ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
RESEARCH ARTICLE – COMPUTER ENGINEERING AND COMPUTER SCIENCE

2024 June

ARCHERY SHOOTINGS SCORE PREDICTION USING EEG SIGNALS

YUNUS EMRE KORKMAZ[1]

[1] Computer Engineering, Engineering Faculty, Eskisehir Technical University, Eskisehir, Turkey.

ABSTRACT

This research focuses on predicting archery shooting scores using EEG signals. The data used consists of EEG recordings from archers during their shots, accompanied by corresponding scores. This study aims to preprocess the data, extract relevant features, and apply various machine learning models to predict the scores accurately. We employ Random Forest Regressor, LSTM, and SVR models and tune their hyperparameters to improve performance. The results indicate that our approach effectively predicts scores, with the LSTM model showing the highest performance after hyperparameter tuning.

Keywords: Archery, Score Prediction, Machine Learning, EEG Signals

1. INTRODUCTION

Archery, a sport that demands high levels of precision, control, and consistency, involves not only physical coordination but also significant cognitive and mental discipline. The mental state of an archer during a shot can profoundly influence their performance [1]. Understanding and predicting this mental state can provide valuable insights for training and performance enhancement. One way to gain such insights is through the analysis of electroencephalography (EEG) signals, which capture the electrical activity of the brain.

EEG signals offer a non-invasive means to study brain function, revealing information about neural activities associated with various cognitive processes such as concentration, relaxation, and stress. These signals can be segmented into different frequency bands, each associated with distinct mental states. For instance, the Delta (4-8 Hz) and Theta (8-12 Hz) bands are often linked with relaxation and light sleep, whereas the Alpha (12-20 Hz) and Beta (20-30 Hz) bands are associated with active thinking and focus. By analyzing these frequency bands, we can potentially correlate specific brain activities with archery performance. [2]

This research focuses on leveraging EEG signals to predict archery shooting scores. We hypothesize that specific patterns in EEG signals, captured during the moments leading up to and during a shot, can be indicative of the archer's performance. To test this hypothesis, we collected EEG data from archers while they performed their shots. This data, along with the corresponding scores, forms the basis of our study.

The primary objectives of this study are to preprocess the EEG data, extract relevant features from the signals, and apply various machine learning models to predict archery scores. Specifically, we employ three models: Random Forest Regressor, Long Short-Term Memory (LSTM) networks, and Support Vector Regressors (SVR). Each model offers unique advantages: Random Forest is known for its robustness and interpretability, LSTM is well-suited for sequential data like EEG signals, and SVR can handle non-linear relationships effectively.

In addition to applying these models with default parameters, we conduct extensive hyperparameter tuning to optimize their performance. This involves systematically adjusting model parameters to improve accuracy and reliability in predictions. The performance of each model is evaluated using metrics such as Mean Squared Error (MSE), R-squared (R2), and Mean Absolute Percentage Error (MAPE).

This research not only aims to achieve high prediction accuracy but also seeks to understand the underlying importance of different EEG frequency bands in predicting archery performance. Through feature importance analysis using the Random Forest model, we identify which frequency bands contribute most significantly to the predictions.

In summary, this study explores the feasibility of using EEG signals for predicting archery shooting scores, offering a novel approach to enhance training and performance evaluation in the sport of archery. By bridging neuroscience and sports analytics, we aim to provide a deeper understanding of the cognitive aspects influencing archery performance and develop tools that can assist athletes in achieving better outcomes.

2. DATASET:

The dataset provided by Eskişehir Technical University Sports Science Faculty (Spor Bilimleri Fakültesi) includes EEG signals recorded during archery shootings and their corresponding scores. The dataset consists of two files:

2.1 EEG WITH NUMERIC LABELS:

The file contains EEG signals for the last second of each shot in 250 Hz frequency, along with the corresponding scores.

Table 1. The structure of file EEG signals and corresponding scores.

eeg_1	eeg_2	...	eeg_250	score
0.7177	-0.2062	...	-1.8025	9
-20.272	-15.527	...	-4.8429	7
...	...	...	...	...

2.2 SCORE VS. DISTANCE:

The file lists the scores and the corresponding distances of the arrows from the center of the target.

