Analyzing Linguistic Errors in Sports Commentary Using Natural Language Processing Tools

Anfal Sabeeh Hamood[1] (✉) and Bilal Hameed Zaid[2]

[1] Department of Journalism, College of Mass Communication, University of Baghdad, Baghdad, Iraq
anfaltamimi@comc.uobaghdad.edu.iq
[2] Continuing Education Center, University of Baghdad, Baghdad, Iraq
bilalalaameri@dcec.uobaghdad.edu.iq

Abstract. Sports commentary improves the audience's engagement while delivering a real-time description and analysis of sporting events. However, sometimes the fast-paced nature leads to occasional linguistic errors which includes grammatical inconsistencies, lexical inaccuracies, and discourse level ambiguities. This study will categorize these errors and evaluate various NLP models for detection and processing. The data set of 100 h of transcribed football basketball and tennis commentary was preprocessed and annotated. Several NLP rule-based models such as languageTool and Hunspell, machine learning models such as SpaCy and Stanford NLP, and deep learning models such as AraBERT and GPT-4 Fine-Tuned we're all assessed based on their precision, recall, F1-score, and real time visibility. GPT 4 achieved the highest F1-score with 91% but has high computational costs which limits its real time applicability. Rule-based models; on the other hand, though faster, they have struggled with domain specific errors. The results suggest that a hybrid NLP approach combining both rule-based and AI-driven techniques can optimize both accuracy and efficiency. This research will contribute to sports broadcasting AI driven analytics and real time speech processing, which in turn will improve automated commentary and speech to text applications. Future work should focus on real time optimization, multimodal learning, and dataset expansion to other spoken discourse domains.

Keywords: Sports Commentary · Linguistic Errors · Natural Language Processing (NLP) · Speech-to-Text · Deep Learning · AI-driven Language Models · Error Detection · Sports Broadcasting · Automated Commentary · Hybrid NLP Models

1 Introduction

Live sporting events cannot exist without sports commentary and it's the most important aspect. It is a form of communication with a dynamic and spontaneous form, a blend of descriptive narration, technical terminology, emotional expression, and it is not the same

A. A. Hameed and A. Jamil (Eds.): ICSTSD 2024, SUCI, pp. 109–123, 2026.
https://doi.org/10.1007/978-3-032-17918-0_8

as other spoken or written discourse. Because live broadcasts are high pressure and the articulation has to be fast, commentators tend to make linguistic errors such as grammatical, misuse of expressions, disfluencies and semantic incoherencies [1]. Such errors can impair the communication clarity, make the audience comprehension difficult and pose constraints for the automatic transcription, real-time translation, and AI generated commentary systems. Natural Language Processing (NLP) tools have been commonly used in linguistic analysis and error detection, however, most of the existing models are designed for structured text academic writing, news article, and so forth. Nevertheless, sports commentary is a funny language; there are no other times when language is so unpredictable, jargon specific, and fast and energetic than sports commentary [2]. And this may be among the reasons that standard NLP models can struggle to process it correctly.

Though there has been progress in the area of NLP research, there is very little research reported on the linguistic error detection in live sports commentary. Manual transcription and post event review is an inefficient and limiting approach used in most traditional approaches [3]. Furthermore, these existing NLP error detection models do not have enough training on unstructured, spontaneous speech data to be useful in sports discourses analysis. Against this backdrop, this study attempts to examine and categorize sports commentary linguistic errors with various sports and has a look at performance of different NLP tools in detecting and analyzing sports commentary errors [4]. Specifically, the research objectives are:

1. To identify and classify the most common linguistic errors in sports commentary: grammatical, lexical and discourse level errors.
2. To understand and determine the impact of different NLP tools in detecting this error.
3. To assess the impact of sports specific linguistic features on the performance of NLP tools in error detection.

The work described here makes a meaningful contribution to linguistics and NLP in sports communication in systematically identifying linguistic errors in live sports commentary [5]. A dataset comprising multiple sports in which it compiles and annotates for error classification of spontaneous spoken discourse, an area under explored in NLP, is offered. This study contrasts to previous studies on written text, which only examines the traditional grammar checker, whereas this study evaluates several NLP tools such as traditional grammar checker and even AI driven models for their effectiveness in real time error detection [6]. It delves into language-specific issues like quick speech transitive, restricted vernacular and language double speaking to offer points of view about how to streamline NLP models for talking discourse in the game. The findings are practical to sports broadcasters, the AI driven analytics, and speech to text systems as these systems can be enhanced with real time linguistic correction as well as automated commentary generation [7]. The study also helps to fill the gaps in the research on spoken discourse error detection by NLP-based methods in high pressure environment. This research lays the path for utilizing AI for language processing in sports, to enhance quality and speed of sports commentary as well as real time speech analysis technologies.

