SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(2), 539.
https://doi.org/10.62617/mcb539

Article

Sports training injuries and prevention measures using big data analysis

Yuan Xue[1], Erjuan Du[2], Zhihong Hou[1,*]

[1] Department of Sports, Qinhuangdao Vocational and Technical College, Qinhuangdao 066100, China
[2] Hebei Province Sports Bureau's Community Sports Center, Shijiazhuang 050041, China
*** Corresponding author:** Zhihong Hou, zhhou_qvc@163.com

CITATION

Xue Y, Du E, Hou Z. Sports training injuries and prevention measures using big data analysis. Molecular & Cellular Biomechanics. 2025; 22(2): 539.
https://doi.org/10.62617/mcb539

ARTICLE INFO

Received: 15 October 2024
Accepted: 25 October 2024
Available online: 23 January 2025

Abstract: This work explores the application of big data technology in monitoring sports training injuries, emphasizing the biomechanical principles underlying injury mechanisms to enhance the accuracy of injury prediction and provide scientific prevention measures. It collects training data from professional sports teams using big data technology and constructs a Bi-directional Long Short-Term Memory (BiLSTM)—Residual Network (ResNet) model through deep learning techniques. In this model, the BiLSTM module captures the temporal sequence features of sports data, while the ResNet module improves the model's expressiveness and stability through residual learning. To establish a clearer connection with mechanobiology, the study discusses the mechanical forces involved in sports injuries, including impact forces, torsional stresses, and their effects on tissues at the cellular level. By integrating biomechanical insights with big data analytics, the research aims to provide a comprehensive understanding of how mechanical stressors contribute to injury risk. The performance of the proposed model in predicting sports injury risks is evaluated, showing an accuracy of 95.72%, a precision of 91.59%, a recall of 85.40%, and an *F*1 score of 88.56%, significantly outperforming existing traditional models and other comparison algorithmsTherefore, the proposed model demonstrates exceptional performance in improving the accuracy of sports injury prediction and providing personalized prevention measures, offering experimental references for the intelligent development of the sports field by bridging sports science and biomechanics.

Keywords: big data analysis; sports injuries; biomechanics; BiLSTM; personalized prevention; deep learning

1. Introduction

Sports training is a crucial component for enhancing athletes' competitive abilities. However, as the intensity and frequency of training increase, the incidence of sports injuries also rises. These injuries can severely impact athletes' health and may have long-term negative effects on their careers and competitive status [1,2]. Although existing traditional models have made certain progress in theory and practice, they still face many challenges in terms of accuracy, generalization ability, and real-time prediction. First, most traditional models rely on limited data sources and empirical methods and lack an in-depth understanding of individual differences among athletes and complex training environments. Besides, the causes of sports injuries are often multi-factorial, involving multiple levels such as physiology, psychology, and environment. This requires prediction models to be able to process and analyze large-scale multi-dimensional data. In addition, as the training intensity of athletes increases, the demand for real-time monitoring and rapid response is also getting higher and higher. It puts forward higher requirements for the real-time prediction ability and computational efficiency of the model. However, existing models are often difficult to balance accuracy and real-time performance and are difficult to meet the practical application needs in a high-intensity training environment. Consequently, data-driven research has become a vital approach for improving injury prevention strategies and has attracted significant attention from scholars in the field. This shift towards a more analytical approach allows for the integration of advanced technologies, such as machine learning and biomechanical analysis, to identify patterns and risk factors associated with injuries. By leveraging large datasets and developing predictive models, researchers can better understand injury mechanisms, optimize training regimens, and implement personalized injury prevention programs tailored to individual athletes' needs. This data-driven strategy not only enhances the effectiveness of prevention measures but also helps in refining overall training methodologies, promoting safer and more efficient training environments for athletes at all levels.

For instance, in recent years, the rapid development of big data technology has provided new perspectives and methods for sports injury research. Big data technology enables the integration and analysis of large-scale, multi-dimensional data from various sources, revealing patterns and trends in sports injury occurrences [3,4]. This data-driven approach not only offers abundant empirical evidence but also identifies potential risk factors and injury patterns, providing strong support for developing scientific prevention measures. By deeply analyzing athletes' training data, incorporating deep learning algorithms, monitoring training conditions in real-time, and issuing warnings for potential injury risks, targeted prevention strategies can be proposed [5]. For instance, Wang et al. [6] demonstrated significant predictive effects of deep learning algorithms in improving resource efficiency and achieving sustainability goals.

Therefore, this work aims to use big data analysis technology to systematically investigate sports training injuries, explore the patterns of sports injury occurrence, and propose effective prevention measures by establishing an injury prediction model. Specifically, the research objectives include integrating multi-source data from sports monitoring devices, health records, and athlete feedback for comprehensive data analysis. Next, a deep learning-based injury prediction model supported by big data is constructed and its effectiveness in practical applications is evaluated. Finally, this work proposes scientifically sound sports injury prevention strategies based on data analysis results to improve training effectiveness and safety for athletes. The significance of this work lies in overcoming the limitations of traditional injury prevention methods through big data technology and offering new research methods and practical guidance for the field of sports medicine. Gaining a deeper understanding of the mechanisms and influencing factors of sports injuries can develop more targeted training programs and prevention measures, effectively reducing the incidence of injuries and enhancing athletes' performance and health.