Table 2. Scores and the distance of the arrows from the center.

score	distance
9	2.5
7	6.2
8	5.1
...	...

3. METHODOLOGY

3.1 DATA PREPARATION AND PREPROCESSING:

In this part, score vs distance data and EEG signals has been merged into a single data frame and a new scoring algorithm has been created according to distance and score information.

3.1.1 CORRECTING DISTANCES:

The distances were adjusted to reflect the actual distances to the target center.

3.1.2 NORMALIZATION:

Min-max scaling was applied to normalize the corrected distances.

3.2 FEATURE EXTRACTION:

Frequency Bands Extraction: Power spectral density was computed for EEG signals in specific frequency bands: Delta (4-8 Hz), Theta (8-12 Hz), Alpha (12-20 Hz), and Beta (20-30 Hz).

3.2 MODELS:

3.2.1 RANDOM FOREST REGRESSOR:

A Random Forest Regressor is trained on the extracted features. The performance of this model is evaluated using mean squared error (MSE), R-squared (R2), and mean absolute percentage error (MAPE).

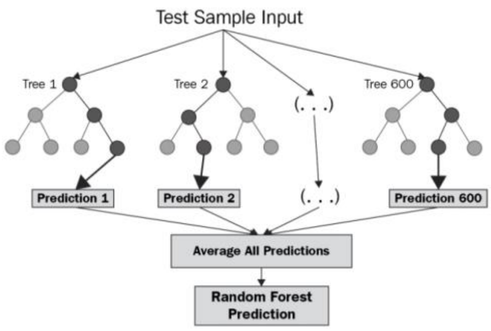

Figure 1. Random Forest Regressor Structure.

3.2.2 LSTM:

An LSTM (Long Short-Term Memory) model is trained on the same features. The LSTM is particularly suited for time series data and sequences, making it a good choice for EEG signal analysis.

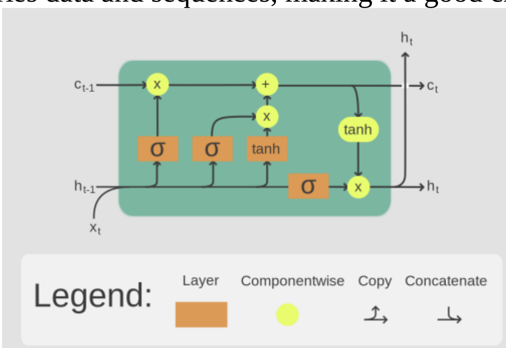

Figure 2. LSTM Model Structure

3.2.3 SUPPORT VECTOR REGRESSOR (SVR):

A Support Vector Regressor is also trained on the extracted features. This model uses a radial basis function kernel to handle non-linear relationships in the data.

3.3 HYPERPARAMETER TUNING:

3.3.1 RANDOM FOREST REGRESSOR TUNING:

Hyperparameter tuning for the Random Forest Regressor is performed using GridSearchCV. This process involves testing different combinations of hyperparameters to find the best configuration for the model.

3.3.2 LSTM TUNING:

The LSTM model undergoes hyperparameter tuning using RandomizedSearchCV. This technique tests a range of hyperparameter values to optimize the model's performance.

3.4 ERROR METRICS:
To evaluate the performance of the machine learning models used in this study, we employ three key error metrics: Mean Squared Error (MSE), R-squared (R2), and Mean Absolute Percentage Error (MAPE). These metrics provide a comprehensive assessment of the models' predictive accuracy and reliability.

3.4.1 MEAN SQUARED ERROR (MSE):
MSE measures the average squared difference between the predicted and actual values. It is defined as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

3.4.2 R-SQUARED (R2):
R2, also known as the coefficient of determination, measures the proportion of variance in the dependent variable that is predictable from the independent variables. It is defined as:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \mu)^2}.$$

3.4.3 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE):
MAPE measures the accuracy of the model's predictions as a percentage and is defined as:

$$\mathrm{MAPE} = 100\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

---

4. RESULTS:
The experiments conducted involved training three models: Random Forest Regressor, LSTM, and Support Vector Regressor (SVR). Each model was initially trained with default parameters, followed by hyperparameter tuning to optimize performance. The performance of each model was evaluated using Mean Squared Error (MSE), R-squared (R2), and Mean Absolute Percentage Error (MAPE). The results are summarized in Table 3 below.