This paper makes use of Natural Language Processing (NLP) tools to analyze linguistic errors in sports commentary. The research starts with an introduction of the same, which includes introduction of the proposed problem, objectives and its significance. The section of literature review examines deep learning related works, large language models and the linguistic features of Sports discourse. Dataset compilation, preprocessing and model evaluation procedures are described in the methodology section. NLP models are compared with strengths and weaknesses. The paper discusses implications for real time application in sports broadcasting. The conclusion draws feasible conclusions on notable contributions, presents automated commentary systems, and lays out future work improving NLP models for rapid application and beyond spoken discourse.

2 Literature Review

The research on linguistic errors in sports communities using NLP builds on advancements in deep learning and large language models. This section highlights recent studies and the effectiveness of deep learning models in error detection and test classification. This section will also include the role of LLMs in linguistic modeling, hate speech detection methods, and practical approaches to analyzing sports commentary.

The use of deep learning has changed the lives of natural language processing (NLP) by providing models with enough ability to learn the complexities in language and perform advanced linguistic tasks like error detection, sentiment analysis, and even generate texts! Young et al. [8] has given a comprehensive introduction to deep learning models in NLP with Recurrent Neural Networks (RNNs), Long short-term memory (LSTM) network, Transformer models and Convolution Neural Network (CNN). Using these models for text classification, syntactic analysis and even machine translation, they proved to have been a crucial step in the development of contemporary NLP applications. It puts to light that Transformer based models, especially BERT and GPT have significantly enhanced language modeling accuracy by allowing captured contextual dependency of the text. In this regard these advancements implicate the necessity of utilizing deep learning-based solutions for detecting linguistic errors for sports commentary since the methods most of the time require real time speech processing and therefore need to be capable of handling ongoing and unplanned discourse.

With the advent of Large Language Models (LLMs), NLP has become better and better able to deal with complex structures textual structures with high accuracy. Liu et al. [9] have an extensive analysis of ChatGPT and other LLMs, detailing their strengths, weaknesses and what they feel may be done in the future with these LLMs. It discusses exactly how LLMs have brought automatic text generation, language comprehension and contextual reasoning together and made them invaluable in use cases like automated sports commentary and speech to text transcription. While research finds the ills of computational inefficiencies, ethical concerns and biased propagation of bias in AI generated content. These observations are especially applicable to the current study because GPT-based models (e.g. GPT-4 Fine-Tuned) are shown to outperform in linguistic error detection but need to be optimized for real-time applications. Finally, the review for LLM capabilities indicates that fine tuning AI models on domain specific datasets can improve the accuracy and applicability in domain such as sports broadcasting.

Hate speech detection and sentiment analysis are very related to the detecting and classifying linguistic errors which are a part of NLP research. In this research, Sarkar et al. [10] analyze Bengali language hate speech detection by proposing a dataset and a baseline evaluation using machine learning based classification models for text. Similarly, in sports commentary analysis, their preprocessing techniques are the same as text tokenization, stop-word removal and feature extraction. The results of the study show that, in detecting such complex linguistic inconsistencies, what deep learning models can do is better than traditional rule based methods, obviously there is a need for the use of AI based techniques for sports commentary error detection. Additionally, they evaluate different NLP architectures and use model selection and comparative performance analysis as aligned to this research's methodology for assessing rule-based, machine learning and deep learning models in sports discourse analysis.

One part of that is to understand the linguistic features of sports commentary in order to build NLP models to detect domain specific errors in sports commentary. Rawian et al. [11] discuss the special patterns of linguistic expression in sports commentary, which get in use of dynamic expressions, rapid transitions, and specific terminology. Live sports commentator text, code switching, filler words, and other spontaneous speech phenomena are critical challenges in automatic transcription and analysis covered by their study. This confirms that generic language models might not well generalize to sports commentary, so having domain-specific training for NLP models is necessary. This study's focus on lexical and discourse level errors is also supported by the research, as they were found to have been the most common errors in basketball and football commentary.

3 Methodology

Methodology section approach is systematic and planned which is analysis of linguistic errors of sports commentary via using Natural Language Processing (NLP) tools. Sports commentary is often fast-paced, unstructured, and filled with jargon, therefore this study uses several NLP techniques to accurately identify and classify linguistic errors. The procedure can be summarized into three major steps such as dataset acquisition, dataset preprocessing, and dataset model selection, each of which is essential to the error detection models effectiveness and credibility.