2. Literature review

Traditional methods of sports injury prevention primarily rely on experience and theoretical knowledge. For example, Cornelissen et al. [7] assessed the injury prevention effectiveness and implementation of the "Warming-up Hockey" program using a mixed-methods approach. The findings reveal the actual effectiveness and promotion challenges of traditional warm-up procedures in reducing sports injuries. MacFarlane et al. [8] explored the factors influencing the cognition, adoption, and implementation of anterior cruciate ligament injury prevention methods in youth sports. They emphasized the application and limitations of traditional prevention strategies in sports environments. Tabben et al. [9] analyzed the implementation barriers and opportunities for injury prevention measures in professional football in Qatar, reflecting the challenges and areas for improvement in traditional prevention methods.

With the development of information technology, big data analysis has gradually been introduced into sports medicine, bringing new opportunities for sports injury research. For instance, Wang et al. [10] used deep transfer learning and multimodal digital twin technology to enhance and diagnose brain MRI image analysis. The findings showcased the potential of big data in sports medicine, particularly in handling complex data and personalized diagnostics. Liu et al. [11] proposed a hybrid design method combining artificial intelligence and big data analysis for visualizing sports data, highlighting the significant role of big data technology in sports data mining and decision support. Nassis et al. [12] reviewed the application of machine learning in predicting football injury risks and emphasized the critical role of big data analysis in identifying and preventing sports injuries, thus advancing precision medicine. Hughes et al. [13] discussed new technologies in sports biomechanics and cautioned about the potential challenges of big data analysis in sports medicine, stressing the importance of careful use of new technologies. Dergaa and Chamari [14] discussed the integration of big data in sports medicine and sports science, highlighting its role in driving future innovations and demonstrating the importance of combining theory and practice.

As machine learning technology rapidly advances, researchers have started experimenting with more complex algorithms, such as Support Vector Machines (SVM), decision trees, random forests, and deep learning models, to build injury prediction models. Park et al. [15] utilized a machine learning model to analyze the risk of Ramp lesions in anterior cruciate ligament injuries. By combining clinical data, the model could effectively predict the occurrence of such injuries, demonstrating the application prospects of machine learning in sports injury risk assessment. Hecksteden et al. [16] used a multi-faceted approach combining screening, monitoring, and machine learning to successfully predict injuries in football, demonstrating the advantages of machine learning in complex prediction tasks. Haller et al. [17] employed comprehensive monitoring methods and machine learning to predict injuries and illnesses in elite youth football players, showing the method’s predictive capability over long-term observation. Kumar et al. [18] reviewed the application of artificial intelligence in sports injury prediction and emphasized the potential of machine learning to improve prediction accuracy and reliability. Ayala et al. [19] conducted innovative research on the early identification of athlete injury risks through machine learning technology. The research results showed that the model could effectively predict potential injury risks and help improve athlete health management and preventive measures.

Although traditional methods of sports injury prevention have achieved some success under theoretical and experiential guidance, recent research has revealed their limitations and challenges in practical applications. For instance, Cornelissen et al. [7] and MacFarlane et al. [8] indicated that traditional warm-up procedures and prevention strategies faced difficulties in promotion and application, reflecting adaptation issues in different sports environments. Similarly, Tabben et al. [9] highlighted barriers and opportunities for improvement in implementing these measures in professional sports. With the rise of big data and artificial intelligence technologies, new opportunities have emerged in the field of sports medicine. Wang et al. [10] and Liu et al. [11] demonstrated the potential of big data in handling complex data and personalized analysis, especially in sports injury prediction and prevention. However, Hughes et al. pointed out challenges in applying big data analysis technologies and the need for careful consideration. The application of machine learning technologies offers more precise tools for injury prediction, although studies by Hecksteden et al. [16] indicate that existing models still have methodological and performance limitations. Therefore, the research innovation lies in integrating multi-dimensional data sources using big data and machine learning technologies to develop more accurate and practical sports injury prediction models. This aims to overcome the limitations of traditional methods and enhance the scientific and effective nature of prevention measures, thus providing new research pathways and practical guidance for the field of sports medicine.

3. Research model

This section provides a detailed explanation of the entire process of collecting and analyzing sports data supported by big data, constructing a sports injury prediction model based on deep learning, and conducting experimental evaluations. By combining big data technology with advanced deep learning models, this work aims to develop an efficient and accurate sports injury prediction system to help athletes prevent injuries during high-intensity training.

3.1. Big data-supported sports data collection and analysis

In modern sports training, data collection has become a core element in predicting and preventing sports injuries. Traditional methods of collecting sports data typically rely on manual recording and limited monitoring devices, which restrict the accuracy and comprehensiveness of the data [20]. However, with the rapid advancement of big data technology, multidimensional data on athletes, such as biomechanics, psychological, and physiological metrics, can now be collected in real-time using advanced wearable devices and sensors. **Table 1** displays the model and specification settings for wearable devices and sensors.

