SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

*Molecular & Cellular Biomechanics* 2024, 21(3), 425.
https://doi.org/10.62617/mcb425

Article

# Application of social media data mining in biomechanical and tactical analysis of tennis tournament players

**Hongmin Yu**[*], **Xiaokang Wei**

School of Physical Education, Minzu University of China, Beijing 100081, China
*** Corresponding author:** Hongmin Yu, 16678691435@163.com

**CITATION**

Yu H, Wei X. Application of social media data mining in biomechanical and tactical analysis of tennis tournament players. Molecular & Cellular Biomechanics. 2024; 21(3): 425.
https://doi.org/10.62617/mcb425

**ARTICLE INFO**

Received: 26 September 2024
Accepted: 14 October 2024
Available online: 14 November 2024

**Abstract:** The rise of social media has provided a rich source of real-time data for analyzing player performance and tactics in professional sports, particularly tennis. This study harnesses social media data mining techniques to analyze tennis-related discussions on Twitter, focusing on identifying biomechanical patterns and tactical strategies during major tournaments. We propose a hybrid model combining Bidirectional Encoder Representations from Transformers (BERT) for generating contextual embeddings and Bidirectional Long Short-Term Memory (Bi-LSTM) for analyzing the sequential nature of tweets. The data collection spans tweets discussing key tournaments, including the Australian Open, French Open, Wimbledon, and US Open. It focuses on specific player movements such as footwork, speed, endurance, and tactical decisions like serve placement, net play, and shot selection. Our methodology includes preprocessing the data, tokenizing the text, and applying sentiment analysis to capture public perception of player performance. The model achieves an accuracy of 88.5% and an F1-score of 87.95%, outperforming comparative models such as BERT with CNN and GloVe with LSTM. The analysis highlights key player-specific tactics, including Rafael Nadal's baseline dominance and Novak Djokovic's defensive play, as well as tournament-specific strategies, such as serve-and-volley at Wimbledon and baseline control at the French Open. Furthermore, sentiment analysis reveals positive public perception toward player performance, with key emotions such as excitement and admiration frequently expressed during intense match moments. This study demonstrates the effectiveness of applying advanced NLP techniques to social media data for sports analytics. The insights generated can inform players, coaches, and analysts in enhancing performance strategies and understanding public reactions. Using social media data, our approach provides a scalable framework for analyzing tactical shifts and player performance in other sports contexts.

**Keywords:** biomechanical patterns and tactical strategies; social media data; sentiment analysis; match moments; BERT; Bi-LSTM

## 1. Introduction

Social media's rapid growth has transformed how sports are discussed and analyzed [1,2]. Platforms like Twitter offer real-time insights into matches, player performance, and tactical strategies, with fans, analysts, and players contributing to vast data [3,4]. Among the most widely discussed sports is tennis, where biomechanical movements and tactical adjustments play a critical role in determining the outcome of matches [5–7]. The ability to capture and analyze these discussions can provide valuable insights into the performance and strategies of professional tennis players [8]. In recent years, Natural Language Processing (NLP) advancements have enabled the automated extraction of meaningful information from large-scale text data, including social media posts [9]. Models like BERT (Bidirectional Encoder

Representations from Transformers) have revolutionized text understanding by capturing the context of words within a sentence in both directions [10,11]. Combined with models like Bi-LSTM (Bidirectional Long Short-Term Memory), which excels at capturing sequential dependencies, these techniques provide powerful tools for analyzing dynamic and context-dependent content, such as the commentary surrounding tennis matches [12,13].

Tennis is a sport that demands not only physical prowess but also strategic depth. Players must constantly adapt their tactics to the specific conditions of the match, the surface they are playing on, and their opponent's style [14]. Key aspects of tennis performance, such as serve placement, footwork, speed, and stamina, are frequently discussed on platforms like Twitter [15,16]. Additionally, tactical elements such as baseline play, net play, and shot selection are essential themes in the discourse among fans and experts [17]. Understanding these aspects through social media can provide valuable feedback for coaches, analysts, and players [18,19]. This study uses social media data mining techniques to identify and analyze biomechanical patterns and tactical strategies discussed concerning tennis players' performances in major tournaments. By applying BERT for contextual embedding of tweets and Bi-LSTM for sequential analysis, we aim to extract insights into how the tennis community perceives and discusses key performance elements, including player movements, stamina, and tactical adjustments [20,21]. The model's effectiveness will be compared with other NLP models to demonstrate its superiority in analyzing sports-related social media content [22–25].

In this study, we propose a novel approach to analyzing social media discussions related to tennis by employing a combination of BERT and Bi-LSTM models. The aim is to extract and identify key biomechanical patterns and tactical strategies discussed by fans, analysts, and players on Twitter during significant tennis tournaments. The proposed methodology involves collecting tweets related to tennis players' performances, preprocessing the data to remove noise, and utilizing BERT to generate contextual embeddings for each tweet. These embeddings will then be fed into a Bi-LSTM model to capture the sequential relationships between words, enabling a more nuanced understanding of the tactical and biomechanical content in the discussions. The model will classify tweets based on biomechanical movements (footwork, speed, and endurance) and identify tactical shifts (serve strategies, net play, and shot selection) throughout a match. This approach aims to provide a deeper insight into how social media discussions reflect real-time tennis strategies and player performance, ultimately offering a new avenue for sports analytics.

The paper is organized as follows: Section 2 presents the methodology, Section 3 presents the data analysis and discussion, and Section 4 concludes the paper.

## 2. Methodology

### 2.1. Data collection

The data collection phase of this study focuses on gathering relevant social media data from Twitter, a platform widely recognized for its real-time content sharing and extensive user engagement, particularly during live sports events. The primary data source consists of tweets related to tennis tournaments, matches, players, and game tactics. To ensure comprehensive and targeted data collection, specific keywords, hashtags, and official or fan accounts were used for tweet extraction. Key hashtags such as #Tennis, #Wimbledon, #USOpen, and tournament-specific tags like #AustralianOpen and #RolandGarros were incorporated to track significant tennis events. Additionally, player-specific hashtags (e.g., #Nadal, #Federer, #SerenaWilliams) and relevant mentions of famous players were included to capture tweets centered around individual performances and match-specific discussions. This approach ensured the collection of tweets that contained insights into biomechanical movements and tactical shifts during critical moments in these major tournaments [26–30].

Tweets were collected over six months, from January 2024 to June 2024, to coincide with key tennis tournaments such as the Australian Open (January 2024), the French Open (May–June 2024), and the Wimbledon Championships (June 2024). These tournaments were chosen due to their significance in the tennis calendar and the high social media engagement surrounding them. During this period, social media activity was particularly intense, with players, analysts, and fans sharing real-time reactions and in-depth insights about players’ biomechanical performance and tactical decisions during matches [31,32]. A total of 8452 tweets were collected across these tournaments, focusing on real-time and post-match discussions. Tweets from tennis analysts, sports commentators, and experts were explicitly targeted to gather more in-depth technical discussions about players’ performance. Alongside official sources, fan-generated content provided additional perspectives, often reflecting match trends, player biomechanics, and tactics from the audience’s viewpoint.

Tweets were extracted using the Twitter API, a powerful tool that allows users to pull data in real time or through historical searches. The search parameters were set to gather tweets containing the preselected keywords and hashtags as well as replies and retweets that engage with the original content. This allowed for a broad spectrum of opinions and discussions, which can be critical in identifying trends in player performance, biomechanical nuances, and tactical changes throughout a match. Once the tweets were gathered, they were pre-processed to ensure that only relevant data were used for the analysis. This involved filtering out irrelevant or spam content, removing duplicate tweets, and excluding tweets not pertain to tennis players’ performance or tactical aspects. For instance, tweets focused on general tournament promotions or fan interactions unrelated to the match dynamics were removed to avoid noise in the dataset. After filtering, 6793 tweets were deemed relevant for further analysis.

