Hit or Miss? Leveraging CPS for Sports Performance

Danya Balakumaran Meenakumari
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA
yka2dk@virginia.edu

Ishita Gupta
Department of Computer Science
University of Virginia
Charlottesville, VA
rsy7tr@virginia.edu

I. Introduction

In the domain of sports analytics, the integration of wearable technology has revolutionized the way performance and training are approached. Wearable devices equipped with advanced sensors alllow the collection of high-resolution motion data, providing unparalleled insights into athletic movements. For a sport like tennis—known for its dynamic, fast-paced, and high-precision nature—such real-time data has immense potential to transform player training and game strategies. This project leverages wearable technology to overcome these challenges by offering real-time, objective, and data-driven feedback. The primary objective of this project is to classify tennis strokes as either hits (successful racket-ball contact) or misses (unsuccessful swings) using smartwatch sensor data. By integrating these data-driven insights into training regimens, this approach seeks to enhance an athlete's understanding of their performance on the court. Beyond tennis, this methodology can be extended to other sports, providing a versatile tool for performance analysis.

II. Methodology

A. Data Collection

The data for this study was collected using the ASUS Zenwatch 2 smartwatch equipped with accelerometer and gyroscope sensors. There were two participants in the experiment, each playing tennis for 10 minutes, swinging withe their right hands. During gameplay, two distinct activities were recorded: *hits*, defined as swings with impact (contact with the ball), and *misses*, defined as swings without impact.

The data collected from the sensors was segmented into sliding windows of varying intervals: 1-second, 2-second, 3-second, and 4-second - to ensure that relevant temporal patterns are captured for further analysis.

B. Data Preprocessing

To prepare the data for machine learning, statistical features were extracted from each sliding window. These features include the mean, median, standard deviation, and root mean square (RMS) of the sensor readings. To ensure uniform scaling and improve model performance, all features were standardized.

C. Model Training

Three machine learning models were trained and evaluated to classify tennis swings as hits or misses. The models included:

- **Random Forest:** An ensemble method using multiple decision trees to improve classification performance.
- **Logistic Regression:** A simple and interpretable linear model for binary classification.
- **Support Vector Machine (SVM):** A robust model designed to maximize the margin between classes.

Each model was trained using the preprocessed features, and their performance was evaluated across the sliding window intervals.

D. Feature Selection

Recursive Feature Elimination (RFE) was employed to identify the most significant features contributing to classification accuracy. This iterative technique systematically removed less important features, ultimately selecting the top three features. These selected features were then used for model training to optimize performance.

III. Results

The models' performance was evaluated based on accuracy across different time window intervals. The table below summarizes the results:

Model	**1-second**	**2-second**	**3-second**	**4-second**
Random Forest	0.9610	0.9506	0.9666	0.9652
Logistic Regression	0.9542	0.9461	0.9600	0.9652
Support Vector Machine	0.9519	0.9551	0.9666	0.9826

TABLE I
Accuracy of models across different time windows.

Table I displays the accuracy of three models across different time intervals.

The Random Forest model has consistently high frequency across all time intervals with best performance at 3-second time interval with(96.67%) with slight dip at 2-second time interval .

The logistic regression model shows slightly lower accuracy compared to Random forest but highlights its peak performance at 4-second interval(96.52%) indicating the consistent and stable performance but not up to the performance of SVM and Random forest at peak intervals.

The SVM model achieved the highest accuracy (98.26%) for the 4-second time window among all the models, which shows its superior performance in classifying hits and misses in tennis.This ensures the models robustness and effectiveness in classifying the classes with the selected features.

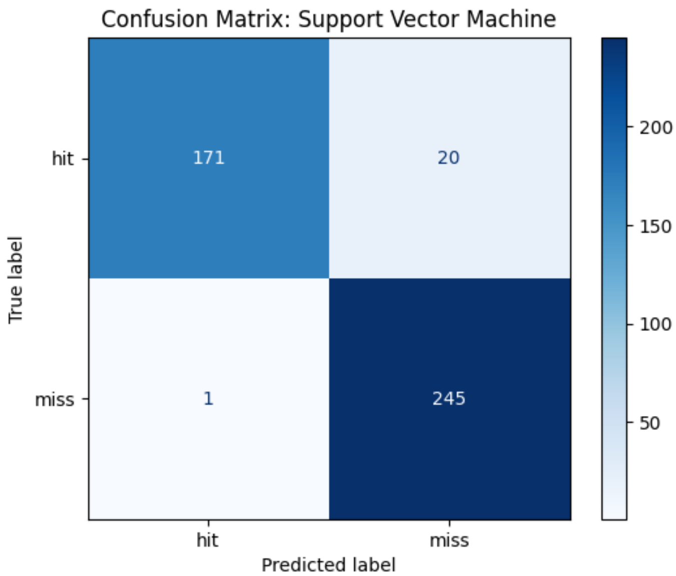

Fig. 1. Classification Matrix for SVM.