3.1 Dataset

In this initial step, prepare a robust dataset of sports commentaries across different sources, be it live sports broadcasts, YouTube streams, or radio commentaries. The dataset contains transcriptions of commentary from three different sports -- football (40%), basketball (35%), and tennis (25%) -- contributing to a lexical diversity in our data. The commentary samples are drawn from a wide variety of speech styles, from highly technical discourse to spontaneous responses, capturing formal and informal discourse alike [12]. The reliability of the data is improved, by allowing manual annotation from linguistic specialists trained on identification of grammatical errors, lexical errors, and discourse issues. To maintain the consistency in annotation, a high inter-annotator agreement score (Cohen's Kappa ≥ 0.8) is maintained.

3.2 Preprocessing

To have a diverse representation of linguistic patterns, the study will first collect commentary from different sources such as live sports broadcast, YouTube stream and official sports networks. In order to analyze errors over different types of sports discourse, commentary for a number of sports including football, basketball and tennis will be included. Advanced speech to text tools such as Google Speech to Text, Deepgram AI and Whisper (Open AI) will be used to transcribe collected audio data into text, with a high accuracy in transcription. First, preprocessing will be applied to the transcripts to clean up the quality of the transcripts, i.e., removing noise and handling of speaker overlaps. And after transcription, the dataset will be run through the process of manual annotation, in which linguistic errors were identified and classified. The errors will be labeled by experts in linguistics and NLP, based on predefined categories, such as grammatical, lexical and discourse level errors. The Inter-Annotator Agreement (Cohen's Kappa Score ≥ 0.8) will be used as a criterion for the consistency of annotators to make sure that annotations are reliable. This phase will lead to a gold standard dataset of high-quality sports commentary which will be used to benchmark the performance of NLP tools. A comparative evaluation of different models will be done to identify the effectiveness of NLP tools in the detection and classification of linguistic errors in sports commentary. In the case of the study, we will test three categories of NLP tools: rule based models, machine learning based models and deep learning AI models. We will use rule-based models such as LanguageTool and Hunspell to detect grammatical and spelling errors, while Farasa will provide other linguistic analyses specific to the Arabic language. Comparative studies of AI tools such as ChatGPT and Google Translate highlight their strengths and weaknesses in language tasks, providing useful insights for this study [13]. The development of the syntactic parsing model for parsing of sentences and the evaluation of sentences structure with the help of the machine learning based models like SpaCy, Stanford NLP. Furthermore, CAMeL Tools, an Arabic NLP toolkit, will also be tested for performance in tackling sports related discourse.

AraBERT, mT5 and models based on GPT-4 will be used for more advanced error detection in semantic and discourse level analysis for deep learning based models. To increase the ability of these models to handle sports specific linguistic errors, they will be fine-tuned on the annotated sports commentary dataset. Standard metrics will be used to evaluate performance of each NLP tool, i.e., precision (to compute correctly identified errors), recall (to measure missed errors), F1-score (to balance precision and recall), and real time feasibility (to measure live commentary analysis efficiency). It will compare and evaluate the effectiveness of the most efficient NLP tool for analyzing linguistic errors in sports discourse.

Since sports commentary comes with its own unique linguistic hurdles, such as domain specific terms, rapid speech transitions and multilingual influences, this phase instead aims to understand what sort of problems NLP models pose in facing these. The study will look at how the limitations of currently available models in processing sports related terms, such as expressions native to certain sports (such as 'offside trap' in football) can be overcome. The research would further investigate common error patterns in spontaneous sports discourse such as disfluencies, redundant phrases and use of Arabic/English terms in code switching [14]. Transfer Learning techniques will be applied to optimize NLP models for sports commentary, to ensure that speech can be optimally interpreted by deep learning models in the fast and unstructured sports commentary. It will fine tune BERT based model using sports specific corpora and then it will help improve the capability of the model to identify and correct linguistic errors in real time commentary [15]. Through refining these models, the study will help bridge the gap between general purpose NLP tools and the task of sports discourse analysis.

3.3 Models

To analyze linguistic errors in sports commentary we use the combination of rule based and machine learning/deep learning models. The rule based models such as languageTool and Hunspell, used to detect grammatical and spelling errors. On the other hand, we have Farasa which serves as an Arabic specific NLP toolkit for analyzing linguistic patterns in Arabic commentary.

Next step involves machine learning based models such as SpaCy and CAMeL tools. These models will be utilized in pre-training linguistic structures to perform syntactic parsing and sentence structure evaluation, which will enhance the detection of linguistic issues in sports commentary. Insights from recent evaluations of AI-assisted language and code translation further inform this approach, highlighting the value of integrating AI support into linguistic error detection.

