playing position group was identified. The model identified each player's weaknesses and strengths, aiming to enhance their future performance.

The objective of another study was to develop an ML model to recognize players' movements and examine specific training factors and the physical performance of the players [25]. The study involved two player training groups: control and experimental. All players underwent 24 training sessions, while the experimental group performed extra practical strength training after each session. Multiple data metrics were gathered from the experiment to construct a data vector serving as the input for a backpropagation neural network (BPNN). Kicking motions were analyzed using the BPNN, with the results demonstrating the efficiency of the proposed model in enhancing player performance. The primary limitation of this work was the small sample size of 116, which could introduce bias into the output of the trained model.

The work of [26] investigated the use of deep reinforcement learning to build an action-value Q-function from football game events. The goal impact metric was defined to quantify players' performance depending on the learned Q-function. A large dataset of action events from several football leagues was constructed to compute the action values for the players. The model was validated using a calibration experiment and a case study application from the English Football League Championship.

Artificial neural networks were employed by [27] to uncover the main performance metrics in professional football that affect outfield players. The study utilized technical data comprising 347 metrics from 966 players. Players were categorized into three main classes (Group 0, Group 1, and Group 2), representing where they played during the following season. The league status prediction accuracy ranged from 61.5% to 8.8%. This represents a significant development, as it demonstrates how ML can be integrated into the recruiting and scouting process in a professional football setting. The absence of physical features was the main drawback of this study.

demonstrated the significance of physiological characteristics as predictors of player performance was demonstrated in [28]. Three different tests were conducted on 16 players to collect the required measurements and metrics. Then, different ML regression models were utilized to forecast which features were significant for each physical performance. The primary limitations of this work were the models' poor prediction accuracy and the restricted sample size. These shortcomings suggest that alternative metrics for predicting players' performance should be explored.

In this study, [29] utilized statistical data to analyze the popularity of and predict the performance of National Basketball Association players. A deep neural network was employed to predict the performance of basketball players. In the data preprocessing phase, two datasets from different sources were merged for use in the proposed models. The validation of these models involved comparing the results with various ML algorithms used in their previous work. The results revealed that the DL model underperformed with the small dataset, highlighting the need for further investigation into additional features to improve prediction accuracy. A summary of the aforementioned papers is presented in Table 1.

Numerous discrepancies and limitations emerged from the data analysis:

1. Limited Dataset Samples: The work done by [24], [25], and [29] highlights the limitation posed by small sample sizes, potentially compromising the representativeness of the data. This highlights how important it is to have big and varied datasets to precisely record player traits and movement patterns across the spectrum of football.

2. Generalization Issue: Training DL models on limited samples derived from a single game [29] points out the challenge of applying it to diverse games.

3. Playing-Position Analysis: This was the main limitation mentioned by [28]. More considerations related to position-specific need to be taken into account to understand the key performance indicators.

4. Information Gap: The absence of athletes' physical capabilities noticed in [28] must be covered to enable thorough analysis and player performance prediction.

5. Practicality and Applicability: Due to particular test requirements and impracticality during training sessions or matches the work of [29] has difficulties in predicting physical performance.

Closing these gaps will contribute to more robust and applicable models for understanding and predicting football players' performance.

This work involves the construction of relatively large and novel datasets encompassing both training sessions and matches. These datasets incorporate physical performance metrics not included in previous studies to provide a comprehensive understanding of the actual performance of football players. In addition, the datasets include a gender metric, which enhances generalizability by facilitating performance evaluation across players on both men's and women's teams. Area is a key metric that aids in assessing performance in specific player positions, as variation in playing techniques exists among forwards, midfielders, and defenders.

Drawing on the information provided above and the literature review, the objectives and contributions of this study can be delineated as follows:

• To implement various DL models to identify the most effective model for the specified physical skills and to conduct a comparative evaluation using diverse metrics.

• To incorporate multiple performance skills of football players into the prediction process.
• To anticipate athlete performance of six distinct skills and to aid coaches in the judicious selection of athletes for games.
• To propose a DL-based model that considers players' performance.

TABLE I. SUMMARY OF THE PREVIOUS WORK

| Reference | Aim | Technique | Limitations |
|---|---|---|---|
| [24] | To anticipate the athletic performance of football players | Regression | Limited dataset samples |
| [25] | To discern the movement patterns of football players | Pattern recognition | Small dataset samples |
| [26] | To assess the comprehensive performance of the player | Regression | Useful during match session |
| [27] | To recognize crucial performance indicators in professional football that impact the league status of outfield players | Regression | lack of information regarding the physical features |
| [28] | To predict the physical performance | Regression | Required specific tests and requirements |
| [29] | To evaluate the effectiveness of applying ML and DL in the sports field | Regression & Classification | Relatively small-scale datasets |

## 3. METHODOLOGY

### 3.1 Datasets

The Accurate and objective information regarding the physical condition of football players is crucial for enabling coaches and players to make knowledgeable decisions regarding the formation of teams and training modalities [30]. The datasets utilized in this work are new and are from two distinct sources: match and training sessions. Each dataset is composed of 38,160 samples, with each record encapsulating specific features of individual players. These two datasets function as inputs for DL models in the classification process, enabling the assessment of the performance level to which a player belongs. Within the framework of our suggested methodology, the system produces distinct player performance classifications as its output. These serve as a reference for adjusting player substitutions or the overall playing strategy during a match and for optimizing training workloads. The datasets employed in this study underwent preprocessing using various techniques to render them suitable for utilization in the models. Fig. 3 illustrates the preprocessing procedures. Initially, area-based clustering was applied to label the dataset, utilizing K-nearest neighbors (KNN) with n = 3 to group players occupying similar areas (such as midfielders, defenders, and forwards). Each group was linked to one class (weak, normal, or active). These were then merged to form the complete labeled dataset. Subsequently, the process involved the identification and elimination of outliers to detect and remove extreme values. The match dataset saw a reduction in the number of samples to 36,421, whereas the training session dataset decreased to 38,145 samples. Next, standard scaling processes were employed to prepare the data for the DL algorithm. Following this, stratified sampling was implemented to guarantee the representation of each subgroup in the final sample, thereby improving the precision and reliability of the results when examining the population as a whole.