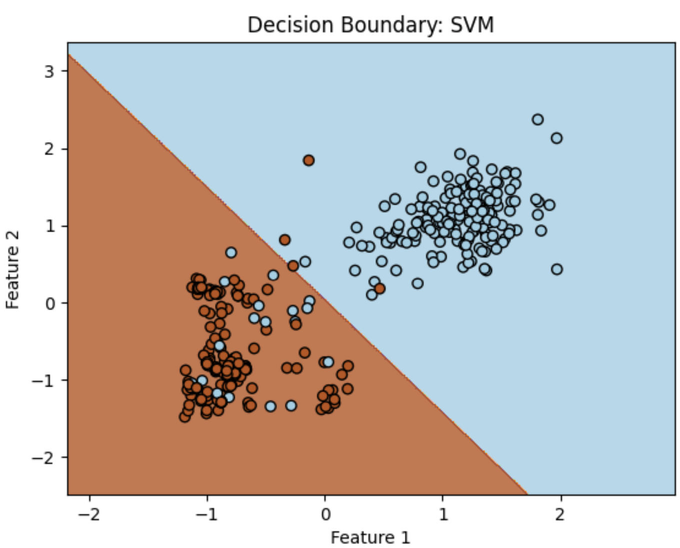

Fig. 2. Decision Boundary for SVM

The confusion matrix in the figure 2 highlights the classification performance of the SVM model. As it is shown, it has 171 true positives that the model predicted the positive

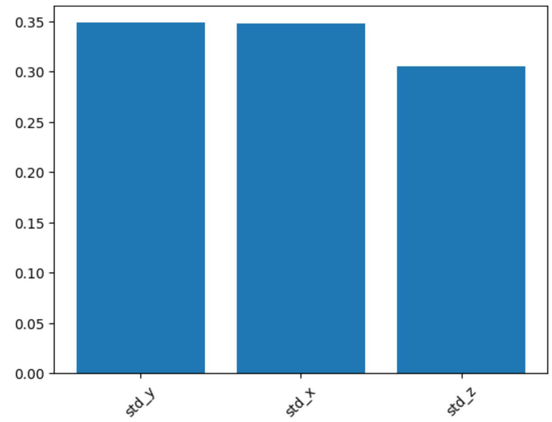

Fig. 3. Selected Features (SVM)

class and 245 true negatives indicating the high number of correct classifications of the wrong class when it is actually positive.With low false positives(1) and false negatives(20) highlights the model ability to predict the large majority of the cases accurately and efficiently.This performance demonstrates the model's high precision and recall, making it a reliable classifier for distinguishing between hits and misses.

The decision boundary plot (Figure 2) illustrates the separation between hits and misses in a two-dimensional feature space. This boundary is determined by the SVM to maximize the margin between the classes. The plot shows clear separation between the two classes with minimal overlap, indicating a good classifier and the distance between the boundary and the nearest data points of each class(margins) are large enough that indicates better generalization.The SVM model successfully delineated the two classes with a clear boundary, ensuring minimal overlap between them.

This visualization reflects the model's ability to generalize well to new data and highlights its effectiveness in separating classes based on the selected features. The low false positive and false negative rates indicate robust performance, critical for real-time sports applications.

IV. Conclusion

This study successfully demonstrated the application of wearable sensor data and machine learning for classifying tennis strokes as hits or misses, providing a solid foundation for sports performance analysis. Future work could involve utilizing multi-sensor systems, such as smart bands, for full-body movement tracking, and expanding the application to monitor tennis-specific metrics like spin rate and racket angle. These additions could allow for targeted performance improvements and real-time implementation across various sports, including racquetball and volleyball, further empowering athletes to achieve their best.

The code for the project can be found on GitHub.