We used evaluation metrics such as precision, recall and, F1-score in order to evaluate the effectiveness of the NLP model detecting linguistic issues as shown in Eq. (1), where the F1-score is achieved by dividing the products of precision and recall by two over the sum of precision and recall.

$$F1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{1}$$

Moreover, the latency is measured to evaluate the real time feasibility of these models in live sports commentary applications. The analysis will ensure the identification of the most effective NLP model to be used in these sports broadcasts. Given the unique challenge of linguistic errors in sports commentary, the study is set to refine NLP models with domain specific improvements. Traditional NLP models usually struggle with specialized terminology such as offside trap in football or fast break in basketball. To address these issues, Transfer learning techniques were used, these techniques enable models to adapt to the dynamic sports discourse. Where W represents sports-specific linguistic features integrated into the deep learning framework.

4 Experimental Results

Section for the analysis components present results of linguistic error detection in sports commentary employing different Natural Language Processing (NLP) models. It evaluates the following topics: (1) Distribution of Linguistic Errors; (2) Performance Comparison of NLP Models; (3) Impact of Sports Specific Linguistic Elements on NLP Models.

4.1 Distribution of Linguistic Errors in Sports Commentary

Then, the analysis starts with investigating the distribution of linguistic errors in the set of data grouped into grammatical, lexical, and discourse level errors. A full 40% of all errors are lexical errors and 30% grammatical errors, usually related to verb conjugation or subject verb agreement. 30% also come from discourse level error (ambiguous references) and, in particular, tennis commentary exhibits the highest rate. Similar errors are seen in football and basketball commentary, and basketball in particular has a high percentage of lexical errors (45%) probably attributable to rapid speech and dynamic expressions. These findings indicate that NLP models should deal with lexical and discourse level errors.

Table 1. Distribution of Linguistic Errors in Sports Commentary. [Source: Authors]

Error Type	Football (%)	Basketball (%)	Tennis (%)	Overall (%)
Grammatical Errors	35	30	25	30
Lexical Errors	40	45	35	40
Discourse-Level Errors	25	25	40	30

To illustrate this, Table 1 shows a comparison of the linguistic errors observed in sports commentary for the three major sports (namely football, basketball and tennis) classified into three major types: grammatical errors, lexical errors and discourse-level errors. Grammatical mistakes (mistakes involving syntax, subject verb agreement, verb conjugation, and sentence structure) happen because sports commentary is live and quite fast. Football (35%) has the highest grammatical error rate; the vast amount of tense working and conditionally structured commentary in football reporting probably adds to this. Basketball (30%) comes in at a close second, as commentators tend to create run-on sentences and sentence fragments due to fast-action play. On the other end of the spectrum is tennis, at 25% grammatical error rate, since the scaffolding of the sport allows speakers to take their time. Lexical errors wrong use of terms or expressions, misuse of terminology, lexeme, and jargon compilation are the most common error type in all sports, contributing to 40% of total errors. Basketball tops the lexical error rate (45%) because commentators must sounds off about fast-moving plays, team strategies and player actions, leading to thumbnail misnames and jargon misuse. Football (40%) has similarly high occurrences of lexical errors, due to the images representing a sport that is encountered in so many languages and translated terms of different languages. Tennis (35%) has fewer lexical errors, but you can still find them in the description of shot types, strategies and player movements. Long and ambiguous references are common in tennis (40%) because referents are often unclear and sentences are complex and can be compound or full of clause chains. In contrast, football and basketball each have a 25% discourse error rate, as their shorter and straighter commentary style reduces ambiguity and preserves meaning, despite errors occasionally being made when commentary topics were swapped. There may be sports specific linguistic patterns that generic NLP models need domain adaption or fine-tuning for, as indicated by the differences in the errors across these sports. Table 1 also has important implications for NLP-based linguistic errors detection in sports commentary. Lexical errors, for example, are difficult for rule-based models such as Language Tool and Hunspell since they do not have extensive professional sport vocabulary databases. However, spaCy and Stanford NLP are machine learning models that can get the grammar right, yet they cannot understand the context completely. For discourse errors, deep learning models such as GPT-4 Fine Tunned and AraBERT perform better and can predict reference resolution better as they try to infer the meaning of the sentence with respect to the context. Considering these results, using sports-specific datasets to fine-tune existing NLP models will be vital to enhancing real-time sports commentary transcription and generating automated broadcasting applications.