Table 1. Model and specification setting for wearable devices and sensors.

Model and specification setting table for wearable devices and sensors	Model	Specification and Characteristics	Collected Data Type
Heart Rate Monitor	Polar H10	Bluetooth compatibility, heart rate zone monitoring, suitable for high-intensity training	Heart rate, heart rate zone (% of maximum heart rate)
Stride Frequency Sensor	Garmin Foot Pod	Records stride frequency and stride length, synchronizes with Garmin devices, suitable for running and cycling	Stride frequency (steps/minute), stride length (meters/step)
Accelerometer	ADXL345	Three-axis acceleration sensor, can measure dynamic motion, range ± 2 g/± 4 g/± 8 g/± 16 g	Acceleration (m/s^2), dynamic motion mode
Electromyography Sensor	MyoMotion	Surface electromyography sensor for monitoring muscle activity, wireless data transmission	Electromyogram signal (μV), muscle activity intensity
GPS Tracker	Garmin Fenix 6	GPS, GLONASS, and Galileo satellite support, high-precision position tracking, suitable for outdoor sports	Position (latitude and longitude), speed (m/s), altitude (meters)
Skin Temperature Sensor	TMP102	Used for monitoring skin surface temperature, I2C interface	Skin temperature (°C), temperature change trend

These devices accurately record various metrics during training, such as heart rate, step frequency, muscle activity, and body posture, and upload the data to the cloud for storage and analysis. This real-time, high-frequency data collection significantly enhances the timeliness and accuracy of the data, providing a rich data foundation for subsequent prediction models.

Big data technology has not only transformed the way sports data are collected but also revolutionized data processing and analysis methods. In sports data analysis, traditional data processing methods often struggle to handle such vast and complex datasets, whereas big data analysis tools like Hadoop and Spark can efficiently process and analyze these data [21,22]. These tools use distributed computing frameworks to break down and execute data processing tasks in parallel, greatly improving the speed and efficiency of data analysis. Additionally, data cleaning and preprocessing techniques, such as outlier detection, data smoothing, and feature engineering, ensure high-quality data, laying a solid foundation for model construction. Big data analysis enables the identification of hidden patterns and potential risk factors in athletes' training processes, providing crucial support for the development of deep learning models.

Therefore, the sports injury prediction model proposed tightly integrates big data technology with deep learning models. By leveraging multidimensional sports data collection and analysis supported by big data, key features closely related to sports injuries are extracted and input into a deep learning-based sports injury prediction model. Big data analysis not only provides high-quality input data for the model but also aids in identifying high-risk behaviors and indicators associated with specific training patterns through retrospective analysis of historical data. Compared to traditional single data source approaches, this big data-based multidimensional analysis offers a more comprehensive reflection of an athlete's overall state, providing rich contextual information for deep learning models and significantly enhancing the accuracy and practicality of sports injury predictions.

3.2. Construction and analysis of a deep learning-based sports injury prediction model

To improve the accuracy of sports injury predictions, this work proposes a deep learning model that integrates a Bidirectional Long Short-Term Memory (BiLSTM) network with a Residual Network (ResNet) [23,24]. The BiLSTM is adept at capturing temporal dependency features within sports data, making it well-suited for analyzing the dynamic changes in athletes' physical states during training. On the other hand, ResNet addresses the vanishing gradient problem in deep neural networks by introducing residual structures, thereby enhancing the model's representational capacity. **Figure 1** illustrates the specific sports injury prediction model based on the BiLSTM-ResNet algorithm.