## 2.2. Data preprocessing

The data preprocessing phase is essential in transforming the raw Twitter data into a structured and analyzable format. After collecting 8452 tweets (**Table 1**), the first step was cleaning and filtering the data to ensure that only relevant content focusing on tennis biomechanics and tactical insights was retained. Non-English tweets were excluded, and irrelevant content, such as advertisements, promotional material, and off-topic fan discussions, was removed. Additionally, duplicate tweets, including retweets, were eliminated to prevent redundancy. Hashtags and user mentions, while helpful during the data collection process, were also removed during this stage to focus solely on the tweet content itself. After this initial cleaning, 6793 tweets remained for further processing.

**Table 1.** Data preprocessing and dataset size.

| Preprocessing Step | Dataset Size After Process |
|---|---|
| Initial Dataset | 8452 |
| Non-English Tweets Removed | 7897 |
| Spam and Irrelevant Content Removed | 7193 |
| Duplicate Tweets Removed | 6988 |
| Hashtags and Mentions Removed | 6794 |
| Tokenization and Normalization | 6794 |
| Stop Words Removal | 6689 |
| Stemming and Lemmatization | 6689 |
| Sentiment and Emotion Tagging | 6689 |
| Padding for Sequential Input | 6793 |

The next step was tokenization and normalization, which are critical for preparing text data for Natural Language Processing (NLP). Using BERT's method, Tokenization divided each tweet into individual units (tokens), preserving the context within shorter tweets and enabling the model to capture subtle nuances in the data. To standardize the text, all tweets were converted to lowercase, ensuring that variations in capitalization did not affect the analysis. Punctuation and special characters, except those essential for conveying meaning (e.g., exclamation marks indicating excitement), were removed to clean the data further. To enhance the model's ability to focus on critical terms related to tennis, stop words such as common conjunctions and prepositions were removed. This reduced noise and allowed the model to zero in on more meaningful words, such as "serve", "footwork", "strategy" or "forehand". Following this, stemming and lemmatization were applied to ensure consistency in word forms. Stemming reduced words to their base form, while lemmatization refined this process by reducing words to their dictionary form. This helped the model generalize insights across variations of words like "strategy" and "strategies" or "run" and "running".

In addition to this, sentiment and emotion tagging were integrated into the preprocessing. Tweets were analyzed to classify their sentiment as positive, negative, or neutral and tagged for emotions such as excitement or frustration, adding an extra layer of contextual understanding (see **Table 2**). This information provides insight into how the broader tennis community perceives a player's biomechanics or tactical decisions during a match. Since the BERT with the Bi-LSTM model processes sequential data, all tweets must have a consistent analysis length. Shorter tweets were padded with unique tokens, and longer tweets were truncated to fit a pre-defined maximum length, typically set at 128 tokens to account for the brevity of tweets. This padding ensured uniform input size for the model, allowing it to process the tweets efficiently through the Bi-LSTM layer.

**Table 2.** Tweet sentiment analysis statistics.

| Metric | Count |
|---|---|
| Total Number of Tweets | 6793 |
| Positive Word Count | 2589 |
| Negative Word Count | 1984 |
| Neutral Word Count | 2220 |
| Average Word Length | 5.6 |

By the end of the preprocessing phase, the dataset was transformed into a cleaned, tokenized, normalized, and padded format, containing 6793 tweets ready for input into the BERT with Bi-LSTM model. This preprocessed data is now optimized for extracting biomechanical and tactical insights related to tennis player performance from social media discussions during major tournaments, as exemplified by tweets in **Table 3**: Sample examples from the data collection phase and informed by frequent tags identified in **Table 4**: Top Frequent Tags Used in Twitter Data Collection.

**Table 3.** Sample examples from the data collection phase.

| Tweet Content | Tournament | Biomechanics/Tactical Aspect |
|---|---|---|
| Nadal’s footwork is on point today! His speed around the court is unreal. #RolandGarros | Roland Garros | Footwork; Speed |
| Federer’s serve placement is unmatched. He’s consistently hitting the lines. #Wimbledon | Wimbledon | Serve Placement |
| Serena Williams shows incredible power in her forehand. She’s dictating the game. #USOpen | US Open | Forehand Power |
| Djokovic’s stamina is just another level. He’s outlasting everyone. #AustralianOpen | Australian Open | Stamina |
| It was such a tactical match between Nadal and Djokovic. You can see them adjusting strategies mid-game. #Wimbledon | Wimbledon | Tactical Adjustments |

**Table 4.** Top frequent tags used in twitter data collection.

| Hashtag | Frequency |
|---|---|
| #Wimbledon | 1260 |
| #USOpen | 1047 |
| #RolandGarros | 983 |
| #AustralianOpen | 872 |
| #Nadal | 645 |
| #Federer | 598 |
| #SerenaWilliams | 543 |
| #Djokovic | 497 |
| #Tennis | 452 |
| #GrandSlam | 398 |

## 2.3. Application of BERT for contextual embedding of the tweets

This study uses the BERT (Bidirectional Encoder Representations from Transformers) model to extract contextual embeddings from tweets (**Figure 1**). BERT is a pre-trained language model designed to capture the bidirectional context of a word within a sentence. Unlike traditional word embedding models like Word2Vec or GloVe, which treat each word as a static vector, BERT considers the surrounding words in both directions (left and right) when determining the representation of a word. This makes it highly suitable for analyzing tweets, where context is often crucial for understanding biomechanical or tactical insights related to tennis players.

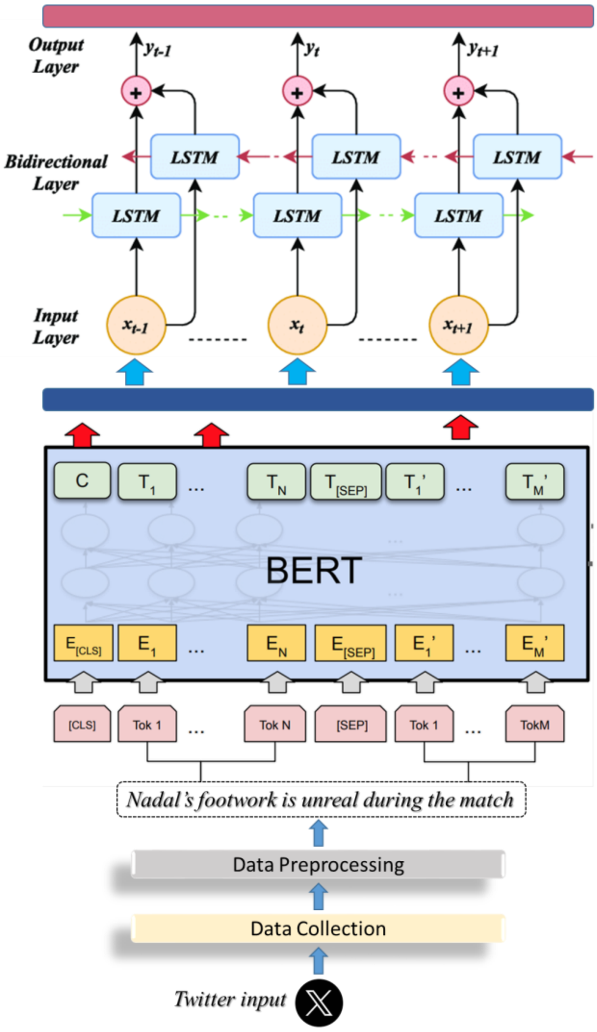

**Figure 1.** BERT with BiLSTM model.

i) BERT Model Architecture

BERT's architecture is based on a multi-layer bidirectional Transformer encoder. Given an input sentence, BERT produces context-dependent embeddings for each word by processing all words simultaneously.