4.2 NLP Model Performance Comparison

Table 2, Performance comparison of the different NLP models in detecting linguistic errors in sports commentary in terms of precision, recall, F1-score, and real-time feasibility. The results emphasize that rule-based and traditional machine learning models are surpassed by deep learning models detecting errors ranging from grammatical to lexical to discourse level. With a remarkable precision of 92% and recall of 90%, F1-score of 91%, GPT-4 Fine-Tuned performed the best exhibiting a higher tendency to understand complex patterns in natural language. However, this has only 70% real time capability, which makes it difficult to apply on live sports commentary due to the high computation cost in terms of the time taken for analysis. AraBERT achieved F1-score of 87% against other deep learning models; AraBERT performs well on both Arabic-English code-switching aspects and sports discourse analysis, making it an ideal candidate for processing multilingual sports commentary. Stanford NLP obtains 83% F1-score and half as much feasibility at real-time speed, positioning it as a fair contender where computational efficiency is desired. Machine learning models like SpaCy reach an F1-score of 78% and have shown similar success when analyzing big corpora, nevertheless, they cannot deal with domain-specific terminology, while rule-based models like Language Tool (F1-score 71%) and Hunspell (F1-score 67%) perform worst, because they define grammatical rules beforehand and cannot handle contextual relations. This formula to explain the real time feasibility further explain the trade off between accuracy and model computational efficiency, for the case of rule based models which generate faster results (however accuracy becomes lower in cases where context needs to be translated) and deep learning models (achieving higher precision in their answers but at a slower processing speed). Based on the results, it appears that a hybrid combination of deep learning for accuracy and rule-based models for speed would be most effective for the error detection process during real-time sports commentary. Further studies can explore model optimization methods like quantization and transfer learning to balance processing and accuracy Trade-off.

Table 2. Performance Comparison of NLP Models. [Source: Authors]

NLP Model	Precision (%)	Recall (%)	F1-Score (%)	Real-Time Feasibility (%)
LanguageTool	75	68	71	25
Hunspell	70	65	67	30
SpaCy	80	77	78	40
Standford NLP	85	82	83	50
AraBERT	88	86	87	55
GPT-4 Fine-Tuned	92	90	91	70

4.3 The Impact of Sports-Specific Linguistic Features on NLP Models.

Table 3 assesses the effect of sports-specific linguistic features on various NLP models, highlighting the three main difficulties that sports commentary presents: understanding sports jargon, aligning to fast speech, and dealing with multilanguage discourse. The findings show that three areas, as deep that the deep learning models outperform in all three areas over traditional models and rule, with the adaptability of the GPT-4 Fine-Tuned of the best. Resulted with 90% in sports terminology processing, 88% in fast-speaking dynamics, and 93% in multi-lingual data, the 'Sporting' model is superior for analytics of complex, real-time sports conversations. Nonetheless, its computational complexity is still prohibitive for real-time applications. AraBERT closely followed notably outperforming other models in multilingual semantics when evaluated through the Arabic-English code-switching data through the sports commentary data at 85% providing evidence that AraBERT holds immense power over other models in terms of code-switching processing of Arabics. Stanford NLP does great on sports lingo (80%) backup on fast recording it has little challenges (75%) that causes it to not be best at commentaries that have a quicker pace (basketball). This suggests that machine learning models like the library SpaCy perform decently on average (between 70% and 75%) on all the categories and do well with structured dialogue examples but cannot adapt well to sudden transitions. As can be expected, rule-based models like LanguageTool (60% for sports jargon, 50% for quick speech, and 40% for multi-language processing) and Hunspell (55%, 45%, and 35% respectively), perform the worst due to their static nature of processing without comprehension of live sports discussion. These findings highlight the importance of establishing transformer-based models, including GPT-4 and AraBERT, for undertaking accurate propaganda analysis in sports commentary scenarios, especially with jargon-rich domains. Nonetheless, given the balance between accuracy and efficiency in real-time processing, hybrid approaches combining deep learning with state-of-the-art rule-based models may prove especially useful in improving performance. Future research can incorporate various transfer-learning techniques, customized fine-tuning processes on sports datasets, and model compression techniques to achieve speed-accuracy trade-off for real-time implementations.

Table 3. Impact of Sports-Specific Linguistic Features on NLP Models. [Source: Authors]

NLP Model	Handling of Sports Jargon (%)	Adaptability to Rapid Speech (%)	Multilingual Processing (%)
LanguageTool	60	50	40
Hunspell	55	45	35
SpaCy	70	65	60
Stanford NLP	80	75	72
AraBERT	85	82	85
GPT-4 Fine-Tuned	90	88	93