### 3.2 Match Session Dataset (MSDS)

The methodology used to create this dataset extended the work of [31] and increased the number of samples to 38,160. It includes five features: ID, gender, area, cross distance (CD), speed (Sp), and activity count (AC). The general statistics and the statistics for each playing position are presented in Table 2 and Table 3. The area represents the main playing positions in football games, with Area 1 representing forwards, Area 2 representing midfielders and area3 corresponding to defenders. Table II and Table III shows the statistics of the MSDS.

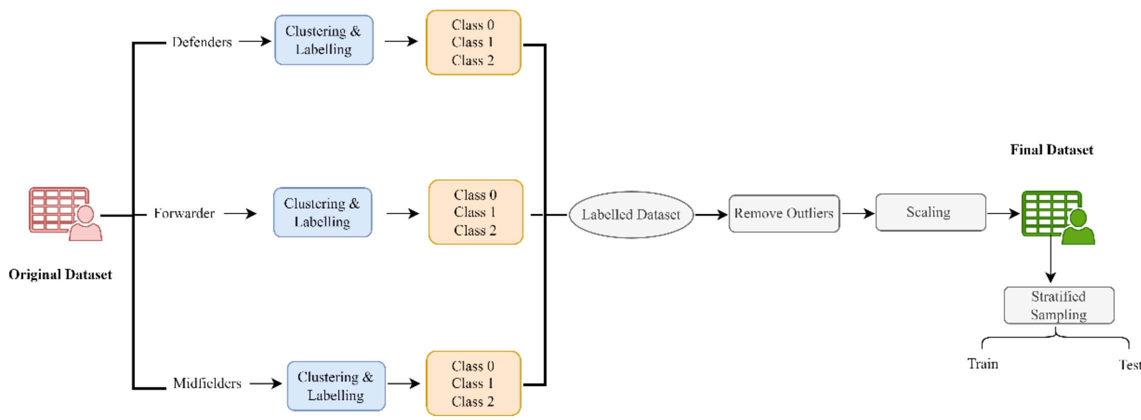


Figure 3. Dataset preprocessing

TABLE II. MSDS GENERAL STATISTICS

| | CD | Sp | AC |
|---|---|---|---|
| Min | 0 | 0 | 0 |
| Max | 1.175 | 7.238 | 58 |
| Mean | 0.542689 | 3.342234 | 17.98207 |
| STD | 0.210487 | 1.296324 | 12.32801 |

TABLE III. MSDS AREA-BASED STATISTICS

| | | | |
|---|---|---|---|
| **Area1** | | | |
| | ***CD*** | ***Sp*** | ***AC*** |
| Min | 0 | 0 | 0 |
| Max | 1.2 | 7.2 | 58 |
| Mean | 0.5 | 3.4 | 17.9 |
| STD | 0.2 | 1.3 | 12.3 |
| **Area2** | | | |
| | ***CD*** | ***Sp*** | ***AC*** |
| Min | 0.032 | 0.2 | 0 |
| Max | 1.147 | 7.064 | 58 |
| Mean | 0.539004 | 3.319536 | 18.15996 |
| STD | 0.206321 | 1.270654 | 12.39821 |
| **Area3** | | | |
| | ***CD*** | ***Sp*** | ***AC*** |
| Min | 0 | 0 | 0 |
| Max | 1.165 | 7.174 | 58 |
| Mean | 0.544036 | 3.350519 | 17.86312 |
| STD | 0.212965 | 1.311602 | 12.27198 |

### *3.3 Training Session Dataset (TSDS)*

This Given the significant privacy concerns associated with football players' data and the unavailability of the required data in existing literature, synthetic data were generated to simulate real values for the purpose of this work. This dataset consists of seven features in addition to the class. The features are ID, gender, area, heart rate (HR), oxygen consumption (O2), steps, and energy. This is an extended version of our previous work [32] with 38,160 samples. General statistics and area-based statistics are presented in Table IV and Table V, respectively.

TABLE IV. TSDS GENERAL STATISTICS

| | HR | O2 | Steps | Energy |
|---|---|---|---|---|
| Min | 73 | 80 | 2 | 45 |
| Max | 116 | 94 | 116 | 98 |
| Mean | 88.05385 | 85.29637 | 39.83015 | 79.0114 |
| STD | 8.904323 | 4.584851 | 22.67519 | 11.36339 |

TABLE V. TSDS AREA-BASED STATISTICS

| | | | | |
|---|---|---|---|---|
| **Area1** | | | | |
| | ***HR*** | ***O2*** | ***Steps*** | ***Energy*** |
| Min | 74 | 80 | 4 | 45 |
| Max | 115 | 94 | 112 | 98 |
| Mean | 88.62385 | 85.27864 | 41.7675 | 79.22896 |
| STD | 9.012097 | 4.595593 | 22.52717 | 11.27419 |
| **Area2** | | | | |

| | *HR* | *O2* | *Steps* | *Energy* |
|---|---|---|---|---|
| Min | 73 | 80 | 2 | 45 |
| Max | 112 | 94 | 102 | 98 |
| Mean | 87.20753 | 85.23566 | 36.54941 | 78.9941 |
| SD | 8.694285 | 4.571574 | 22.24257 | 11.35143 |
| **Area3** | | | | |
| | *HR* | *O2* | *Steps* | *Energy* |
| Min | 74 | 80 | 4 | 46 |
| Max | 116 | 94 | 116 | 98 |
| Mean | 88.61452 | 85.36589 | 42.13968 | 78.91994 |
| STD | 8.990138 | 4.59177 | 22.78071 | 11.41819 |

### *3.4 Performance evaluation models (PEM)*

in this work, FNNS [33] were used to evaluate players' physical performance. since the datasets used in this work are new and no similar studies are available, ml algorithms were used for validation. the general block diagram of the proposed model is shown in Fig. 4. in our previous work, state-of-the-art ml algorithms were used. in this study, the algorithms with the highest accuracy were selected for the validation procedure: decision tree (DT) [34] is selected for the MSDS dataset, and KNN [35] and gaussian naïve bayes (GNB) [36] were selected for the TSDS dataset. new deep neural networks were designed for each session. in the absence of standard criteria for selecting the appropriate architecture for these networks, we began with a simple architecture and refined it through trial and error until we reached the best-fit model for the corresponding dataset. different architectures and numerous hyperparameter tunings were tested. the configurations of the best-fit model are illustrated in Table VI and Table VII. the models have four layers. the first is the input layer, with six nodes corresponding to the number of features in the MSDS dataset, while there are seven nodes for the TSDS dataset. the second layer for the two models has 64 nodes, and the second hidden layer has 32 nodes with the rectified linear unit (Relu) activation function. the output layers consist of three nodes, one for each class, utilizing the Softmax activation function. this evaluation model was implemented using Keras (version 2.15.0) with a Tensorflow backend. the training process was repeated multiple times, using different numbers of epochs, batch sizes, and varied optimizers with a default learning rate of 0.01. the models were trained using google Colab in a CPU environment.