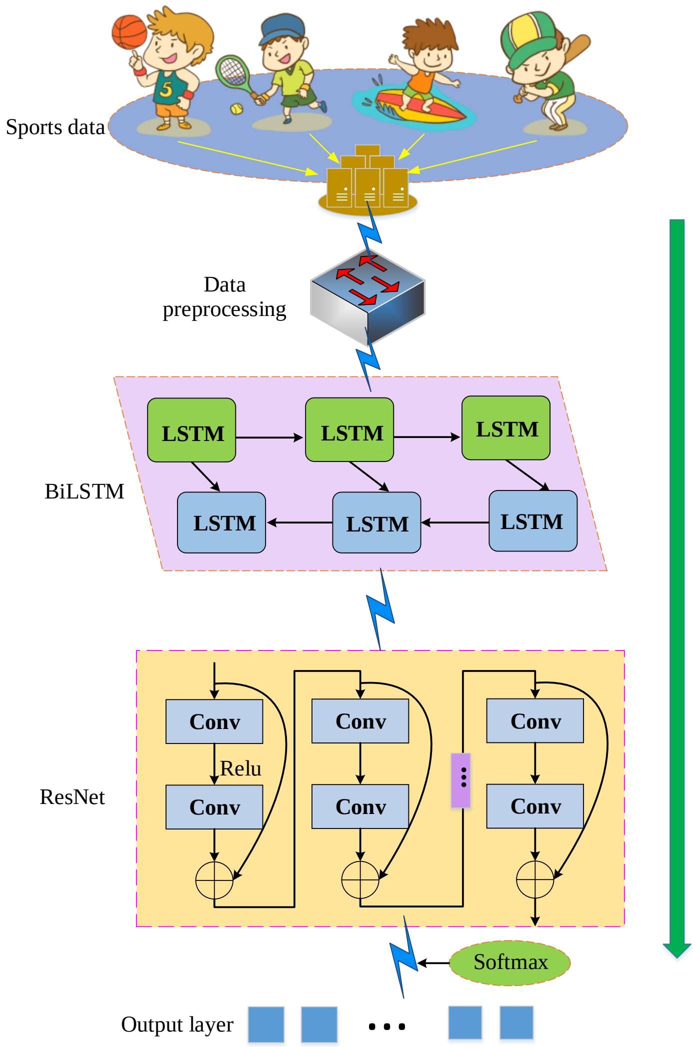

Figure 1. Framework of the sports injury prediction model based on the BiLSTM-ResNet algorithm.

Figure 1 illustrates the framework of the prediction model, which integrates BiLSTM and ResNet deep learning techniques. The model begins with the collection of time-series data from sports sensors, such as heart rate, step frequency, and acceleration. After these data are standardized, they are fed into the BiLSTM module. In the BiLSTM module, the forward LSTM layer extracts features from the beginning of the time series to the current position, while the backward LSTM layer extracts information in reverse, from the end of the time series to the current position. This bidirectional feature extraction capability allows the BiLSTM to capture complex temporal dependencies within the sports data, enhancing the understanding of the dynamic states of the athletes. The feature extraction performed by the bidirectional LSTM units generates high-dimensional feature vectors enriched with temporal information. These vectors reflect the physiological changes and potential risks of athletes during different training phases, providing crucial data support for subsequent injury predictions. In the BiLSTM module, the hidden state h_t at time t can be represented as shown in Equations (1) and (2):

$$\overrightarrow{h}_t = \overrightarrow{LSTM}\left(\overrightarrow{W}_t, \overrightarrow{h}_{t-1}, \overrightarrow{b}_t, \overrightarrow{c}_{t-1}\right) \tag{1}$$

$$\overleftarrow{h}_t = \overleftarrow{LSTM}\left(\overleftarrow{W}_t, \overleftarrow{h}_{t-1}, \overleftarrow{b}_t, \overleftarrow{c}_{t+1}\right) \tag{2}$$

W and b represent the weights associated with the gate units and memory cells, respectively, while c_t and h_t denote the state of the memory cell and the hidden state of the LSTM at time t. The arrows → and ← signify the forward and backward temporal features, respectively.

To further enhance the model's performance, the temporal feature vectors extracted by the BiLSTM are fed into the ResNet module. The ResNet module consists of multiple residual blocks, each comprising two convolutional layers and a shortcut connection. The BiLSTM feature vectors processed through these residual blocks allow for the extraction of deep-level features from the sports data. ResNet introduces residual blocks to mitigate the vanishing gradient problem in deep neural networks, thereby improving the model's training stability and expressive power. The design philosophy of residual blocks involves using shortcut connections to pass the input signal directly to deeper layers of the network, avoiding information loss and difficulties associated with training deep networks. This approach captures complex patterns and nonlinear relationships in the sports data, thereby enhancing the model's ability to predict sports injuries.

After the feature extraction by the BiLSTM and ResNet modules, a fusion layer integrates these two sets of features. The fusion process involves concatenating or applying weighted averages to the temporal features generated by the BiLSTM and the deep features extracted by the ResNet, resulting in a comprehensive feature representation. These fused features are then input into a fully connected layer for the final classification or regression task, predicting the risk of sports injuries.

3.3. Experimental evaluation

To evaluate the effectiveness of the model constructed, experimental data are sourced from the training records of the S professional sports team. Data collection took place from September 2023 to November 2023, encompassing detailed physiological and kinematic indicators of different athletes across various training sessions. These data are gathered using advanced wearable devices and sensors, capturing multi-dimensional information such as heart rate, stride frequency, acceleration, and electromyographic signals. The data spans an entire training cycle, ensuring data comprehensiveness and representativeness. Additionally, the dataset includes records of injuries sustained by athletes to mark the occurrence of sports injuries. In the data preprocessing stage, the raw data are cleaned and standardized. In data cleaning, the collected raw data are thoroughly inspected first to identify and handle missing values. For missing data, multiple strategies are adopted. For a small amount of missing values, mean or median filling is used, for a large range of missing data, time series interpolation methods are used to maintain the time continuity of the data. In addition, outlier detection is also performed on the data to identify extreme or inconsistent data points that may affect model performance. Outlier detection is carried out through statistical analysis methods such as *Z*-score and interquartile range (IQR) methods. First, the *Z*-score of each feature is calculated to identify outliers that are far from other data points. Then, the IQR method is used to further filter out values that are outside the normal range. These outliers are processed according to specific situations, such as replacement, deletion, or transformation, to reduce their impact on model training. Finally, different types of input features are normalized to meet the input requirements of the model. The dataset is then split into training and test sets in an 8:2 ratio.