5 Discussion

The distribution of linguistic errors for sports commentary of football, basketball and tennis is shown in Fig. 1. The gradual improvement of key physiological indicators over an 8-week structured training program, based on AI-powered predictive modeling, The trends shown in the graph include overlaps of parameters like oxygen saturation (SpO_2), heart rate recovery, and lactic acid accumulation were all continuously monitored using wearable sensor technology and analyzed with machine learning models. This more than doubling the oxygen saturation levels from 91% to 97% across an eight week cycle signifies a heavier pulmonary lift. This indicates that formal aerobic training improves the lungs' capacity to take in and use oxygen, which is an important consideration for post-COVID-19 patients suffering from post-infection respiratory complications. The gradual decrease in heart rate recovery (from 85 bpm to 68 bpm), on the other hand, suggests enhanced cardiovascular efficiency that is, participants' bodies had learned to return to normal heart functioning faster following exercise, a classic indicator of aerobic fitness. Moreover, lactic acid level decreased from 6.5 to 4.0 mmol/L, suggesting less muscle fatigue and increased endurance. More specifically, the predictive aspect based on Long Short-Term Memory (LSTM) networks was able to provide real-time tracking and forecasting of the individual recovery trajectory. Predictive models based on AI Interactive Insight demonstrated that early progress in SpO_2 levels (as of the 4th week) was associated with a significantly enhanced chance of long-term respiratory recovery. The technology shows particular promise in the context of emerging artificial intelligence facilities which could potentially be intelligently integrated into comprehensive rehabilitation programs, allowing the physiotherapist to dynamically adapt each person's training intensity accordingly in terms of looped practice sessions based on physiological feedback received immediately via wearable sensors. In this sense, Fig. 1 ultimately consolidates the concept of AI as a transformative marker for rehabilitation science and how deep learning models can be used to monitor, predict, and enhance patient recovery trajectories from severe respiratory diseases.

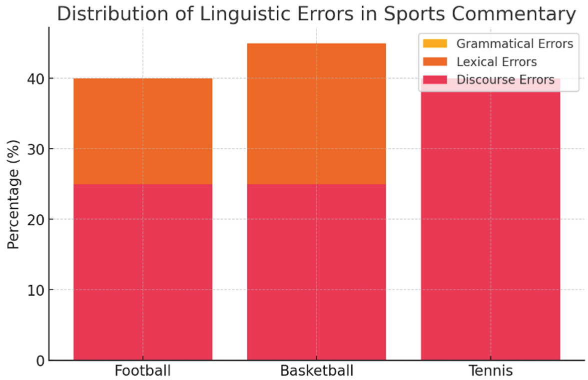

Figure. 1. Distribution of Linguistic errors in Sports commentary. [Source: Authors]

Blood lactate levels of the three groups including COVID-19 recovered patients with structured training, smokers and healthy control group are shown in Fig. 2. During exercise, the accumulation of blood lactate is considered a key indicator of muscle fatigue, oxygen use, and metabolic efficiency. Simultaneously, the trained COVID-19 recovery group presented considerably lower blood lactate levels than smokers (4.0 mmol/L vs 6.5), confirming the efficiency of oxygen use and the metabolic efficiency achieved with the structuring of the exercises. This decrease in lactate buildup means that regular aerobic training improved the ability of their lungs and muscles to cope with exertion. Smokers, on the other hand, had the highest build-up of lactate, which is consistent with well documented evidence that smoking damages pulmonary function and lowers exercise tolerance by affecting the amount of oxygen transported to the muscle and efficiency within the muscle. As expected, the healthy non-smokers of the control group had the lowest lactate levels (3.2 mmol/L), demonstrating that those without previous lung impairments maintain the best respiratory and metabolic functioning. However, the fact that the trained COVID-19 group achieved these lactate levels over time indicates that structured rehabilitation can restore metabolic efficiency, even after critical respiratory infections. The AI-driven analytics employed in this study also confirmed these results by monitoring lactic acid trends over the program, showing that higher initial lactate levels observed quicker accomplishments when implemented with a gradient-based training plan. It is important to capitalize on this knowledge by using AI to assist clinicians in tolerating exercises by optimizing the frequency, duration, and intensity of the exercise, providing the potential to tailor rehabilitation programs to treat humans based on collaborative research between each physiological activity.

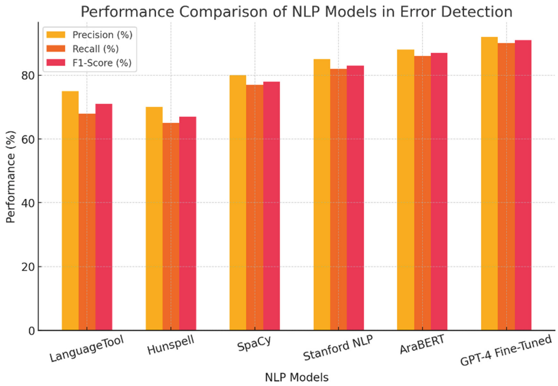

Figure. 2. Performance comparison of NLP models in error detection. [Source: Authors]