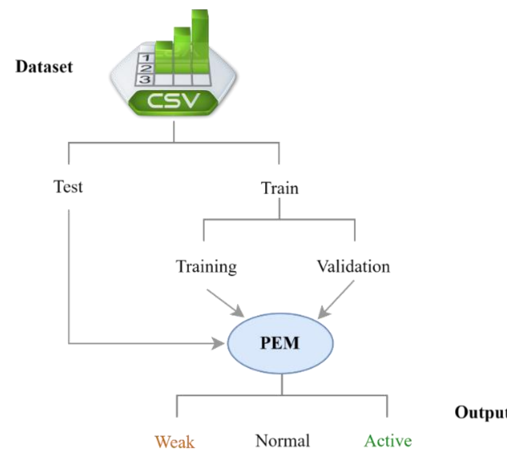


Figure 4. Block diagram of PEM

TABLE VI. FNN MODEL FOR MSDS

| **Model: "sequential" MSDS** | | | |
|---|---|---|---|
| ***Layer(type)*** | ***Output Shape*** | ***Activation Function*** | ***#Paramametres*** |
| ***dense (Dense)*** | (None, 64) | Relu | 488 |
| ***Dense_1 (Dense)*** | (None, 32) | Relu | 2080 |
| ***dense _2(Dense)*** | (None, 3) | Softmax | 99 |
| Total params: 2691 (10.26 KB)<br>Trainable params: 2691 (10.26 KB)<br>Non-trainable params: 0 (0.00 Byte) | | | |

TABLE VII. FNN MODEL FOR TSDS

| **Model: "sequential" TSDS** | | | |
|---|---|---|---|
| ***Layer(type)*** | ***Output Shape*** | ***Activation Function*** | ***#Paramametres*** |
| ***dense (Dense)*** | (None, 64) | Relu | 512 |
| ***Dense_1 (Dense)*** | (None, 32) | Relu | 2080 |
| ***dense _2(Dense)*** | (None, 3) | Softmax | 99 |
| Total params: 2691 (10.51 KB)<br>Trainable params: 2691 (10.51 KB)<br>Non-trainable params: 0 (0.00 Byte) | | | |

### *3.5 Performance evaluation models (PEM)*

The framework of the proposed model is illustrated in Fig. 5. Two modules are available in this scheme: training and match sessions. Each module has its specialized dataset. The next steps include training the DL model and selecting the best performance architecture.

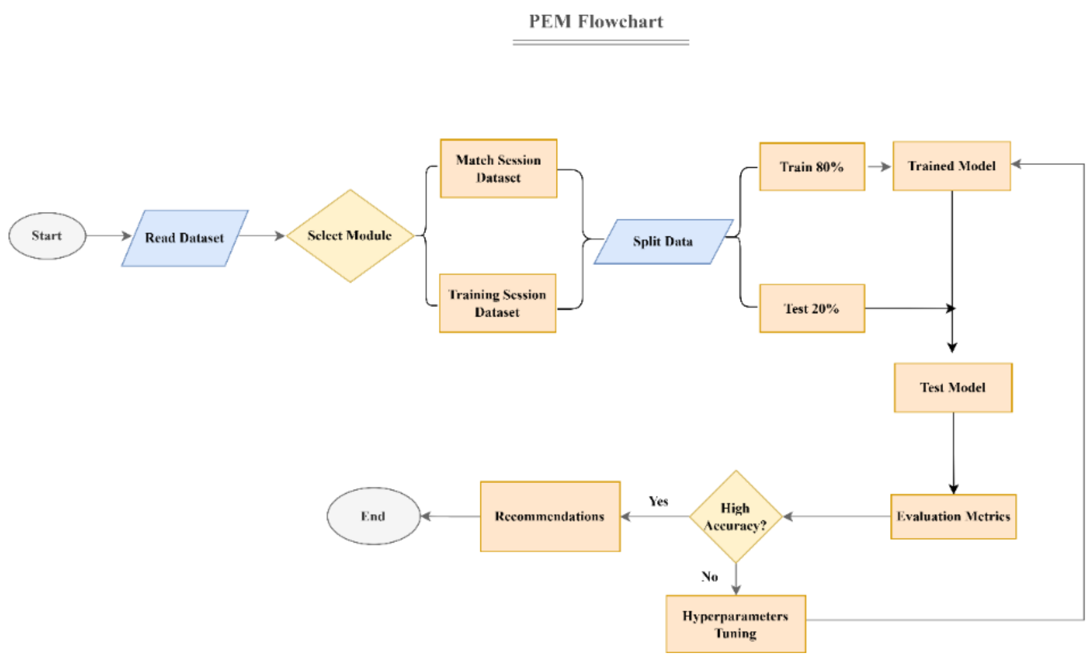


Figure 5. PEM flowchart

## 4. RESULTS AND DISCUSSION

### *4.1 statistical analysis*

The data presented in Table 8 examine the performance differentials among players during matches and training sessions across various positions—forwards, defenders, and midfielders—using multiple features, such as CD, HR, Sp, O2, AC, steps, and energy. The provided data include mean values and standard deviations for each feature within each position. For the MSDS dataset, CD is fairly consistent across all positions, with mean values hovering around 0.5. Sp displays variability but maintains comparable mean values across forwards, defenders, and midfielders. AC exhibits uniformity across positions, with an average of 18 ± 12.3. The large standard deviations, particularly in AC, suggest notable variability within each position.