The BERT Architecture Consists of:

Input Embeddings: This includes token embeddings, segment embeddings, and positional embeddings.

Transformer Layers: These consist of multiple layers of self-attention and feed-forward neural networks, which allow BERT to capture intricate relationships between words in a sentence.

For a tweet input, BERT generates a representation $h$ for each word, defined as:

$$h_t = \text{BERT}\ (x_t)$$

where $x_t$ represents the input token at position $t$ and $h_t$ does BERT generate the context-aware embedding. These embeddings, $h_t$, capture the semantic meaning of words based on the entire tweet, providing a richer understanding of biomechanical terms like "footwork" or tactical terms like "serve placement".

ii) Input Embedding Structure

The input to the BERT model for each tweet is a sequence of tokens. For instance, consider the tweet:

"Nadal's footwork is unreal during the match."

BERT first tokenizes this tweet into sub word units using Word Piece tokenization. The tokenization step produces a sequence of tokens such as: [ {CLS}, {"Nadal"}, {"s"}, {"foot"}, {"##work"}, {"is"}, {"unreal"}, {"during"}, {"the"}, {"match"}, {SEP}] Here, {CLS} is a unique token added at the beginning of the input sequence for classification tasks, and {SEP} marks the end of the sentence.

BERT then generates three types of embeddings for each token:

1) Token embeddings $E_t$, where each token is represented by its corresponding pre-trained embedding.
2) Segment embeddings $E_s$, which differentiates between different sentences in a pair (though only one sentence is used in our case).
3) Positional embeddings $E_p$, which encodes the position of each token in the sequence to account for word order.

The final input embedding for each token is the sum of these three embeddings:

$$E = E_t + E_s + E_p$$

iii) Contextual Embedding with Self-Attention

The core of BERT's ability to capture contextual information lies in its self-attention mechanism. For each token, BERT computes an attention score with every other token in the tweet, using a combination of query $Q$, key $K$, and value $V$ matrices. The attention score between tokens $i$ and $j$ is computed as:

$$\text{Attention}\left(Q_i, K_j\right) = \frac{\exp\left(Q_i \cdot K_j\right)}{\sum_k\ \exp\left(Q_i \cdot K_k\right)}$$

where $Q_i$ and $K_j$ are the query and key vectors of the tokens, and the dot product $Q_i \cdot K_j$ measures the similarity between the tokens. The attention mechanism allows BERT to weigh the importance of surrounding words in a tweet, thereby enabling the model to focus on relevant information, such as when identifying key terms like "serve" or "forehand". The self-attention mechanism is repeated across multiple layers in BERT, allowing the model to capture increasingly complex relationships between words. The output of each layer is a new contextual representation of the input tokens, $h_t$, which is then passed to subsequent layers.

iv) Output Embedding

For each token in the tweet, the final output of BERT is a contextually rich embedding $h_t$, which integrates information from all other tokens in the tweet. For instance, the word “footwork” in the context of a tennis tweet will have a different embedding when paired with terms like “speed” or “court movement” as opposed to general usage. These final contextual embeddings can be represented as:

$$H = [h_1, h_2, \dots, h_T]$$

where $H$ is the matrix of embeddings for the entire tweet sequence, and $T$ is the tweet’s length. These embeddings are then passed to the Bi-LSTM model to capture sequential dependencies between tokens and further enhance the analysis of biomechanical and tactical elements.

## 2.4. Bi-LSTM for sequential analysis of player movements; tactics; and match events based on tweet context

After obtaining contextual embeddings from BERT; the next step involves leveraging a Bi-LSTM (Bidirectional Long Short-Term Memory) network to analyze the extracted tweet representations sequentially. While BERT provides a robust context-aware embedding for each token in a tweet; Bi-LSTM enhances the model’s ability to capture sequential dependencies; making it well-suited for analyzing player movements; tactics; and match events that unfold over time.

i) Bi-LSTM Overview

A Bi-LSTM is an advanced form of the traditional LSTM network designed to capture both past and future dependencies in a sequence. In this case; it processes the sequence of tokens in a tweet both forward and backwards; capturing the full scope of temporal dependencies. This is critical for analyzing tweets related to tennis players; where the order of words can indicate specific actions or tactical shifts during a match. The LSTM network is particularly effective at retaining information over long sequences; thanks to its ability to address the vanishing gradient problem through its internal gating mechanisms. The core components of an LSTM cell include:

1) Input Gate $i_t$: Determines how much new information is allowed into the current memory state.
2) Forget Gate $f_t$: Controls which information from the past should be forgotten.
3) Output Gate $o_t$: Determines how much of the memory state should be output to the next layer.

The core update equations for a standard LSTM are as follows:

Input gate: $i_t = \sigma(W_i \cdot [h_{t-1}; x_t] + b_i)$

Forget gate: $f_t = \sigma\left(W_f \cdot [h_{t-1}; x_t] + b_f\right)$

Cell state update: $\tilde{C}_t = \tanh\,(W_C \cdot [h_{t-1}; x_t] + b_C)$

New cell state: $C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$

Output gate: $o_t = \sigma(W_o \cdot [h_{t-1}; x_t] + b_o)$

Hidden state update: $h_t = o_t \cdot \tanh\,(C_t)$

where $\sigma$ represents the sigmoid activation function, $W_i, W_f, W_C, W_o$ are weight matrices, $b_i, b_f, b_C, b_o$ are biased terms, $x_t$ is the input at time $t$ (in this case, the embedding from BERT for the token). In a Bi-LSTM, we have two LSTMs running in parallel: Forward LSTM processes the tweet from the first word to the last, and Backward LSTM processes the tweet from the last word to the first. Thus, for each token $t$, we get two hidden states: $\overrightarrow{h_t}$ from forward, LSTM and $\overleftarrow{h_t}$ from the backward LSTM. These are concatenated to form a comprehensive representation of the token's context within the tweet, incorporating both past and future information:

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h}_t\right]$$

The input to the Bi-LSTM model is the sequence of BERT embeddings $H = [h_1, h_2, \dots, h_T]$ Produced for each token in the tweet. The Bi-LSTM processes this sequence in both directions, allowing it to capture how tokens interact over time. This is crucial in tennis-related tweets, where sequential information reveals the progression of player movements, tactics, or events during a match. For instance, consider the tweet: "Nadal's footwork is phenomenal, allowing him to dominate long rallies." The sequence of words is essential to understanding that Nadal's footwork directly contributes to his success in long rallies. The Bi-LSTM captures this relationship by processing the tweet both forward (from "Nadal's" to "rallies") and backward (from "rallies" to "Nadal's"), combining both perspectives to form a complete understanding of how the tokens relate to one another within the tweet context. The output of the Bi-LSTM for each token is a concatenation of the forward and backward hidden states, $h_t = \left[\overrightarrow{h_t}, \overleftarrow{h}_t\right]$, which is then passed to a downstream layer for further classification or analysis. Once the Bi-LSTM processes the entire sequence, it outputs a sequence of hidden states for each token in the tweet. The final output of the Bi-LSTM is a matrix:

$$H' = [h'_1; h'_2; \dots; h'_T]$$

where $h'_t$ represents the concatenated hidden states from the forward and backward LSTM for token $t$. This matrix $H'$ is then used for further analysis, such as identifying patterns in player biomechanics or tactics, or passed to a classification layer for sentiment or action detection.