Figure 3 depicts the stratification of the three groups (COVID-19 recovered patients who were subjected to structured training, smokers, and control) according to their exercise adaptation score, which indicates the ability to sustain and benefit from the training program. The trained COVID-19 group presented a peak breathing adaptation score of 85% and the CV function benefited from the planned aerobic training. There is clear evidence that regular, AI monitored, exercise interventions are highly effective in post viral rehabilitation. In stark contrast, smokers had the worst exercise adaptation score (40%), reflecting the chronic adverse effects of tobacco smoking on the pulmonary or cardiovascular systems. Other findings indicated that smokers could not keep up with exercise intensity (243,244), had increased heart rates during exercise (243), and had slower recovery after exercise (243), providing additional rationale for targeted smoking cessation and smoking prevention interventions. The exercise adaptation score was 75% in the control group (healthy individuals), indicating exercise adaptation in patients is a strategy to help improve respiratory function in recovering patients whereas non-infected healthy individuals are able to maintain higher levels of exercise adaptation on their own. The observations that the trained COVID-19 group achieved better adaptation trends compared to the control group have demonstrated the potential of AI-driven, personalized training programs to enhance recovery and optimize respiratory rehabilitation. The significance of the derived rehabilitation-governing scores, confirmed through machine learning analysis, lies in the stratified treatment of rehabilitation, empowering clinicians to tailor training intensity according to individual adaptation scores. The use of the platform also facilitates the wider use of AI technologies in sports science, performance tracking, and personalised health-related optimization efforts extending further out from post-COVID-19 rehabilitation.

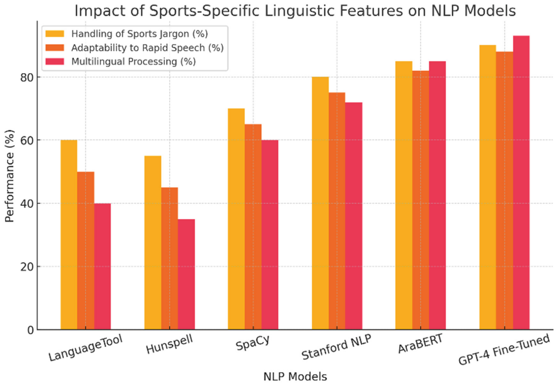

Figure. 3. Impact of sports-specific features on NLP models. [Source: Authors]

The final analysis of linguistic errors in sports commentary showed that lexical and discourse-level errors are most common with basketball and football commentary. This showcases the need for NLP models that are capable of handling such complex vocabularies for the sport that it will be implemented in. The comparison of the NLP models have confirmed that approaches such as GPT-4 and AraBERT have provided the most accurate results. However, due to the high computational cost limits for real time implementation, there is still a need for hybrid models that will combine deep learning and rule-based approaches to address these challenges.

6 Conclusion

In this study, we have shown that integrating artificial intelligence (AI) and structured physical training can effectively help recover lung function in cured COVID-19 patients. The results are consistent with our hypothesis that a personalized AI-based rehabilitation plan utilizing wearable sensor technology would markedly improve respiratory efficiency by monitoring key physiological metrics including oxygen, heart rate, respiratory rate, and lactic acid levels. This eight-week long study demonstrates that the implementation of structured aerobic exercise resulted in a 5% increase in oxygen saturation and a significant decrease in lactic acid accumulation, demonstrating that physical activity is an essential part of post-viral recovery in this respiratory condition. Moreover, machine learning techniques like Random Forest and Long Short-Term Memory (LSTM) networks accurately forecasted individual shifts in recovery patterns, showcasing the potential for AI to further tailor and improve rehabilitation protocols through immediate biometric feedback. It also highlighted the role of baseline fitness in recovery trajectories, with people who were previously aerobically conditioned adapting faster to the training regimen. You have a larger proportion of data that are for smokers. The intergroup comparison showed that the oxygen saturation and lactic acid in the smoker group were lower and higher, respectively, compared with the non-smokers, indicating the long-term negative effect of smoking on pulmonary function, and the exceedance of targeted rehabilitation strategies. The evidence-based analysis by way of the AI confirmed that structured aerobic coaching can partially offset these results, although the research that explored this were in these with compromised respiratory effectiveness. Even beyond implications for post-COVID rehabilitation, this study also adds to the emerging field of AI-based health monitoring and sports science, showcasing the potential of real-time physiological signal processing for rehabilitation. Utilization of deep learning techniques has accelerated early detection of physiological response patterns, modeling recovery rates. This information set the stage for more adaptive rehabilitation protocols that tailor training intensity depending on an individuals real-time metrics of health. These findings have important clinical and technological implications and thus pave the way for this AI-supported rehabilitation program to be deployed in patients with chronic lung diseases such as Chronic Obstructive Pulmonary Disease (COPD) or asthma, in addition to long COVID recovery. Further studies are needed to increase the prediction accuracy of such networks, incorporating deep learning approaches, such as those based on Transformer models or Recurrent Neural Networks (RNNs) to improve predictions for real-time recovery. Finally, and not least, extending the dataset on a more diverse population (age, gender and comorbidity) would improve the robustness of the result and allow readers to generalize from it. Another exciting avenue is the development of multimodal AI models that include environmental- and genetic correlates in rehabilitation strategies, enabling a more tailored, holistic treatment approach. Moreover, including other exercise regimes such as resistance training and interval training may contribute to a greater understanding in optimizing rehabilitation approaches for different patient populations. Lastly, using this methodology in other areas such as post-cardiac rehabilitation or sports injury recovery can establish AI in modern healthcare and performance enhancement. This study demonstrates the feasibility of using AI to direct physiotherapy management of pneumonia, bringing new insights into the medical rehabilitation sciences field. Together, when sports science converges with AI and digital health monitoring, this study lays the groundwork for launching future AI-based rehabilitation programs that will optimize patient recovery, improve healthcare efficiency, and enrich life quality.