For the TSDS dataset, the physiological responses and performance metrics are consistent across positions, suggesting a standardized training protocol. However, variations in the steps taken may warrant further investigation. The dataset provides a comprehensive overview of player responses during the training session, forming a basis for understanding positional differences in physiological demands and performance metrics.

TABLE VIII. STATISTICAL DIFFERENCES OF PLAYERS' PERFORMANCE IN TRAINING AND MATCH SESSION

| | Match Session Dataset | | | | Training Session Dataset | | | |
|---|---|---|---|---|---|---|---|---|
| ***Position*** | ***Feature*** | ***Min ± SD*** | ***Max ± SD*** | ***Mean ± SD*** | ***Feature*** | ***Min ± SD*** | ***Max ± SD*** | ***Mean ± SD*** |
| ***Forwarder*** | ***CD*** | 0 ± 0.2 | 1.18 ± 0.2 | 0.54 ± 0.2 | ***HR*** | 74 ± 9 | 115 ± 9 | 89 ± 9 |
| | ***Sp*** | 0 ± 0.3 | 7.2 ± 0.3 | 3.3 ± 0.3 | ***O2*** | 80 ± 5 | 94 ± 5 | 85 ± 5 |
| | ***AC*** | 0 ± 12.3 | 58 ± 12.3 | 18 ± 12.3 | ***Steps*** | 4 ± 23 | 112 ± 23 | 42 ± 23 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | ***Energy*** | 45 ± 11 | 98 ± 11 | 79 ± 11 |
| ***Defenders*** | ***CD*** | 0.03 ± 0.2 | 1.15 ± 0.2 | 0.5 ± 0.2 | ***HR*** | 73 ± 9 | 112 ± 9 | 87 ±9 |
| | ***Sp*** | 0.2 ± 1.3 | 7.1 ± 1.3 | 3.3 ± 1.3 | ***O2*** | 80 ± 5 | 94 ± 5 | 85 ± 5 |
| | ***AC*** | 0 ± 12.3 | 58 ± 12.3 | 18 ± 12.3 | ***Steps*** | 2 ± 22 | 102 ± 22 | 37 ± 22 |
| | | | | | ***Energy*** | 45 ± 11 | 98 ± 11 | 79 ± 11 |
| ***Midfielders*** | ***CD*** | 0 ± 0.2 | 1.7 ± 0.2 | 0.5 ± 0.2 | ***HR*** | 74 ± 9 | 116 ± 9 | 87 ± 9 |
| | ***Sp*** | 0 ± 1.3 | 7.8 ± 1.3 | 3.4 ± 1.3 | ***O2*** | 80 ± 5 | 94 ± 5 | 85 ± 5 |
| | ***AC*** | 0 ± 12.3 | 58 ± 12.3 | 18 ± 12.3 | ***Steps*** | 4 ± 23 | 116 ± 23 | 42 ± 23 |
| | | | | | ***Energy*** | 46 ± 11 | 98 ± 11 | 79 ± 11 |

### *4.2 DL results*

The In addition to conventional ML methods, FNNs were applied to predict the performance of football players. FNNs are distinct from traditional ML models like DT and KNN; the challenge lies in identifying the optimal FNN architecture and finely tuning the hyperparameter set, given the vast number of potential FNN configurations. In this work, two FNN models for classification were designed: the first for the MSDS dataset and the second for the TSDS dataset. Numerous factors were tested to achieve more accurate models, including experimenting with different epochs with varying batch sizes, utilizing the Google Colab environment, and exploring various optimizer algorithms. The learning rate remained unchanged at 0.01. The factors, along with their corresponding test results, are shown in Table IX, Table X, Fig. 6 and Fig. 7. The configurations for the models are explained in the previous section. The classifiers were evaluated using metrics including accuracy, precision, recall, and F1 score [37].

The results of the MSDS dataset are presented in Table IX and Fig. 6. They indicate high performance and a well-trained classifier. Various configurations were tested in terms of epochs and batch sizes. The trade-off between computational resources and model performance was taken into account during the training phase. Table X and Fig. 7 demonstrate the high performance of the TSDS model across all metrics and various configurations. The overall performance of the models' during training was excellent, and they were subsequently evaluated on a test set, as depicted in Table X.

TABLE IX. ACCURACY MATRIX OF THE MSDS MODEL

| #Epochs | Batch-size | Class | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|---|---|
| 50 | 30 | Class 0 | 1.000 | 1.00 | 1.00 | **1.00%** |
| | | Class 1 | 1.000 | 1.00 | 1.00 | |
| | | Class 2 | 1.000 | 1.00 | 1.00 | |
| 15 | 10 | Class 0 | 1.000 | 1.000 | 1.000 | 0.999% |
| | | Class 1 | 1.000 | 1.000 | 1.000 | |
| | | Class 2 | 1.000 | 1.000 | 1.000 | |
| 10 | 5 | Class 0 | 1.000 | 0.997 | 0.999 | 0.997% |
| | | Class 1 | 0.999 | 0.996 | 0.997 | |
| | | Class 2 | 0.989 | 0.998 | 0.994 | |

TABLE X. ACCURACY MATRIX OF THE TSDS MODEL

| # Epochs | Batch size | Class | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|---|---|
| 50 | 50 | C 0 | 0.9928 | 0.9981 | 0.9954 | 99.66% |
| | | C 1 | 0.9986 | 0.9961 | 0.9973 | |
| | | C 2 | 0.9986 | 0.9954 | 0.9970 | |
| 50 | 20 | C 0 | 0.9977 | 0.9889 | 0.9933 | 99.46% |
| | | C 1 | 0.9965 | 0.9958 | 0.9961 | |
| | | C 2 | 0.9883 | 0.9995 | 0.9939 | |
| | | C 0 | 0.9709 | 0.9992 | 0.9849 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 30 | 32 | C 1 | 0.9989 | 0.9855 | 0.9922 | 98.92% |
| | | Cs 2 | 0.9991 | 0.9818 | 0.9903 | |

