A Deep Learning Approach to Tennis Injury Prediction

ANONYMOUS AUTHOR(S)

Despite the widespread adoption of AI technologies in various sectors, their integration into sports performance optimization and injury prevention remains underexplored. This study focuses on utilizing AI to analyze and interpret collegiate tennis player's data–such as weight, workload, nutrition, and sleep–to accurately predict injury risks and enhance athletes' career longevity through optimized rest and training regimens. The aim of this research is to leverage Artificial Intelligence (AI), to bridge the gap between rapid technological advancements and their application in sports performance.

To address this challenge, we propose to develop a deep-learning algorithm capable of predicting potential injuries and future performance metrics. We employed a methodology involving the distribution of 8 WHOOP devices among D1 tennis players of both genders and the comprehensive user data includes sleep duration, workout/match intensity and length, and physiological parameters. The study further explores the application of tailored advice generated by the ChatGPT API to mitigate identified injury risks, thereby offering a novel approach to personalized athlete care and training optimization.

We anticipate that our findings will demonstrate the significant benefits of AI in transforming athlete training and health management. The expected outcomes include an accurate algorithm for injury risk prediction and the ability to deliver customized injury prevention advice. This technology benefits both athletes and coaches. Athletes can better understand and manage their training load, optimizing performance while minimizing injuries. Coaches gain valuable insights into athletes' stress levels, enabling them to tailor training programs for maximum benefit and longevity and modify their lineups to get the most competitive team for upcoming matches.

CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Additional Key Words and Phrases: Security, Penetration testing, operational systems

1 INTRODUCTION

2 OVERVIEW

Give an informal description of the designed system, using the workflow picture. Add an example of how a user can use the system here to help people follow up.

3 SET UP

In this section, we will introduce how we collect the necessary data.

3.1 Participant

3.2 Data Collection

The data we collect consists of three sources: Whoop devices, questionnaires, and physician report.

3.2.1 Whoop data. Whoop is ...., it tracks players' sleep data, training data....

3.2.2 Questionnaire.

3.2.3 Physician report.

3.3 Data pre-processing

4 INJURY PREDICTION

We design a model to predict injury probability scores using various machine learning techniques.

4.1 Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP) is a deep learning neural network model designed to predict an athlete's physical capability based on a sequence of physiological and activity metrics. This model leverages the temporal nature of the collected data to capture complex patterns that might indicate injury risk or performance potential.

4.1.1 Model Architecture. The MLP model consists of three fully connected layers:

(1) Input Layer: Accepts a flattened sequence of 7 days of data.
(2) Hidden Layer: Contains 128 units with ReLU activation.
(3) Output Layer: Produces a single value with sigmoid activation.

The model architecture can be represented as follows:

```
class MLPModel(nn.Module):
    def __init__(self, input_size, hidden_size, output_size):
        super(MLPModel, self).__init__()
        self.fc1 = nn.Linear(input_size, hidden_size)
        self.fc2 = nn.Linear(hidden_size, hidden_size // 2)
        self.fc3 = nn.Linear(hidden_size // 2, output_size)
        self.relu = nn.ReLU()
        self.sigmoid = nn.Sigmoid()

    def forward(self, x):
        out = self.relu(self.fc1(x))
        out = self.relu(self.fc2(out))
        out = self.sigmoid(self.fc3(out)) * 100  # Scale output to 0-100
        return out
```

4.1.2 Input Data. The model takes a flattened sequence of 7 days of data as input. Each day is represented by 49 numerical features, including metrics such as Activity Strain, Heart Rate, Sleep Duration, and various other physiological and activity-related measurements. The input data is normalized using StandardScaler before being fed into the model to ensure consistent scaling across features.

4.1.3 Output. The model produces a single scalar value representing the predicted physical capability percentage for the day following the input sequence. This value is scaled to range from 0 to 100, providing an easily interpretable metric for athletes and coaches.

4.1.4 Training Process. The model was trained on a dataset of 82,804 sequences, derived from the WHOOP device data collected from D1 tennis players. The training process involved:

(1) Data Preparation: Cleaning and normalizing the input data.
(2) Sequence Creation: Generating 7-day sequences from the time-series data.
(3) Model Training: Using Adam optimizer and Mean Squared Error (MSE) loss function.

The training was conducted for 100 epochs with a batch size of 32. The process showed a general decrease in both training and validation loss over time.

4.1.5 Results. The final epoch of training reported a training loss of 0.0388 and a validation loss of 0.6391. The model's predictions on the test set ranged from approximately 58% to 80% physical capability, providing a nuanced view of athletes' potential performance states.

4.1.6 *Discussion.*

The MLP model demonstrates the potential of using deep learning techniques to predict physical capability based on multi-day sequences of physiological and activity data. This approach allows for the capture of complex patterns and interactions between various metrics that may not be apparent in simpler models.

The range of predictions (58%–80%) suggests that the model is able to differentiate between various states of physical readiness, which could be valuable for both injury prevention and performance optimization. However, the relatively high validation loss indicates that there may be room for improvement, possibly through further feature engineering, model architecture refinement, or the incorporation of additional relevant data sources.

In the context of our tennis injury prediction system, the MLP model serves as a sophisticated tool for assessing an athlete's overall physical state. This information can be used in conjunction with other models and expert knowledge to provide comprehensive insights into injury risk and performance potential.

5 INJURY ADVISING

Based on the injury prediction, our design connects the LLMs to give suggestion for the player.

6 EVALUATION

In this section, we will evaluate the injury prediction effect based on various models.

6.1 XGB

6.2 Neural Network

7 RELATED WORK

8 CONCLUSION

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009