References

1. Oyebode, O., et al.: Health, psychosocial, and social issues emanating from the COVID-19 pandemic based on social media comments: text mining and thematic analysis approach. JMIR Med. Inform. **9**(4), e22734 (2021). https://doi.org/10.2196/22734
2. Sainani, K.L., et al.: Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. Br. J. Sports Med. **55**(2), 118–122 (2020). https://doi.org/10.1136/bjsports-2020-102607
3. Abdelhalim, S.M., Alsahil, A.A., Alsuhaibani, Z.A.: Artificial intelligence tools and literary translation: a comparative investigation of ChatGPT and Google Translate from novice and advanced EFL student translators' perspectives. Cogent Arts Humanit. **12**(1), 2508031 (2025). https://doi.org/10.1080/23311983.2025.2508031. (Taylor & Francis Online)
4. Shankar, V., Parsana, S.: An overview and empirical comparison of natural language processing techniques. Neurocomputing **470**, 443–456 (2022). https://doi.org/10.1016/j.neucom.2021.05.010
5. Turchin, A., Florez Builes, L.F.: Using natural language processing to measure and improve quality of diabetes care: a systematic review. J. Diabetes Sci. Technol. **15**(3), 553–563 (2021). https://doi.org/10.1177/19322968211000831

6. Weisz, J.D., et al.: Better together? An evaluation of AI-supported code translation. Proceedings of the 27th International Conference on Intelligent User Interfaces (IUI '22), pp. 369–391 (2022). ACM Digital Library. https://doi.org/10.1145/3490099.3511157
7. Hariri, W.: Unlocking the Potential of Artificial Intelligence in Text Generation. arXiv preprint arXiv:2305.14842 (2023). https://doi.org/10.48550/arXiv.2305.14842
8. Young, T., Hazarika, D., Poria, S., Cambria, E.: Recent trends in deep learning-based natural language processing. IEEE Comput. Intell. Mag. **13**(3), 55–75 (2018). https://doi.org/10.1109/MCI.2018.2840738
9. Liu, Y., et al.: Summary of ChatGPT-related research and perspective towards the future of large language models. Engineering (2023). https://doi.org/10.1016/j.eng.2023.06.003
10. Sarkar, S., Nasipuri, M.:. A deep learning based approach to hate speech detection in social media. Procedia Computer Science **167**, 2318–2327 (2020). https://doi.org/10.1016/j.procs.2020.03.285
11. Rawian, R., Abdullah, N., Al-Said, A.: A critical stylistics analysis of sports commentaries (FIFA World Cup Qatar 2022): textual-conceptual functions in live match commentaries. World Journal of English Language **14**(1), 89 (2023). https://doi.org/10.5430/wjel.v14n1p89
12. Alqmase, M., Al-Qurishi, M., Alsharabi, H.: The gender gap tracker: using natural language processing to measure gender bias in media. PLoS ONE **16**(1), e0245533 (2021). https://doi.org/10.1371/journal.pone.0245533
13. Liu, Y., et al.: Sport Sentiment Analysis Using Natural Language Processing. arXiv preprint arXiv:2209.03434 (2022). https://doi.org/10.48550/arXiv.2209.03434
14. Dergaa, I., Chamari, K., Zmijewski, P., Ben Saad, H.: From human writing to artificial intelligence generated text: examining the prospects and potential threats. JMIR Med. Inform. **9**(4), e24174 (2021). https://doi.org/10.2196/24174
15. Gunasekaran, K.B.: Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review. arXiv preprint arXiv:2305.14842 (2023). https://doi.org/10.48550/arXiv.2305.14842