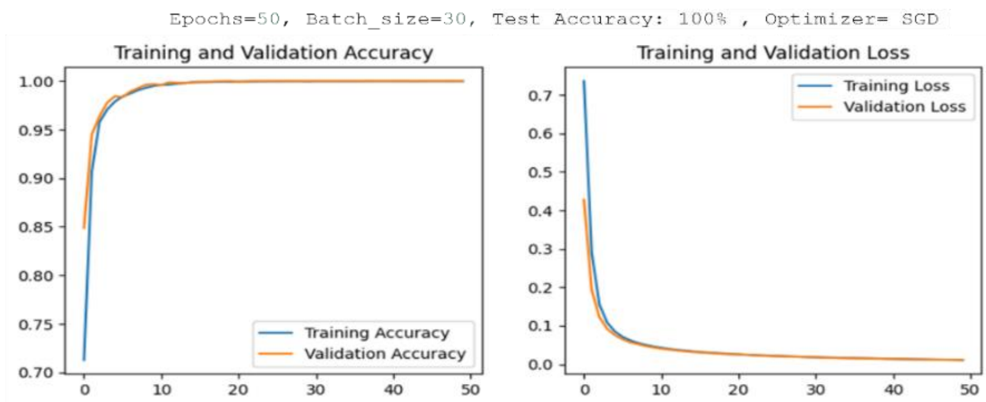


Figure 6. TSDS learning curves

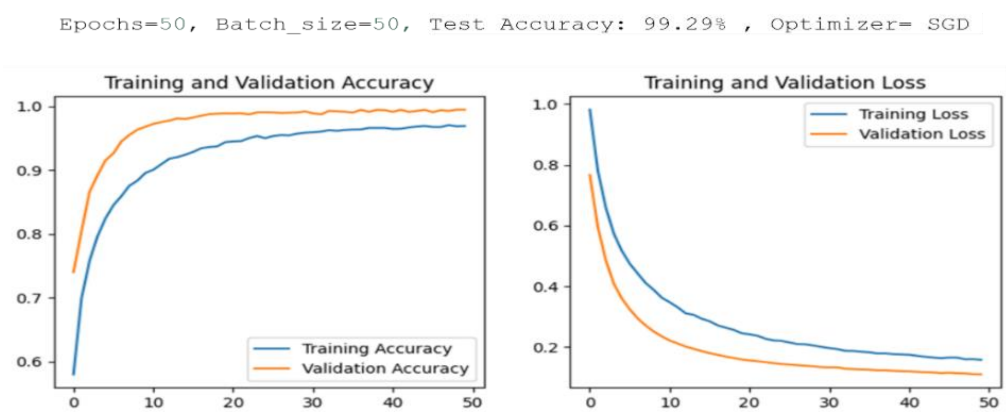


Figure 7. TSDS learning curves

By analyzing the data presented in Table XI reveals several key observations; the MSDS dataset exhibits a comparatively easier learning curve compared to the TSDS dataset, achieving higher accuracy with fewer epochs and smaller batch sizes. The Adam optimizer demonstrated superior performance compared to the stochastic gradient descent (SGD) optimizer; this was evident in its ability to attain 100% accuracy on the MSDS dataset across varying batch sizes and epochs. In contrast, the SGD optimizer reached 100% accuracy only with 50 epochs and 30 batch sizes. It is noteworthy, however, that the SGD optimizer outperformed the Adam optimizer in terms of learning curve smoothness, which can be attributed to differences in learning rate strategies employed by the two optimizers. Furthermore, the GPU environment did not appear to confer a significant advantage over the CPU environment, as the accuracy levels are comparable or slightly lower on the GPU than on the CPU. This suggests that the dataset may not be sufficiently large or complex to harness the parallel processing capabilities of the GPU effectively. Additionally, the smoothness of the learning curve appears to be contingent on both the batch size and the dataset. Larger batch sizes tend to yield smoother learning curves, contributing to a reduction in the variance of gradient updates.

### *4.3 Model validation*

Since the used datasets are new and the new approach used for classifying the performance level for the players; the evaluation was performed using five metrics: accuracy, precision, recall, F1 score, and test time. These constitute the primary metrics for assessing model performance in classification problems. These metrics are defined by [37], wherein TP denotes true positives, FP signifies false positives, TN represents true negatives, and FN stands for false negatives.

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

TABLE XI. RESULTS OF DIFFERENT HYPERPARAMETERS TUNNING

| # Epochs | Batch- size | Dataset | Environment | Optimizer | Accuracy |
|---|---|---|---|---|---|
| 15 | 10 | MSDS | GPU | Adam | 100% |
| 10 | 10 | MSDS | CPU | Adam | 99.97% |
| 10 | 5 | MSDS | GPU | Adam | 100% |
| 10 | 30 | MSDS | GPU | Adam | 99.89% |
| 10 | 40 | MSDS | CPU | Adam | 99.99% |
| 50 | 40 | MSDS | CPU | Adam | 100% |
| 50 | 30 | MSDS | CPU | SGD | 100% |
| 50 | 10 | TSDS | CPU | Adam | 99.27% |
| 40 | 32 | TSTD | CPU | Adam | 99.23% |
| 20 | 32 | TSDS | CPU | Adam | 99.24% |
| 10 | 32 | TSDS | CPU | Adam | 99.08% |
| 50 | 50 | TSDS | CPU | SGD | 99.29 |

The time denotes the total number of seconds required to evaluate the model's performance on the test set. In the preceding section, we applied PEM models to assess the performance of football players using both the MSDS and TSDS datasets. While the FNN used in this work demonstrated high precision in the training and validation phases, cross-validation can provide a more robust estimation of model performance. For instance, one cross-validation method involves splitting the dataset into five parts (fivefold). The model is trained on four parts, with the remaining part used for validation. This procedure is iterated five times, utilizing each stage for both training and validation purposes. The outcomes are subsequently averaged to yield a precise assessment of model performance. Cross-validation serves the additional purpose of gauging whether the model exhibits overfitting to the training data, a critical consideration when employing FNN models. Employing this methodology enhances our comprehension of the model's efficacy in classifying players' skill levels and confidence.