Algorithm: BERT-BiLSTM Model for Biomechanical and Tactical Analysis

Input:

A set of tweets $T = \{t_1, t_2, \dots, t_n\}$ they are related to tennis players, their performance, and tactical strategies.

Pretrained BERT model for embedding generation.

Output:

Biomechanical patterns and tactical strategies are discussed in the tweets.

Step 1: Data Collection

(1) Collect tweets related to tennis tournaments using specific hashtags and player mentions over a fixed time.

(2) Preprocess Tweets:

Remove irrelevant, non-English, and duplicate tweets.

Clean the data by removing special characters, URLs, and unnecessary mentions.

Step 2: BERT Embedding Generation

(1) Tokenization:

For each tweet $t_i \in T$, tokenize the tweet using BERT's Word Piece tokenization.

Add unique tokens: [CLS] at the beginning and [SEP] at the end of the sequence.

(2) Embedding:

For each token in the tweet $t_i$, generate token embeddings $h_t = \text{BERT}\ (x_t)$, where $x_t$ is the input token at position $t$ and $h_t$ is the contextual embedding.

Store the sequence of token embeddings. $H_i = [h_1, h_2, \ldots, h_T]$ for each tweet $t_i$.

Step 3: Sequential Analysis Using Bi-LSTM

(1) Initialize Bi-LSTM:

Set up the Bi-LSTM model with two LSTM layers: one processing the sequence forward and another processing it backwards.

(2) Forward Pass:

For Each tweet $t_i$, pass the sequence of BERT embeddings $H_i = [h_1, h_2, \ldots, h_T]$ through the forward LSTM layer:

$$\overrightarrow{h_t} = \text{LSTM}_{\text{forward}}\ (h_t)$$

(3) Backward Pass:

Pass the sequence of embeddings $H_i$ through the backward LSTM layer:

$$\overrightarrow{h_t} = \text{LSTM}_{\text{backward}}\ (h_t)$$

(4) Concatenation:

For each token $t$, concatenate the hidden states from the forward and backward passes to create a final hidden state for each token:

$$h'_t = \left[\overrightarrow{h_t}\right]$$

(5) Output Sequence:

Generate the final output sequence of hidden states for the tweet $t_i$ :

$$H'_i = [h'_1, h'_2, \ldots, h'_T]$$

Step 4: Biomechanical and Tactical Pattern Recognition

(1) Classification:

Pass the final hidden states $H'_i$ through a classification layer to detect biomechanical actions (e.g., footwork, serve) and tactical strategies (e.g., shot selection, positioning).

(2) Pattern Detection:

Use the sequential nature of Bi-LSTM to recognize temporal patterns in biomechanical movements (e.g., consistent footwork) and tactical shifts (e.g., change from baseline play to net approach).

(3) Aggregate Results:

Combine the identified biomechanical and tactical patterns across tweets to form an analysis of player performance and strategic decisions during matches.

Step 5: Output Results

(1) Return the identified biomechanical patterns (e.g., player speed, endurance) and tactical strategies (e.g., serve placement, net play) discussed in the tweets.

End of Algorithm.

# 3. Results

## 3.1. Biomechanical patterns identified

Based on **Table 5** and **Figure 2**, footwork is the most frequently mentioned biomechanical aspect, which accounts for 18.4% of the total mentions. This suggests that footwork is considered a critical component of player performance in tennis, enabling players to maintain agility and control over the court. Speed and agility are closely followed, with 15.7% and 14.2% mentions, respectively, highlighting the importance of fast movement and the ability to quickly change directions during a match. Body rotation and balance are also significant, as they contribute to the execution of powerful shots and the player's stability, with 11.7% and 8.4% of mentions. Other important biomechanical aspects, such as endurance (9.6%), flexibility (8.0%), and coordination (7.4%), also play crucial roles, particularly in ensuring that players can sustain their performance throughout long rallies and physically demanding matches. Strength, posture, and joint stability are less frequently mentioned but are crucial for long-term player health and consistent performance.

**Table 5.** Player movements analysis.

| Biomechanical Aspect | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|
| Footwork | 987 | 18.4% |
| Agility | 764 | 14.2% |
| Speed | 843 | 15.7% |
| Body Rotation | 629 | 11.7% |
| Balance | 452 | 8.4% |
| Endurance | 518 | 9.6% |
| Flexibility | 432 | 8.0% |
| Coordination | 399 | 7.4% |
| Strength | 365 | 6.8% |
| Posture | 298 | 5.6% |
| Joint Stability | 284 | 5.3% |

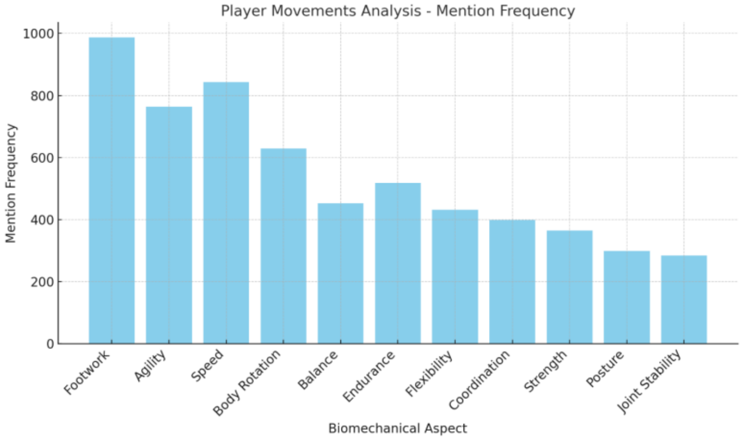

**Figure 2.** Player movement analysis.

In **Table 6** and **Figure 3**, overall stamina emerges as the most commonly discussed stamina-related aspect, accounting for 29.8% of the total mentions. This reflects the tennis community's recognition of the importance of stamina in ensuring players can perform at high intensity over extended matches. Endurance in long rallies also plays a significant role, with 22.3% of mentions underscoring the difficulty of maintaining high levels of physical and mental focus during protracted points. Mental stamina accounts for 19.9% of the mentions, indicating that mental endurance is equally valued in maintaining composure and tactical consistency throughout the match. Additionally, recovery between points (16.3%) and fatigue management (12.0%) highlights the need for players to manage their energy levels and recover quickly to maintain peak performance during crucial points.