The experiments are conducted on a high-performance computing platform equipped with an NVIDIA Tesla V100 GPU, running the Ubuntu 20.04 operating system. The deep learning framework is built using the Python programming language, specifically leveraging the TensorFlow and Keras libraries. During model training, the hyperparameter settings of both the BiLSTM and ResNet architectures play a critical role in determining the final performance. **Table 2** provides the detailed hyperparameter settings.

Table 2. Hyperparameter settings.

Hyperparameter	Value	Hyperparameter	Value
Number of Hidden Units	128	Initial Learning Rate	0.001
Activation Function	tanh	Learning Rate Adjustment Strategy	Dynamic Learning Rate Adjustment
Number of BiLSTM Layers	2–10	Batch Size	64
Number of Residual Blocks	4	Number of Training Epochs	50
Number of Convolutional Layers	2 Layers per Residual Block	Dropout Rate	0.5
Convolutional Kernel Size	3 × 3	Regularization Type	L2 Regularization
Batch Normalization	Yes	L2 Regularization Parameter	0.01
Optimization Algorithm	Adam		

Table 2 shows that in the construction process of the BiLSTM-ResNet model proposed, the selection of hyperparameters has a decisive impact on the performance of the model. The number of hidden units and layers is carefully selected to achieve the best balance between accuracy and efficiency. When choosing hidden units, fewer units may lead to insufficient expressive ability of the model, while too many units may cause overfitting and increase the computational burden of model training. Selecting 128 hidden units can provide sufficient model capacity for the task while maintaining good generalization ability. In terms of layer selection, a configuration

of a 2-layer BiLSTM network and 4 ResNet blocks is adopted. This design is based on considerations of the difficulty of training deep networks and performance improvement. Too few layers may not be able to fully capture the complex patterns in the data, while too many layers may lead to problems such as vanishing gradients or exploding gradients during the training process, affecting the stability and convergence speed of the model. Through cross-validation and multiple experiments, it is determined that a 2-layer BiLSTM can effectively capture long-term dependencies in time series data. Besides, 4 ResNet blocks provide sufficient depth for the model to learn complex features in the data, while avoiding the training difficulties brought by overly deep network structures. In addition, dynamic learning rate adjustment strategies and L2 regularization techniques are also adopted to further improve the generalization performance of the model. The dynamic learning rate adjustment strategy allows the model to learn quickly in the early stage of training, and in the later stage of training, it can fine-tune the weights to avoid local minima. L2 regularization helps reduce model complexity and prevent overfitting. Through these carefully selected hyperparameters, the proposed model not only performs well on the training set but also shows strong predictive ability and generalization on unknown data.

The accuracy and stability of the BiLSTM and ResNet fusion model developed is further analyzed for predicting sports injuries. Based on the experimental results, the performance of the proposed model is evaluated against BiLSTM, LSTM, and the model proposed by Haller et al. [17] on the test set [25]. The evaluation metrics include accuracy, precision, recall, $F1$ score, root mean squared error (RMSE), and operational efficiency.

4. Results and discussion

4.1. Performance comparison analysis of different algorithms

First, the performance of the proposed model is evaluated and compared with BiLSTM, LSTM, and the model proposed by Haller et al. [17] on the test set. **Figure 2** shows the results for accuracy, precision, recall, and $F1$ score.

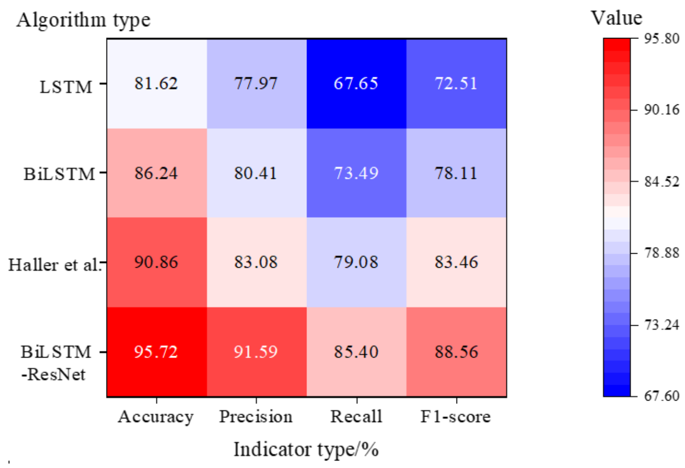

Figure 2. Accuracy results of sports injury prediction under different algorithms.