In addition, three ML algorithms were employed to verify the results as part of the validation procedures. The outcomes of the work by [32] were used to choose the ML models. Table XII presents the accuracy and computation time of applying the adopted models to the datasets. Notably, the DL model attains exceptional accuracy within an exceedingly low computation time, demonstrating its worth in real-time decision-making scenarios. For the TSDS, the KNN and GNB models slightly lag with the accuracy as compared to the DL model's impressive accuracy and computation time. The robust classification abilities of DL are highlighted by the results establishing it as a viable model for time-sensitive applications. The trade-off between computational time and accuracy is required when choosing a model for a particular task.

Table XIII presents the standard evaluation metrics (accuracy, precision, recall, and F1 score) for each fold. The precision of the model was assessed, while the recall measured how accurately the TP was predicted. The balance between precision and recall was quantified by the F1 score.

TABLE XII. DL AND ML RESULTS

| | Model | Accuracy | Time |
|---|---|---|---|
| **Match session dataset** | DT | 100 % | 0.001578 s |
| | DL | 100 % | 0.00012 s |
| **Training session dataset** | GNB | 70 % | 0.002865 s |
| | KNN | 68 % | 0.412976 s |
| | DL | 99.29 % | 0.00023 s |

Given the importance of both precision and recall in the measurement of accuracy, the F1 score was utilized for model evaluation. Fold 5 demonstrated the highest F1 score and accuracy, reaching 99.86%. The model displayed some confusion between Class 2 and Class 0, as the classes share predominantly similar values in terms of position and playing area. Fig. 8 and Fig. 9 display the evaluation metrics from the fivefold validation and provide a clear visualization of the model's performance through the confusion matrix.

*E-mail:author's email*

TABLE XIII. MSDS 5-FOLD CROSS-VALIDATION RESULTS

| MSDS | | | | | |
|---|---|---|---|---|---|
| | ***Fold1*** | ***Fold2*** | ***Fold3*** | ***Fold4*** | ***Fold5*** |
| **Accuracy** | 99.6568 | 99.79406 | 99.656 | 99.93135 | 99.86271 |
| **Precision** | 99.6595 | 99.79458 | 99.659 | 99.93148 | 99.86321 |
| **Recall** | 99.6568 | 99.79406 | 99.656 | 99.93135 | 99.86271 |
| **F1-score** | 99.6570 | 99.79408 | 99.657 | 99.93135 | 99.86271 |

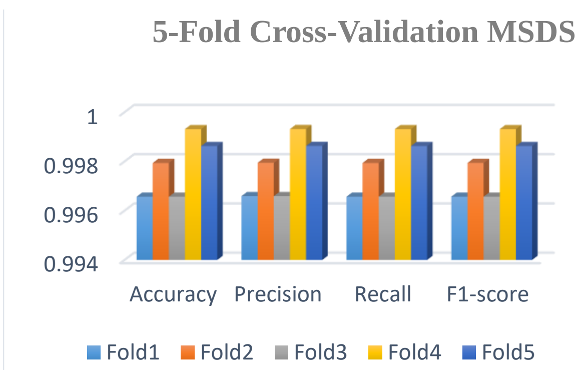


Figure 8. MSDS performance results

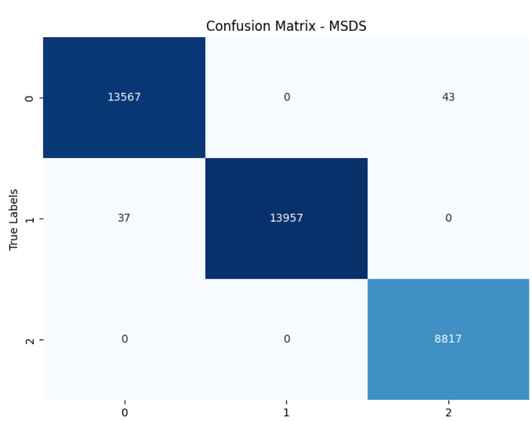


Figure 9. MSDS confusion matrix

The results presented in Table XIV, Fig. 10, and Fig. 11 consistently depict the superior performance of the TSDS model. The accuracy across all clusters varies from 98.93% to 99.46%, indicating that the model can make accurate predictions in all tests. Although the F1 score values showed a stable, minor increase, the F1 score reflects the balanced nature of combining accuracy and recall. It implies an equilibrium in the models' ability to effectively prevent false positives and capture true positives. In this context, the TSDS model performance seems strong across the majority of types, high in terms of accuracy, and balanced in terms of precision and recall. However, a close examination of the confusion matrix can provide even more specific insights into the model's behavior. High and relatively stable performance across the vast majority of classes conclude that the model is obviously suitable for use in the specified datasets and classification tasks.

TABLE XIV. TABLE 14. TSDS 5-FOLD CROSS-VALIDATION RESULTS

| TSDS | | | | | |
|---|---|---|---|---|---|
| | ***Fold1*** | ***Fold2*** | ***Fold3*** | ***Fold4*** | ***Fold5*** |
| **Accuracy** | 98.925 | 99.3183 | 99.318 | 99.449 | 99.4625 |
| **Precision** | 98.933 | 99.318 | 99.324 | 99.450 | 99.46354 |
| **Recall** | 98.925 | 99.3183 | 99.318 | 99.449 | 99.4625 |
| **F1-score** | 98.925 | 99.317 | 99.317 | 99.448 | 99.4622 |