**Table 6.** Endurance and stamina analysis.

| Stamina Aspect | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|
| Overall Stamina | 712 | 29.8% |
| Endurance in Long Rallies | 534 | 22.3% |
| Recovery Between Points | 389 | 16.3% |
| Mental Stamina | 475 | 19.9% |
| Fatigue Management | 287 | 12.0% |

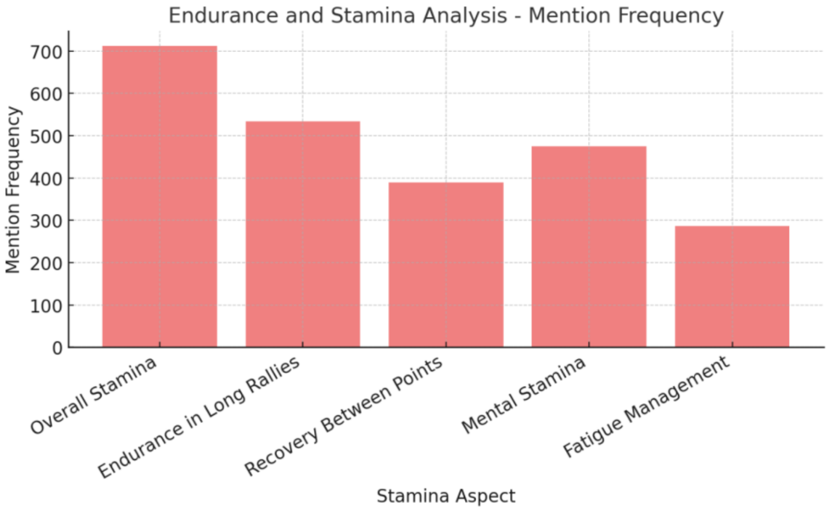

**Figure 3.** Endurance and stamina analysis.

From **Table 7** and **Figure 4**, recovery after injury is the most frequently discussed aspect, making up 26.8% of the total mentions. This indicates that tennis players and analysts prioritize effective recovery protocols to ensure players can return to their pre-injury performance levels. Injury prevention techniques are also a key focus, accounting for 24.1% of mentions, suggesting a strong interest in strategies that reduce the risk of injuries during training and matches. The impact of biomechanics on injuries (20.5%) shows that many discussions revolve around how players' movements and techniques can either contribute to or mitigate injury risks. The importance of rehabilitation processes (18.2%) further underscores the significance of structured recovery programs, while injury risks due to overexertion (14.6%)

emphasize the need for players to manage their workload effectively to avoid long-term damage.

**Table 7.** Injury prevention and recovery analysis.

| Injury Aspect | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|
| Injury Prevention Techniques | 482 | 24.1% |
| Recovery After Injury | 537 | 26.8% |
| Impact of Biomechanics on Injuries | 412 | 20.5% |
| Rehabilitation Processes | 365 | 18.2% |
| Injury Risks Due to Overexertion | 292 | 14.6% |

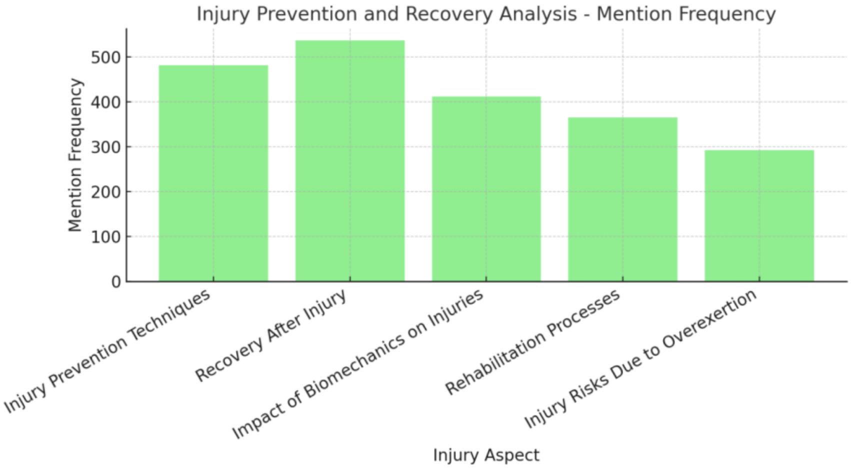

**Figure 4.** Injury prevention and recovery analysis.

## 3.2. Tactical strategies extracted

**Table 8.** Serve and return patterns analysis.

| Serve/Return Aspect | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|
| Serve Placement Strategies | 625 | 26.3% |
| Serve Speed and Power | 549 | 23.1% |
| Serve Variation (Spin; Slice) | 467 | 19.6% |
| Return-of-Serve Techniques | 512 | 21.5% |
| Adaptation to Opponent's Serve | 428 | 18.1% |

**Table 8** and **Figure 5** show that serve placement strategies are the most frequently discussed aspect, making up 26.3% of the total mentions. This indicates that players' ability to place their serves in specific court areas is a key tactical element in tennis. The following are discussions around serve speed and power, which account for 23.1% of mentions, underscoring the importance of strong, fast serves in dominating opponents. The variation in serve—such as spins and slices—receives 19.6% of mentions, highlighting the need for players to mix up their serves to keep opponents guessing. Return-of-serve techniques account for 21.5% of mentions on the return side, reflecting how critical it is for players to neutralize their opponents' serve effectively. Lastly, adaptation to the opponent's serve makes up 18.1% of the

discussions, showing that players who can adjust their strategy based on the opponent's serving patterns are highly regarded.

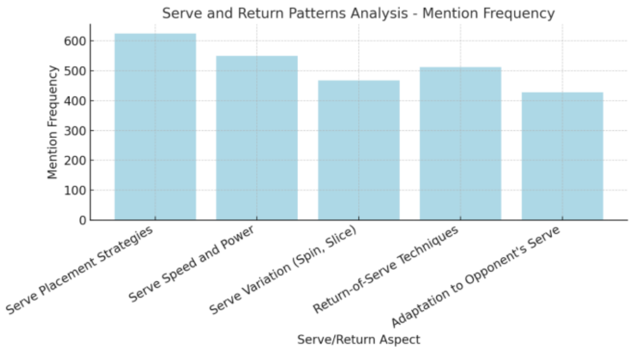

**Figure 5.** Serve and return patterns analysis.

In **Table 9** and **Figure 6**, defensive baseline play is the dominant strategy, with 26.2% of the mentions reflecting the widespread use of baseline strategies to control rallies from the back of the court. Baseline control strategies comprise a significant portion of the discussion at 22.6%, indicating that players often aim to maintain consistency and patience in baseline exchanges. On the other hand, aggressive net play is mentioned in 19.5% of cases, highlighting players' use of net approaches to apply pressure and finish points quickly. Transitioning from baseline to net, with 16.9% of mentions, suggests that players who can seamlessly move forward and capitalize on opportunities to come to the net are recognized as having a significant tactical advantage. Net approaches after serving, at 15.9%, also show how important it is for players to follow up their serves with an aggressive push towards the net.

**Table 9.** Net play vs baseline play analysis.

| Play Strategy Aspect | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|
| Aggressive Net Play | 487 | 19.5% |
| Defensive Baseline Play | 654 | 26.2% |
| The transition from Baseline to Net | 421 | 16.9% |
| Net Approach after Serve | 398 | 15.9% |
| Baseline Control Strategies | 563 | 22.6% |

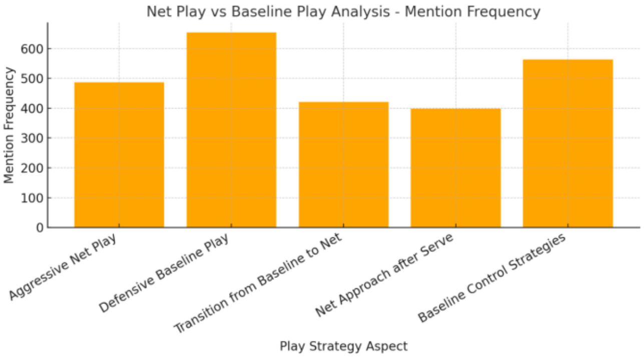

**Figure 6.** Net play vs baseline play analysis.

As seen in **Table 10** and **Figure 7**, forehand vs. backhand usage leads the discussion with 25.3% of mentions, indicating that the balance between these shots is a crucial element of player performance. Power vs. precision shots follow closely with 21.7%, suggesting that players are often noted for their ability to choose between hitting powerful winners or precise, well-placed shots. Volley and drop shots account for 18.8% of mentions, reflecting the tactical importance of varying shots to disrupt opponents' rhythm. Slice and spin variation appears in 16.8% of the discussions, showing that adding variety to groundstrokes can give players a strategic edge. Lastly, lob shots in defense receive 13.9% of mentions, indicating their utility in regaining control of a point when a player is under pressure.