Figure 2 demonstrates the significant advantages of the BiLSTM-ResNet algorithm proposed for predicting sports injuries. In terms of accuracy, the BiLSTM-ResNet algorithm achieves 95.72%. It is significantly higher than the model proposed by Haller et al. at 90.86%, BiLSTM at 86.24%, and LSTM at 81.62%, indicating its outstanding performance in overall prediction accuracy. Regarding precision, the BiLSTM-ResNet algorithm also outperforms other comparative algorithms with a precision of 91.59%, indicating a lower false-positive rate when accurately identifying sports injury events. For recall, the BiLSTM-ResNet algorithm achieves 85.40%, surpassing the models of Haller et al., BiLSTM, and LSTM, showing a stronger ability to recognize actual injury events. In the comprehensive evaluation metric of *F*1 score, the BiLSTM-ResNet algorithm leads with a score of 88.56%, demonstrating its excellent performance in balancing precision and recall. Overall, the BiLSTM-ResNet model algorithm proposed significantly outperforms other algorithms across all metrics, proving its powerful capability and potential applications in sports injury prediction.

Furthermore, a comparison is made of the RMSE and operational efficiency of the proposed model algorithm under 2–10 layers of the BiLSTM model, as shown in **Figure 3**.

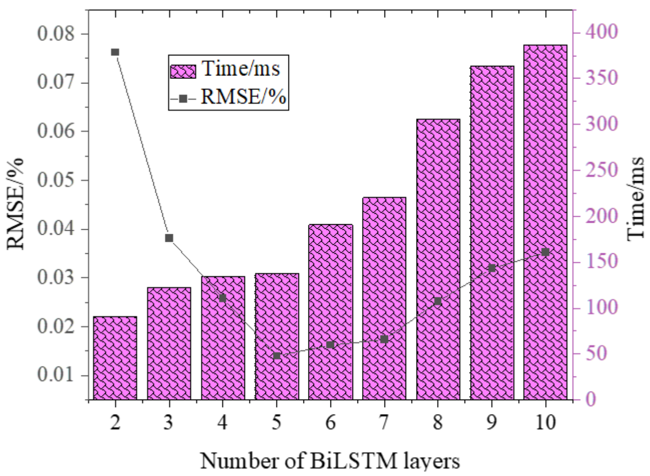

Figure 3. Changes in RMSE and operational efficiency of the model with different BiLSTM layers.

Figure 3 illustrates the evaluation results of the proposed model algorithm by comparing the RMSE and operational efficiency of the BiLSTM model with varying numbers of layers for predicting sports injury risk. It can be observed that as the number of BiLSTM layers increases, the RMSE first decreases and then increases. Starting from 0.0762% with 2 layers, it gradually decreases to a minimum of 0.0140% at 5 layers, indicating that a deeper network initially effectively reduces prediction errors. However, when the number of layers continues to increase beyond 6, the RMSE begins to rise again, reaching 0.0352% at 10 layers, suggesting that too many layers may lead to overfitting or difficulties in training the model. Meanwhile, the operational time also increases significantly with the number of layers, rising from 91.21 milliseconds at 2 layers to 386.60 milliseconds at 10 layers. Although the 5-layer BiLSTM demonstrates the best performance in terms of RMSE, its operational efficiency is relatively low. This indicates that in practical applications, a balance between accuracy and efficiency is necessary to select the optimal network structure.

4.2. Discussion

The BiLSTM-ResNet model algorithm developed demonstrates excellent performance in predicting injuries during sports training. Compared to the research by Meng and Qiao [26], this work significantly improves the accuracy and precision of predicting sports injury risk by integrating deep learning and big data analysis techniques. Unlike traditional predictive models, the BiLSTM-ResNet model achieves a prediction accuracy of over 95% for sports injuries. This shows significant advantages across various performance metrics and captures potential risk factors for sports injuries more accurately. Compared with the traditional LSTM baseline algorithm, the multi-dimensional analysis based on big data analysis provides a more comprehensive reflection of the overall state of athletes. This provides more abundant context information for deep learning models. This aligns with the findings of Zhang et al. [27] and Fathi et al. [28]. Therefore, the model proposed not only improves the effectiveness of preventive measures but also helps improve the overall training method and provides a safer and more efficient training environment for athletes at all levels. This data-driven strategy not only improves the effectiveness of preventive measures but also helps improve the overall training method, providing stronger guarantees for the health and competitive state of athletes. The superior performance of the proposed model allows coaches and athletes to identify potential injuries earlier, enabling them to adjust training plans accordingly to prevent serious injuries and ensure the health and safety of the athletes.

When analyzing the potential implementation challenges of the BiLSTM-ResNet model in the real sports environment, data privacy and the need for dedicated equipment must be considered. First, the physiological and training data of athletes is very sensitive personal information. Therefore, when collecting and analyzing these data, strict data protection regulations must be complied with. This may require additional investment in data security infrastructure and ensuring that all data processing activities comply with privacy standards such as GDPR. In addition, the proposed model relies on advanced wearable devices and sensors to collect multi-dimensional data, which may require sports organizations to invest in expensive hardware. This demand for dedicated equipment may become an obstacle for small or underfunded sports teams to implement the model. Therefore, future work should explore how to reduce these thresholds by using more economical or existing technologies while still maintaining the accuracy and practicability of the model.