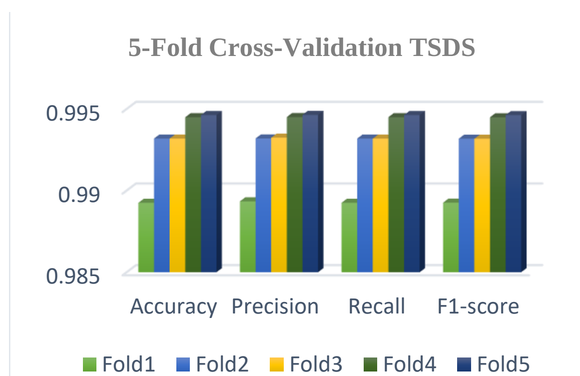


Figure 10. TSDS performance results

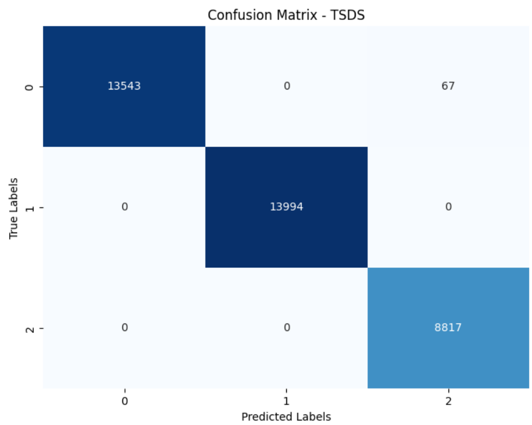


Figure 11. TSDS confusion matrix

## 5. CONCLUSION

In this article, we introduced the PEM method, a DL-based approach designed for classifying the performance of football players during both training sessions and matches. Our findings reveal that DL-based models outperform traditional ML models in the performance classification task. The proposed models exhibited an accuracy exceeding 99% in the performance classification process. Additionally, our models underscore the significance of distinct physical features for players during training sessions and matches. Furthermore, the work illustrated that players occupying various field positions exhibit disparate performance levels and engage in distinct activities. We also demonstrate the pivotal role of test time in DL models, particularly concerning their viability for online applications. We expect our overall findings to provide valuable support for coaching staff and team management in enhancing player performance in matches and training, thereby guiding them to concentrate on key areas for improvement. Our future endeavors involve the integration of these models into real-time matches. Recognizing and addressing challenges related to real-time data processing and model deployment is imperative for laying the groundwork for forthcoming research initiatives. Moreover, a more intricate examination of position-specific performance patterns could reveal potential differences in playing styles or strategies among them.

## REFERENCES


[1] H. A. Ahmed, M. a. N. Al-Hayanni, and M. S. Croock, "Intelligent and secure real-time auto-stop car system using deep-learning models," International Journal of Electrical and Computer Engineering Systems, vol. 15, no. 1, pp. 31–39, Jan. 2024. doi: 10.32985/ijeces.15.1.4.

[2] S. H. Ahmed and A. N. Rashid, "Prediction of single object tracking based on learning approach in wireless sensor networks," 2021 14th International Conference on Developments in eSystems Engineering (DeSE), Dec. 2021, doi: 10.1109/dese54285.2021.9719540.

[3] M. A. Uthaib and M. S. Croock, "Multiclassification of license plate based on deep convolution neural networks," International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering, vol. 11, no. 6, p. 5266, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5266-5276.

[4] A. Lakhan et al., "Restricted Boltzmann machine assisted secure serverless edge system for internet of medical things," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 673–683, Feb. 2023, doi: 10.1109/jbhi.2022.3178660.

[5] K S. Buyrukoğlu and S. Savaş, "Stacked-Based Ensemble Machine Learning model for positioning footballer," Arabian Journal for Science and Engineering, vol. 48, no. 2, pp. 1371–1383, Apr. 2022, doi: 10.1007/s13369-022-06857-8.

[6] L. Javadpour, J. Blakeslee, M. Khazaeli, and P. Schroeder, "Optimizing the best play in basketball using deep learning," Journal of Sports Analytics, vol. 8, no. 1, pp. 1–7, Mar. 2022, doi: 10.3233/jsa-200524.

[7] T. Satyapanich and A. Somkheawwan, Predicting game results for football league using deep learning. 2023. doi: 10.1109/icsec59635.2023.10329770.

[8] J. AlMulla, M. T. Islam, H. R. H. Al-Absi, and T. Alam, "SoccerNet: A Gated Recurrent Unit-based model to predict soccer match winners," PloS One, vol. 18, no. 8, p. e0288933, Aug. 2023, doi: 10.1371/journal.pone.0288933.

[9] M. Stoeve, D. Schuldhaus, A. Gamp, C. Zwick, and B. M. Eskofier, "From the laboratory to the field: IMU-Based shot and pass detection in football training and game scenarios using deep learning," Sensors, vol. 21, no. 9, p. 3071, Apr. 2021, doi: 10.3390/s21093071.

[10] Dunne, M. (2021). A comparative study on deep & machine learning techniques used for football injury prediction & prevention. Ph.D. dissertation, National College of Ireland, Dublin.

[11] H. Song, N. X.-Y. Han, C. E. Montenegro-Marin, and S. Krishnamoorthy, "Secure prediction and assessment of sports injuries using deep learning based convolutional neural network," Journal of Ambient Intelligence & Humanized Computing/Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 3, pp. 3399–3410, Mar. 2021, doi: 10.1007/s12652-020-02560-4.

[12] A. Ati, P. Bouchet, and R. B. Jeddou, "Using multi-criteria decision-making and machine learning for football player selection and performance prediction: A systematic review," Data Science and Management, vol. 7, no. 2, pp. 79–88, Jun. 2024, doi: 10.1016/j.dsm.2023.11.001.

[13] C. Mou, "The attention mechanism performance analysis for football players using the internet of things and deep learning," IEEE Access, p. 1, Jan. 2024, doi: 10.1109/access.2024.3350036.

[14] S. Akan and S. Varlı, "Use of deep learning in soccer videos analysis: survey," Multimedia Systems, vol. 29, no. 3, pp. 897–915, Dec. 2022, doi: 10.1007/s00530-022-01027-0.

[15] I. S. Gillani et al., "Yolov5, Yolo-x, Yolo-r, Yolov7 Performance Comparison: A Survey," Artificial Intelligence and Fuzzy Logic System, Sep. 2022, doi: 10.5121/csit.2022.121602.

[16] X. Li and R. Ullah, "An image classification algorithm for football players' activities using deep neural network," Soft Computing, vol. 27, no. 24, pp. 19317–19337, Oct. 2023, doi: 10.1007/s00500-023-09321-3.

[17] P. Chmura et al., "Is there meaningful influence from situational and environmental factors on the physical and technical activity of elite football players? Evidence from the data of 5 consecutive seasons of the German Bundesliga," PloS One, vol. 16, no. 3, p. e0247771, Mar. 2021, doi: 10.1371/journal.pone.0247771.

[18] D. Carnevale et al., “Executive Functions, Physical Abilities, and Their Relationship with Tactical Performance in Young Soccer Players,” Perceptual and Motor Skills, vol. 129, no. 5, pp. 1477–1491, Jul. 2022, doi: 10.1177/00315125221112236.