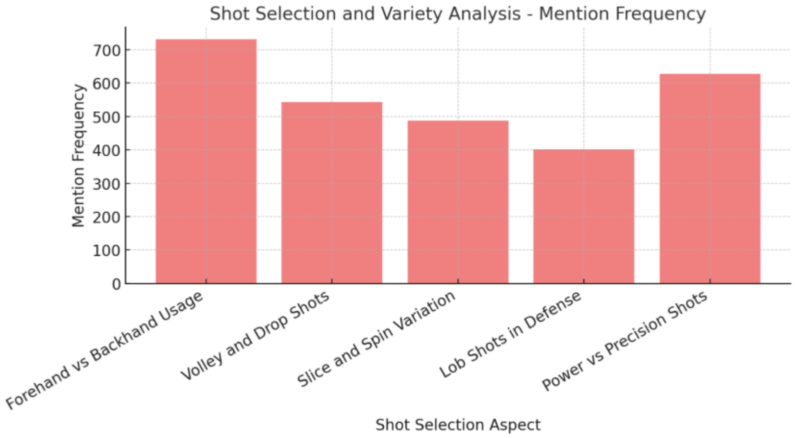

**Figure 7.** Shot selection and variety analysis.

**Table 10.** Shot selection and variety analysis.

| Shot Selection Aspect | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|
| Forehand vs Backhand Usage | 732 | 25.3% |
| Volley and Drop Shots | 543 | 18.8% |
| Slice and Spin Variation | 487 | 16.8% |
| Lob Shots in Defense | 402 | 13.9% |
| Power vs Precision Shots | 628 | 21.7% |

### 3.3. Sentiment and public perception

Based on **Table 11**, positive sentiment dominates the discussion surrounding player performance, accounting for 30.2% of the mentions. This suggests that fans and analysts generally express favorable views of players' efforts and achievements during matches. Neutral sentiment follows closely at 23.5%, indicating a significant number of objective or analytical observations rather than emotional responses. Negative sentiment makes up 18.7%, reflecting criticisms or frustrations with specific aspects of a player's game, possibly related to underperformance or tactical errors. Additionally, emotions of excitement are mentioned 12.9% of the time, typically during key moments of the match where players excel or make spectacular plays, while frustration accounts for 11.1%, often associated with missed opportunities or costly mistakes.

**Table 11.** Sentiment toward player performance analysis.

| Sentiment Type | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|
| Positive Sentiment | 1328 | 30.2% |
| Negative Sentiment | 824 | 18.7% |
| Neutral Sentiment | 1034 | 23.5% |
| Excitement | 569 | 12.9% |
| Frustration | 487 | 11.1% |

**Table 12** shows that excitement is the most frequently expressed emotion, with 27.6% of the occurrences being. This reflects the thrilling nature of tennis matches, especially during intense rallies or unexpected turnarounds. Frustration is the next most common emotion at 21.9%, often arising when players struggle or fail to meet expectations. Admiration, making up 17.3% of the mentions, highlights fans' respect and awe for the skill and athleticism of top players. Surprise at 15.1% represents reactions to unexpected outcomes or performances, such as upsets or tactical shifts during critical moments in a match. Lastly, at 18.1%, disappointment reflects the emotions tied to missed opportunities by individual players and overall match outcomes.

**Table 12.** Emotion tagging analysis.

| Emotion Type | Emotion Count | Percentage of Total Occurrences (%) |
|---|---|---|
| Excitement | 689 | 27.6% |
| Frustration | 547 | 21.9% |
| Admiration | 432 | 17.3% |
| Surprise | 376 | 15.1% |
| Disappointment | 458 | 18.1% |

## 3.4. Temporal patterns in tactical shifts

**Table 13** shows that early set aggression is a prominent tactical pattern, accounting for 23.1% of the total occurrences. This suggests that players start matches aggressively to establish dominance and set the pace early. Breakpoint strategy is another significant tactical shift, with 22.2% of mentions indicating that players often employ specific tactics when facing break points, as these moments can determine the outcome of a set or match. Mid-set adjustments become essential as matches progress, with 20.4% of occurrences, as players tweak their strategies based on the opponent's performance and match dynamics. Final set tactical changes are seen in 18.0% of discussions, highlighting the importance of making critical adjustments in the match's final stages. With 16.6% of mentions, late-set defensive play suggests that players often switch to a more conservative approach to protect a lead or avoid making mistakes under pressure.

**Table 13.** Match phase analysis.

| Match Phase Aspect | Occurrence Count | Percentage of Total Occurrences (%) |
|---|---|---|
| Early Set Aggression | 542 | 23.1% |
| Mid-Set Adjustments | 478 | 20.4% |
| Late Set Defensive Play | 389 | 16.6% |
| Final Set Tactical Changes | 423 | 18.0% |
| Break Point Strategy | 521 | 22.2% |

In **Table 14**, defensive play after winning a break is the most frequent tactical shift, accounting for 24.5% of occurrences. This indicates that players often switch to a more defensive style after securing a break to maintain their advantage. Increased risk-taking during tie-breaks, with 22.6% of mentions, highlights how players are willing to take more calculated risks in high-stakes situations to secure a crucial win. Aggression after losing a game is mentioned 21.4% of the time, suggesting that players frequently respond to setbacks by becoming more aggressive in their gameplay. Strategic slowdowns to regain composure occur in 19.3% of the cases, showing that players sometimes deliberately slow down the match's pace to reset their strategy and composure. Lastly, switches to net play in decisive points are seen in 16.4% of the discussions, indicating that players use this tactic to finish points quickly and decisively.

**Table 14.** Momentum shifts analysis.

| Tactical Shift Aspect | Occurrence Count | Percentage of Total Occurrences (%) |
|---|---|---|
| Aggression After Losing a Game | 467 | 21.4% |
| Defensive Play After Winning a Break | 534 | 24.5% |
| Increased Risk-Taking During Tie-Breaks | 492 | 22.6% |
| Switch to Net Play in Decisive Points | 358 | 16.4% |
| Strategic Slowdown to Regain Composure | 421 | 19.3% |

## 3.5. Frequent keywords and hashtags

As seen in **Table 15**, the hashtag #Tennis is the most frequently used, making up 15.8% of the total hashtags. This suggests that the general term "tennis" dominates discussions, broadly categorizing tweets about matches, players, and performances. More specific biomechanical and tactical terms are also prominently featured, such as #Footwork (10.9%) and #ServePlacement (9.3%), highlighting the focus on players' movement and serve tactics. #NetPlay and #BaselineStrategy are also key hashtags, with 7.9% and 8.7% of the total mentions, reflecting the tactical variations between net approaches and baseline control. Terms such as #Backhand (7.0%), #Forehand (7.5%), and aspects related to player physicality like #Speed, #Agility, and #Endurance, each making up around 6.0% to 6.4%, indicate the continued emphasis on specific technical skills.