Next, there are also potential differences when applying the proposed model to different sports. Different sports have different injury patterns and risk factors, which poses a challenge to the universality of the proposed model. For example, contact sports such as rugby may have more impact-related injuries, while endurance sports such as long-distance running may face more overuse-related injuries. The proposed model needs to be able to recognize and adapt to these specific injury patterns to provide customized preventive measures for each sport. This may require additional analysis of a large amount of historical injury data for each sport to ensure that the model can capture the risk factors unique to each sport. In addition, the model may need to be fine-tuned for specific sports to optimize its predictive ability. This adaptability is the key to improving the practical applicability of the model in different sports environments.

Finally, the computational requirements of the model are also an important aspect that potential users need to consider. The proposed BiLSTM-ResNet model is a deep learning model that requires significant computational resources for training and prediction, especially when processing large amounts of data. This may require high-performance computing hardware such as GPU and corresponding software support. For sports organizations with limited resources, this may be an important consideration. However, with the development of cloud computing services, these computing resources become more accessible and cost-effective. Future work can explore how to further optimize the computational efficiency of the model, such as using model compression or more efficient deep learning architectures to achieve wider application in resource-constrained environments.

Additionally, the proposed model provides a solid theoretical foundation for developing more personalized and scientifically-based preventive measures. By thoroughly analyzing historical sports data and real-time monitoring data, the BiLSTM-ResNet model can identify the injury risks faced by specific athletes during particular training phases, thereby offering precise preventive recommendations to support the long-term development of athletes.

5. Conclusion

The BiLSTM-ResNet model developed has achieved significant research outcomes in predicting sports training injuries, demonstrating strong capabilities in enhancing prediction accuracy and identifying injury risks. By integrating deep learning and big data technologies, this work successfully improves the prediction accuracy of injury risks to 95.72%, significantly surpassing traditional models. Moreover, the model provides personalized preventive measures for athletes, effectively reducing the occurrence of sports injuries. However, the work also has certain limitations, such as the diversity and representativeness of the data sources, which may impact the model’s generalizability. Additionally, the complexity of the model and the increased runtime with more layers limit its efficiency in practical applications. Future research should focus on expanding the diversity and scale of the dataset, optimizing the model’s computational efficiency, and exploring more data-driven preventive strategies to further enhance the model’s practicality and reliability.

Author contributions: Conceptualization, YX and ED; methodology, ED and ZH; software, ZH and YX; validation, ZH, YX and ED; formal analysis, ZH, YX and ED; investigation, ZH and ED; resources, ZH; data curation, YX; writing—original draft preparation, YX and ED; writing—review and editing, YX and ED; visualization, YX; supervision, ZH; project administration, ED; funding acquisition,