Table 3. Model Outputs for Different Performance Metrics.

Model	MSE	R2	MAPE
Random Forest (Initial)	0.020297	0.023993	0.240657
LSTM (Initial)	0.019571	0.058923	0.240657
SVR (Initial)	0.019801	0.047852	0.240657
Random Forest (Tuned)	0.019065	0.083226	0.238599
LSTM (Tuned)	0.019079	0.082555	0.236162
SVR (Tuned)	0.019471	0.063735	0.237602

4.1 FEATURE IMPORTANCE ANALYSIS:

An analysis of feature importance was conducted using the Random Forest model. The importance of each frequency band as a feature in predicting archery scores is ranked in Table 4 below.

Table 4. Feature Importance Ranking (Random Forest).

Feature	Importance
Band (12-20 Hz)	0.3592440910524334
Band (20-30 Hz)	0.32340654047181994
Band (4-8 Hz)	0.3173493684757467

According to this, the most meaningful frequency band in this task is 12-20 Hz. (Beta band)

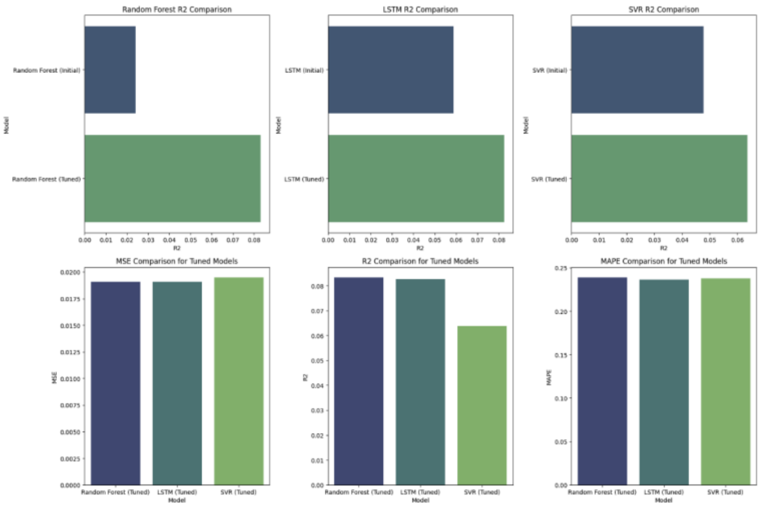

Figure 3. Bar Chart Results from related work

The results demonstrate that hyperparameter tuning improved the performance of all models, with the Random Forest model achieving the highest R2 score of 0.083226 and the lowest MAPE of 0.238599 after tuning. The LSTM model also showed significant improvement with an R2 score of 0.082555 and the lowest MAPE of 0.236162 among all models.

5. DISCUSSION

One of the primary challenges encountered in this research was the limited amount of data available for training and testing the models. The small dataset size restricts the ability of the models to generalize well to unseen data, which may affect their predictive performance.

Moreover, the dataset exhibited an imbalance in the distribution of scores, with certain scores appearing more frequently than others. This imbalance can lead to biased model predictions, as the models may become more attuned to predicting the more frequent scores accurately, while underperforming on the less frequent scores. Techniques such as resampling or using different performance metrics were considered to mitigate these issues, but the inherent limitation of the dataset size remained a significant challenge.

Future work should focus on collecting a more extensive and balanced dataset, potentially including data from a wider range of archers, and shooting conditions. This would enhance the robustness and generalizability of the predictive models, leading to more accurate and reliable score predictions based on EEG signals.

REFERENCES

[1] Vendrame, E., Belluscio, V., Truppa, L., Rum, L., Lazich, A., Bergamini, E., & Mannini, A. (2022). Performance assessment in archery: a systematic review. *Sports Biomechanics*, 1-23.

[2] Khosla, A., Khandnor, P., & Chand, T. (2020). A comparative analysis of signal processing and classification methods for different applications based on EEG signals. *Biocybernetics and Biomedical Engineering*, *40*(2), 649-690.