**Table 15.** Hashtag analysis.

| Hashtag | Occurrence Count | Percentage of Total Hashtags (%) |
|---|---|---|
| #Tennis | 1260 | 15.8% |
| #Footwork | 872 | 10.9% |
| #ServePlacement | 745 | 9.3% |
| #NetPlay | 632 | 7.9% |
| #BaselineStrategy | 689 | 8.7% |
| #Backhand | 564 | 7.0% |
| #Forehand | 598 | 7.5% |
| #Speed | 512 | 6.4% |
| #Agility | 478 | 6.0% |
| #Endurance | 489 | 6.1% |

In **Table 16**, Novak Djokovic is the most frequently mentioned player, accounting for 16.2% of the total player-specific discussions, primarily due to his excellence in return of serve, endurance, and defensive play. Rafael Nadal follows with 15.4% mentions, driven by his aggressive topspin, forehand dominance, and baseline control, which have been hallmarks of his clay-court dominance. Roger Federer is also widely discussed (14.6%) for his serve placement, net play, and shot precision, reflecting his all-court mastery, particularly on grass. Serena Williams (13.4%) is recognized for her powerful serve, strong forehand, and court coverage, emphasizing her aggressive play style. Other players like Naomi Osaka, Ashleigh Barty, and Stefanos Tsitsipas also feature prominently in discussions, with specific mentions of their key tactics, such as baseline play for Osaka, slice variation for Barty, and serve and volley for Tsitsipas.

**Table 16.** Player-specific tactics analysis.

| Player Name | Key Tactical Terms | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|---|
| Rafael Nadal | Aggressive Topspin; Forehand Dominance; Baseline Control | 678 | 15.4% |
| Roger Federer | Serve Placement; Net Play; Shot Precision | 642 | 14.6% |
| Serena Williams | Powerful Serve; Strong Forehand; Court Coverage | 589 | 13.4% |
| Novak Djokovic | Return of Serve; Endurance; Defensive Play | 711 | 16.2% |
| Naomi Osaka | Baseline Play; Defensive Agility; Power Shots | 455 | 10.4% |
| Ashleigh Barty | All-court game; Slice Variation; Net Play | 398 | 9.1% |
| Stefanos Tsitsipas | One-Handed Backhand; Serve and Volley; Forehand Dominance | 472 | 10.8% |
| Simona Halep | Aggressive Baseline Play; Counterpunching; Defensive Footwork | 321 | 7.3% |
| Dominic Thiem | Heavy Topspin; Baseline Play; Physical Endurance | 287 | 6.6% |
| Daniil Medvedev | Flat Groundstrokes; Tactical Serve; Baseline Control | 359 | 8.2% |

In **Table 17**, Wimbledon is the most frequently discussed tournament, making up 26.3% of the mentions. This reflects the unique tactical focus required for grass-court play, quick points, and serve dominance, which defines the playing style at

Wimbledon. The French Open (Roland Garros) follows closely at 24.0%, with discussions centered on clay court dominance, topspin, and endurance, as players must adapt to slower courts and longer rallies. The US Open (22.7%) and Australian Open (22.1%) are also widely mentioned, with the US Open focusing on hard court tactics, powerful groundstrokes, and return of serve, while the Australian Open is noted for its fast courts, serve and volley strategies, and aggressive baseline play. Each tournament's specific playing conditions drive the tactical adaptations of the players, as reflected in the discussions.

**Table 17.** Tournament-specific tactics analysis.

| Tournament Name | Key Tactical Strategies | Mention Frequency | Percentage of Total Mentions (%) |
|---|---|---|---|
| Australian Open | Fast Courts, Strong Serve and Volley, Aggressive Baseline Play | 768 | 22.1% |
| French Open (Roland Garros) | Clay Court Dominance, Topspin, Endurance, Court Coverage | 834 | 24.0% |
| Wimbledon | Grass Court Play, Net Play, Quick Points, Serve Dominance | 912 | 26.3% |
| US Open | Hard Court, Powerful Groundstrokes, Return of Serve, Strategic Play | 789 | 22.7% |

## 3.6. Model performance metrics

**Table 18.** BERT with Bi-LSTM compared to other models.

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| BERT with Bi-LSTM | 88.5 | 87.9 | 88.0 | 87.95 |
| BERT with CNN | 85.7 | 84.3 | 85.1 | 84.70 |
| GloVe with LSTM | 82.3 | 81.5 | 81.9 | 81.70 |
| FastText with Bi-LSTM | 84.5 | 83.8 | 84.0 | 83.90 |
| Word2Vec with GRU | 81.9 | 80.7 | 81.3 | 81.00 |

**Table 18** and **Figure 8** show that the BERT with Bi-LSTM model outperforms other models across all key metrics. With an accuracy of 88.5%, it demonstrates superior performance in identifying biomechanical patterns and tactical strategies from tweets. Its precision is 87.9%, indicating a high degree of correctness in the model's predictions, while its recall stands at 88.0%, showing its ability to identify relevant instances accurately. The F1-score of 87.95% reflects the balance between precision and recall, confirming that the model is highly effective in analyzing social media data for tennis performance insights. When compared to the BERT with CNN model, which has an accuracy of 85.7%, the BERT with Bi-LSTM model shows a noticeable improvement, particularly in its ability to capture sequential dependencies in text, as indicated by the higher F1-score of 87.95% compared to 84.70% for BERT with CNN. Similarly, models like GloVe with LSTM and FastText with Bi-LSTM achieve lower performance, with accuracies of 82.3% and 84.5%, respectively, highlighting the strength of using BERT embeddings combined with Bi-LSTM's sequential analysis capabilities. The Word2Vec with GRU model has the lowest performance, with an accuracy of 81.9%, precision of 80.7%, and F1-score of 81.00%,

further demonstrating that the combination of BERT and Bi-LSTM provides a more practical approach for analyzing tennis-related tweets, especially in understanding context and sequences in textual data.

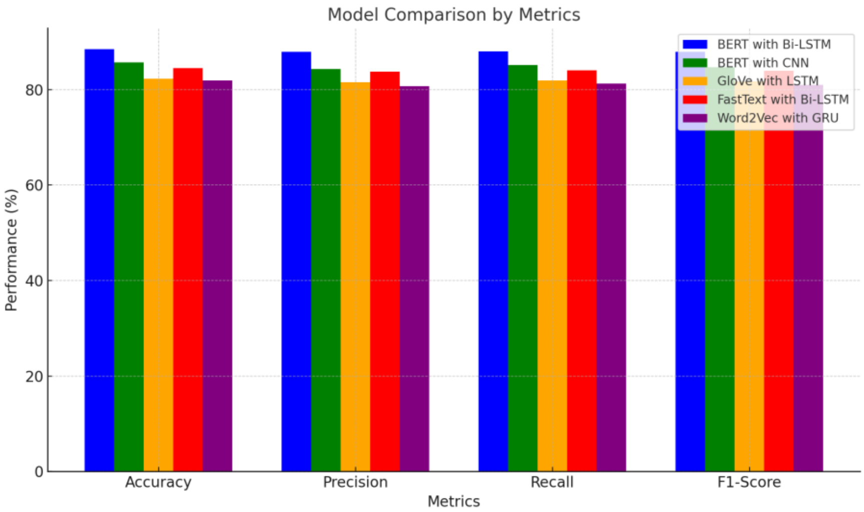

**Figure 8.** BERT with Bi-LSTM compared to other models.

## 4. Conclusion and future work

In this study, as discussed on Twitter, we successfully applied social media data mining techniques to analyze tennis players' biomechanical patterns and tactical strategies. By utilizing BERT for contextual embedding and Bi-LSTM for sequential analysis, our model was able to identify key performance metrics and tactical adjustments in real-time tennis discussions. The results indicate that this approach provides deeper insights into player performance, particularly footwork, serve placement, endurance, and strategic decision-making during matches. Additionally, we identified differences in tactics across tournaments and players, revealing how playing surfaces and match contexts influence performance. Our model outperformed several other NLP models, showcasing its effectiveness in extracting relevant sports analytics information from unstructured social media data. This research contributes to the growing field of sports analytics and offers a framework for applying techniques similar to those used in other sports domains.