[19] S. González-Víllora, J. Serra-Olivares, J. C. Pastor-Vicedo, and I. T. Da Costa, “Review of the tactical evaluation tools for youth players, assessing the tactics in team sports: football,” SpringerPlus, vol. 4, no. 1, Nov. 2015, doi: 10.1186/s40064-015-1462-0.

[20] Q. Yi et al., “Evaluation of the technical performance of football players in the UEFA Champions League,” International Journal of Environmental Research and Public Health/International Journal of Environmental Research and Public Health, vol. 17, no. 2, p. 604, Jan. 2020, doi: 10.3390/ijerph17020604.

[21] E. Wakelam, V. Steuber, and J. Wakelam, “The collection, analysis and exploitation of footballer attributes: A systematic review,” Journal of Sports Analytics, vol. 8, no. 1, pp. 31–67, Mar. 2022, doi: 10.3233/jsa-200554.

[22] M.-Á. Gómez, C. Lago, M.-T. Gómez, and P. Furley, “Analysis of elite soccer players’ performance before and after signing a new contract,” PloS One, vol. 14, no. 1, p. e0211058, Jan. 2019, doi: 10.1371/journal.pone.0211058.

[23] J. Więckowski and W. Sałabun, “Evaluation of football players’ performance based on Multi-Criteria Decision Analysis approach and sensitivity analysis,” in Lecture notes in computer science, 2023, pp. 602–613. doi: 10.1007/978-981-99-8067-3_45.

[24] S. Manish, V. Bhagat, and R. Pramila, “Prediction of Football Players Performance using Machine Learning and Deep Learning Algorithms,” 2021 2nd International Conference for Emerging Technology (INCET), May 2021, doi: 10.1109/incet51464.2021.9456424.

[25] X. Cao, X. Zhao, H. Tang, N. Fan, and F. Zereg, “Football players’ strength training method using image processing based on machine learning,” PloS One, vol. 18, no. 6, p. e0287433, Jun. 2023, doi: 10.1371/journal.pone.0287433.

[26] G. Liu, Y. Luo, O. Schulte, and T. Kharrat, “Deep soccer analytics: learning an action-value function for evaluating soccer players,” Data Mining and Knowledge Discovery, vol. 34, no. 5, pp. 1531–1559, Jul. 2020, doi: 10.1007/s10618-020-00705-9.

[27] D. Barron, G. Ball, M. T. Robins, and C. Sunderland, “Artificial neural networks and player recruitment in professional soccer,” PloS One, vol. 13, no. 10, p. e0205818, Oct. 2018, doi: 10.1371/journal.pone.0205818.

[28] T. Bongiovanni et al., “Importance of anthropometric features to predict physical performance in elite youth soccer: a machine learning approach,” Research in Sports Medicine, vol. 29, no. 3, pp. 213–224, Aug. 2020, doi: 10.1080/15438627.2020.1809410.

[29] N. H. Nguyen, D. T. A. Nguyen, B. Ma, and J. Hu, “The application of machine learning and deep learning in sport: predicting NBA players’ performance and popularity,” Journal of Information and Telecommunication, vol. 6, no. 2, pp. 217–235, Sep. 2021, doi: 10.1080/24751839.2021.1977066.

[30] T. Yang, G. Yuan, and J. Yan, “Health analysis of Footballer using big data and deep learning,” Scientific Programming, vol. 2021, pp. 1–8, Jun. 2021, doi: 10.1155/2021/9608147.

[31] B. M. Merzah, M. S. Croock, and A. N. Rashid, “Football player tracking and performance analysis using the OpenCV Library,” Mathematical Modelling and Engineering Problems/Mathematical Modelling of Engineering Problems, vol. 11, no. 1, pp. 123–132, Jan. 2024, doi: 10.18280/mmep.110113.

[32] B. M. Merzah, M. S. Croock, and A. N. Rashid, “Intelligent classifiers for football player performance based on machine learning models,” International Journal of Electrical and Computer Engineering Systems, vol. 15, no. 2, pp. 173–183, Feb. 2024, doi: 10.32985/ijeces.15.2.6.

[33] V. K. Ojha, A. Abraham, and V. Snel, “Metaheuristic design of feedforward neural networks: A review of two decades of research,” Engineering Applications of Artificial Intelligence, vol. 60, pp. 97–116, Apr. 2017, doi: 10.1016/j.engappai.2017.01.013.

[34] J. R. Quinlan, “Induction of decision trees,” Machine Learning, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/bf00116251.

[35] J. Goldberger, G. E. Hinton, S. T. Roweis, and R. R. Salakhutdinov, “Neighbourhood components analysis,” Neural Information Processing Systems, vol. 17, pp. 513–520, Dec. 2004.

[36] N. Bayes, “An essay towards solving a problem in the doctrine of chances,” Biometrika, vol. 45, no. 3–4, pp. 296–315, Jan. 1958, doi: 10.1093/biomet/45.3-4.296.

[37] H. Dalianis, “Evaluation Metrics and evaluation,” in Springer eBooks, 2018, pp. 45–53. doi: 10.1007/978-3-319-78503-5_6.

**Baydaa M. Merzah** earned her MSc (2017) in computer engineering at Yildiz Technical University, College of Electric and Electronic engineering. She finished the B.Sc. degree in computer Science from University of Baghdad (2004). She is a staff member of Al-Nahrain University since 2008. The research interests include Software, Machine Leaning, Deep Learning and Computer Vision.

**Muayad S. Croock** earned his Ph.D. (2012) in Computer Engineering at Newcastle University, School of Electrical, Electronics and Computer Engineering. He finished his M.Sc. in Computer Engineering from University of Technology -Iraq (2003), and his B. Sc. Degree in Computer Engineering from the University of Technology (2003). Dr. Muayad is a staff member at University of Technology since 2004, where he currently holds a professor in Software Engineering. His research interests include Software Engineering, Security and WSN.

**Ahmed N. Rashid** Associated Professor at University of Anbar, College of Computer Science and Information Technology. Head of the Department of Computer Networks Systems. His research interest: Sensor Networks, Neural Networks and Ad hoc Networks and RFID Application.