ZH and YX. 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. Podlog, L., Wagnsson, S., & Wadey, R. (2024). The impact of competitive youth athlete injury on parents: a narrative review. Sport in Society, 27(8), 1332-1355.
2. Gerçek, H., Işık, İ. D., Gürel, M. N., Pekyavaş, N. Ö., & Altıntaş, A. (2023). Comparison of Sports Injury Anxiety in Athletes Doing Sports on Different Surfaces. International Journal of Disabilities Sports and Health Sciences, 6(1), 1-7.
3. Guelmami, N., Fekih-Romdhane, F., Mechraoui, O., & Bragazzi, N. L. (2023). Injury Prevention, Optimized Training and Rehabilitation: How Is AI Reshaping the Field of Sports Medicine. New Asian Journal of Medicine, 1(1), 30-34.
4. Seçkin, A. Ç., Ateş, B., & Seçkin, M. (2023). Review on Wearable Technology in sports: Concepts, Challenges and opportunities. Applied Sciences, 13(18), 10399.
5. Eid, A. I. A., Miled, A. B., Fatnassi, A., Nawaz, M. A., Mahmoud, A. F., Abdalla, F. A., ... & Mohamed, I. B. (2024). Sports Prediction Model through Cloud Computing and Big Data Based on Artificial Intelligence Method. Journal of Intelligent Learning Systems and Applications, 16(2), 53-79.
6. Wang, Z., Deng, Y., Zhou, S., & Wu, Z. (2023). Achieving sustainable development goal 9: A study of enterprise resource optimization based on artificial intelligence algorithms. Resources Policy, 80, 103212.
7. Cornelissen, M. H., Kemler, E., Baan, A., & van Nassau, F. (2023). Mixed-methods process evaluation of the injury prevention Warming-up Hockey programme and its implementation. BMJ Open Sport & Exercise Medicine, 9(2), e001456.
8. MacFarlane, A. J., Whelan, T., Weiss-Laxer, N. S., Haider, M. N., Dinse, S. A., Bisson, L. J., & Marzo, J. M. (2024). Factors associated with awareness, adoption, and implementation of anterior cruciate ligament injury prevention in youth sports. Sports health, 16(4), 588-595.
9. Tabben, M., Verhagen, E., Warsen, M., Chaabane, M., Schumacher, Y., Alkhelaifi, K., ... & Bolling, C. (2023). Obstacles and opportunities for injury prevention in professional football in Qatar: exploring the implementation reality. BMJ Open Sport & Exercise Medicine, 9(1), e001370.
10. Wang, J., Qiao, L., Lv, H., & Lv, Z. (2022). Deep transfer learning-based multi-modal digital twins for enhancement and diagnostic analysis of brain mri image. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20(4), 2407-2419.
11. Liu, A., Mahapatra, R. P., & Mayuri, A. V. R. (2023). Hybrid design for sports data visualization using AI and big data analytics. Complex & Intelligent Systems, 9(3), 2969-2980.
12. Nassis, G., Verhagen, E., Brito, J., Figueiredo, P., & Krustrup, P. (2023). A review of machine learning applications in soccer with an emphasis on injury risk. Biology of sport, 40(1), 233-239.
13. Hughes, G. T., Camomilla, V., Vanwanseele, B., Harrison, A. J., Fong, D. T., & Bradshaw, E. J. (2024). Novel technology in sports biomechanics: Some words of caution. Sports Biomechanics, 23(4), 393-401.
14. Dergaa, I., & Chamari, K. (2024). Big Data in Sports Medicine and Exercise Science: Integrating Theory and Practice for Future Innovations. Tunisian Journal of Sports Science and Medicine, 2(1), 1-13.
15. Park, Y. B., Kim, H., Lee, H. J., Baek, S. H., Kwak, I. Y., & Kim, S. H. (2023). The clinical application of machine learning models for risk analysis of ramp lesions in anterior cruciate ligament injuries. The American Journal of Sports Medicine, 51(1), 107-118.
16. Hecksteden, A., Schmartz, G. P., Egyptien, Y., Aus der Fünten, K., Keller, A., & Meyer, T. (2023). Forecasting football injuries by combining screening, monitoring and machine learning. Science and medicine in football, 7(3), 214-228.
17. Haller, N., Kranzinger, S., Kranzinger, C., Blumkaitis, J. C., Strepp, T., Simon, P., ... & Stöggl, T. (2023). Predicting injury and illness with machine learning in elite youth soccer: a comprehensive monitoring approach over 3 months. Journal of Sports Science & Medicine, 22(3), 476.
18. Kumar, G. S., Kumar, M. D., Reddy, S. V. R., Kumari, B. S., & Reddy, C. R. (2024). Injury Prediction in Sports using Artificial Intelligence Applications: A Brief Review. Journal of Robotics and Control (JRC), 5(1), 16-26.

19. Ayala, R. E. D., Granados, D. P., Gutiérrez, C. A. G., Ruíz, M. A. O., Espinosa, N. R., & Heredia, E. C. (2024). Novel Study for the Early Identification of Injury Risks in Athletes Using Machine Learning Techniques. Applied Sciences, 14(2), 570.
20. Sun, F., Zhu, Y., Jia, C., Zhao, T., Chu, L., & Mao, Y. (2023). Advances in self-powered sports monitoring sensors based on triboelectric nanogenerators. Journal of Energy Chemistry, 79, 477-488.
21. De Fazio, R., Mastronardi, V. M., De Vittorio, M., & Visconti, P. (2023). Wearable sensors and smart devices to monitor rehabilitation parameters and sports performance: an overview. Sensors, 23(4), 1856.
22. Yang, J., Meng, C., & Ling, L. (2024). Prediction and simulation of wearable sensor devices for sports injury prevention based on BP neural network. Measurement: Sensors, 33, 101104.
23. Afsar, M. M., Saqib, S., Aladfaj, M., Alatiyyah, M. H., Alnowaiser, K., Aljuaid, H., ... & Park, J. (2023). Body-worn sensors for recognizing physical sports activities in Exergaming via deep learning model. IEEE Access, 11, 12460-12473.
24. Wang, L., Ji, W., Wang, G., Feng, Y., & Du, M. (2024). Intelligent design and optimization of exercise equipment based on fusion algorithm of YOLOv5-ResNet 50. Alexandria Engineering Journal, 104, 710-722.
25. Wang, T. Y., Cui, J., & Fan, Y. (2023). A wearable-based sports health monitoring system using CNN and LSTM with self-attentions. Plos one, 18(10), e0292012.
26. Meng, L., & Qiao, E. (2023). Analysis and design of dual-feature fusion neural network for sports injury estimation model. Neural Computing and Applications, 35(20), 14627-14639.
27. Zhang, J. Y., Yang, X. K., Ren, J. J., Li, L. J., Zhang, D. D., Gu, J., & Xiong, W. H. (2024). Terahertz recognition of composite material interfaces based on ResNet-BiLSTM. Measurement, 233, 114771.
28. Fathi, M., Shah-Hosseini, R., & Moghimi, A. (2023). 3D-ResNet-BiLSTM Model: A Deep Learning Model for County-Level Soybean Yield Prediction with Time-Series Sentinel-1, Sentinel-2 Imagery, and Daymet Data. Remote Sensing, 15(23), 5551.