Future work could focus on expanding the dataset to include a broader range of tournaments and player profiles, as well as integrating video analysis further to enhance the understanding of tactical shifts and biomechanical performance.

**Author contributions:** Conceptualization, HY and XW; methodology, HY; software, XW; validation, HY and XW; formal analysis, HY; investigation, HY; resources, XW; data curation, XW; writing—original draft preparation, writing—review and editing, HY and XW; visualization, XW; supervision, XW; project administration, HY; funding acquisition, XW. All authors have read and agreed to the published version of the manuscript.

**Ethical approval:** Not Applicable.

**Conflict of interest:** The authors declare no conflict of interest.

## References

1. Muhsal; F.; Jaitner; D.; & John; J. (2023). # picturesofchange: Physical self-representations in social media as a sign of change in sports-and movement culture: An integrative review with educational implications. Current Issues in Sport Science (CISS); 8(3); 006-006.
2. Hachtmann; F. (2022). Emerging trends in computer-mediated communication and social media in sport: Theory and practice. The emerald handbook of computer-mediated communication and social media; 269-284.
3. Weimar; D.; Soebbing; B. P.; & Wicker; P. (2021). Dealing with statistical significance in big data: The social media value of game outcomes in professional football. Journal of Sport Management; 35(3); 266-277.
4. Pather; S. (2021). The impact of digital media platforms on sports reporting and audience engagement: A case study of Twitter in South Africa. University of Johannesburg (South Africa).
5. Crespo; M.; Martínez-Gallego; R.; & Filipcic; A. (2024). Determining the tactical and technical level of competitive tennis players using a competency model: a systematic review. Frontiers in Sports and Active Living; 6; 1406846.
6. Sampaio; T.; Oliveira; J. P.; Marinho; D. A.; Neiva; H. P.; & Morais; J. E. (2024). Applications of Machine Learning to Optimize Tennis Performance: A Systematic Review. Applied Sciences; 14(13); 5517.
7. Fett; J.; Oberschelp; N.; Vuong; J. L.; Wiewelhove; T.; & Ferrauti; A. (2021). Kinematic characteristics of the tennis serve from the ad and deuce court service positions in elite junior players. PLoS One; 16(7); e0252650.
8. Carboch; J.; Brenton; J.; Reischlova; E.; & Kocib; T. (2023). Anticipatory information sources of serve and returning of elite professional tennis players: A qualitative approach. International Journal of Sports Science & Coaching; 18(3); 761-771.
9. Olivetti; E. A.; Cole; J. M.; Kim; E.; Kononova; O.; Ceder; G.; Han; T. Y. J.; & Hiszpanski; A. M. (2020). Data-driven materials research enabled by natural language processing and information extraction. Applied Physics Reviews; 7(4).
10. Lin; J.; Nogueira; R.; & Yates; A. (2022). Pretrained transformers for text ranking: Bert and beyond. Springer Nature.
11. Koroteev; M. V. (2021). BERT: a review of applications in natural language processing and understanding. arXiv preprint arXiv:2103.11943.
12. Sun; Y.; & Platoš; J. (2023). Attention-based Stacked Bidirectional Long Short-term Memory Model for Word Sense Disambiguation. ACM Transactions on Asian and Low-Resource Language Information Processing.
13. Zeberga; K.; Attique; M.; Shah; B.; Ali; F.; Jembre; Y. Z.; & Chung; T. S. (2022). [Retracted] A Novel Text Mining Approach for Mental
14. Kolman; N. (2023). Unravelling tennis performance: creating monitoring tools to measure and understand technical and tactical skills.
15. Srihi; S.; Jouira; G.; Ben Waer; F.; Rebai; H.; Majdoub; A.; & Sahli; S. (2022). Postural balance in young tennis players of varied competition levels. Perceptual and Motor Skills; 129(5); 1599-1613.
16. Li; X. (2022). Biomechanical analysis of different footwork foot movements in table tennis. Computational Intelligence and Neuroscience; 2022(1); 9684535.
17. Bergström; L. (2020). “Play ball!”: A Study of Speech Variations and Characteristics of UK Sports Commentary.
18. Norris; L. A.; Didymus; F. F.; & Kaiseler; M. (2020). Understanding social networks and social support resources with sports coaches. Psychology of Sport and Exercise; 48; 101665.
19. Martin; D.; O Donoghue; P. G.; Bradley; J.; & McGrath; D. (2021). Developing a framework for professional practice in applied performance analysis. International Journal of Performance Analysis in Sport; 21(6); 845-888.
20. Indumathi N et al., Impact of Fireworks Industry Safety Measures and Prevention Management System on Human Error Mitigation Using a Machine Learning Approach, Sensors, 2023, 23 (9), 4365; DOI:10.3390/s23094365.
21. Parkavi K et al., Effective Scheduling of Multi-Load Automated Guided Vehicle in Spinning Mill: A Case Study, IEEE Access, 2023, DOI:10.1109/ACCESS.2023.3236843.
22. Ran Q et al., English language teaching based on big data analytics in augmentative and alternative communication system, Springer-International Journal of Speech Technology, 2022, DOI:10.1007/s10772-022-09960-1.
23. Ngangbam PS et al., Investigation on characteristics of Monte Carlo model of single electron transistor using Orthodox Theory, Elsevier, Sustainable Energy Technologies and Assessments, Vol. 48, 2021, 101601, DOI:10.1016/j.seta.2021.101601.
24. Huidan Huang et al., Emotional intelligence for board capital on technological innovation performance of high-tech enterprises, Elsevier, Aggression and Violent Behavior, 2021, 101633, DOI:10.1016/j.avb.2021.101633.

25. Sudhakar S, et al., Cost-effective and efficient 3D human model creation and re-identification application for human digital twins, Multimedia Tools and Applications, 2021. DOI:10.1007/s11042-021-10842-y.
26. Prabhakaran N et al., Novel Collision Detection and Avoidance System for Mid-vehicle Using Offset-Based Curvilinear Motion. Wireless Personal Communication, 2021. DOI:10.1007/s11277-021-08333-2.
27. Balajee A et al., Modeling and multi-class classification of vibroarthographic signals via time domain curvilinear divergence random forest, J Ambient Intell Human Comput, 2021, DOI:10.1007/s12652-020-02869-0.
28. Omnia SN et al., An educational tool for enhanced mobile e-Learning for technical higher education using mobile devices for augmented reality, Microprocessors and Microsystems, 83, 2021, 104030, DOI:10.1016/j.micpro.2021.104030 .
29. Firas TA et al., Strategizing Low-Carbon Urban Planning through Environmental Impact Assessment by Artificial Intelligence-Driven Carbon Foot Print Forecasting, Journal of Machine and Computing, 4(4), 2024, doi: 10.53759/7669/jmc202404105.
30. Shaymaa HN, et al., Genetic Algorithms for Optimized Selection of Biodegradable Polymers in Sustainable Manufacturing Processes, Journal of Machine and Computing, 4(3), 563-574, https://doi.org/10.53759/7669/jmc202404054.
31. Hayder MAG et al., An open-source MP + CNN + BiLSTM model-based hybrid model for recognizing sign language on smartphones. Int J Syst Assur Eng Manag (2024). https://doi.org/10.1007/s13198-024-02376-x
32. Bhavana Raj K et al., Equipment Planning for an Automated Production Line Using a Cloud System, Innovations in Computer Science and Engineering. ICICSE 2022. Lecture Notes in Networks and Systems, 565, 707–717, Springer, Singapore. DOI:10.1007/978-981-19-7455-7_57.