**Construction of risk assessment model and intelligent prevention system for sports injury in adolescent physical training**

Cong Li 1,a；Quan Zhang1,b\*

1. *Pai Chai University, Daejeon, 35345, South Korea*

*aEmail:*Lc2347710@163.com

*bEmail:* zhangquan577615@163.com

Abstract: This study developed a risk assessment model for sports training injuries in adolescents and designed an intelligent prevention system. The model integrates four-dimensional indicators including physiological, technical, environmental, and psychological factors, utilizing machine learning algorithms to enable precise prediction. The system employs a three-tier architecture to develop core functional modules, achieving real-time monitoring, personalized interventions, and multi-party collaborative protection. Validation results demonstrate the model's 92% accuracy rate, with the system significantly reducing injury incidence rates, providing technical support for scientific adolescent training.

Key words: youth sports; sports injury; risk assessment; machine learning; intelligent prevention system

 foreword ：

Sports injuries in youth athletic training have become increasingly prominent, not only hindering athletes 'competitive development but also threatening their long-term health. Traditional experience-based protective measures suffer from delays and subjective limitations. Current research predominantly focuses on adult professional athletes, lacking specialized risk assessment tools tailored to adolescents' developmental characteristics[1]. This study aims to establish a multi-dimensional injury risk prediction model and develop an intelligent proactive prevention system. By integrating sports science, data science, and information technology, this work quantifies risk factors and enables real-time dynamic monitoring, achieving proactive injury prevention. These efforts hold practical value for ensuring youth sports safety and enhancing scientific training methodologies.

1. Construction of risk assessment model for sports injuries in adolescent physical training

(1) Design of risk assessment index system

1. Physiological indicators

The developmental stage of adolescents' physical systems fundamentally influences sports injury risks. Differences in musculoskeletal maturity determine the threshold range for tissue load-bearing capacity. During rapid growth phases, individuals with incompletely closed growth plates are particularly susceptible to growth plate injuries from excessive impact training[2]. Cardiovascular fitness directly affects exercise endurance; athletes with low maximal oxygen uptake (VO2max) are more prone to technical deformities and injuries caused by fatigue during high-intensity training. Muscle strength imbalances are prevalent among adolescents, with individuals having a hamstring-to-quadiceps strength ratio below 0.6 showing significantly increased anterior cruciate ligament injury risks. Flexibility metrics—such as sitting forward bend test results—below age-standard values by 20% indicate restricted joint mobility, thereby increasing the likelihood of muscle strains.

2. Motor skill indicators

Technical movement standardization plays a decisive role in injury prevention. Incorrect landing cushioning actions subject the knee joint to ground reaction forces exceeding eight times body weight, significantly increasing anterior cruciate ligament (ACL) rupture risks. Insufficient kinetic chain transmission efficiency leads to energy leakage, forcing local tissues to compensate through compensatory work that causes overuse injuries. For instance, swimmer impingement syndrome often originates from deviated stroke trajectories. Training duration reflects neuromuscular control maturity: adolescents with less than two years of training demonstrate 37% higher coordination errors in complex movements compared to seasoned athletes. Specialized technical stability can be quantitatively assessed using inertial measurement units (IMUs). In basketball scenarios, a trunk lateral swing angle exceeding 10 degrees during sudden stop-and-shoot shots indicates core instability risks.

3. Psychological indicators

Psychological states influence physiological functions through neuroendocrine pathways. Athletes scoring above the 75th percentile on the Pre-competition Anxiety Scale (CSAI-2) exhibit abnormally elevated cortisol levels, leading to impaired muscle tone regulation[3]. Attentional distraction causes visual information processing delays exceeding 0.3 seconds, with soccer goalkeepers experiencing a 40% increase in save reaction time under distracted conditions. Deviations in motivation levels trigger overtraining tendencies, while positive responses on the Self-Forcing Scale (SCOFF) correlate with a 5.2-fold higher incidence of overuse injuries. Psychological fatigue leads to risk assessment errors, with decision-making errors increasing by 60% after four consecutive hours of training. Social pressure in team sports may induce risky behaviors, as adolescent athletes attempt dangerous moves 35% more frequently under peer observation.

Table 1: Core indicators and monitoring methods for injury risk assessment

|  |  |  |  |
| --- | --- | --- | --- |
| Indicator category | Core parameters | Measurement method | Risk threshold |
|  physiological index  | Bone age difference | Wrist X-ray evaluation | > 1.5 years (advance/late) |
|  | Muscle balance ratio | Isovel strength test | Hamstrings/quadriceps <0.6 |
| Skill indicators | Coefficient of variation of action | 3D motion capture | Key Angle standard deviation>8° |
|  | Technology consolidation index | Coach assessment scale | New movements are learned <20 times |
|  environment pointer  | Hardness of the surface |  Shore durometer  | ASHD value> 90 |
|  | Wind and humidity pressure index (WBGT) |  environment monitoring instrument  | >28℃ |
| Psychological indicators | Focus | Eye tracking gaze duration | <200ms/key stimulus |
|  | Training Stress Scale (TSS) | Psychological questionnaire | > 7 points (out of 10) |

(2) Data acquisition and preprocessing methods

1. Multimodal data acquisition technology

The inertial measurement unit (IMU) sensor captures real-time joint acceleration data at 200Hz sampling rate. A high-risk alert is triggered when the ankle inversion angular velocity exceeds 800°/s during football directional changes. Surface electromyography (sEMG) monitors muscle activation timing, with a time difference greater than 30ms between the medial and lateral thigh muscles indicating abnormal patellar trajectory. Smart textiles embedded with piezoresistive fibers track real-time sole pressure distribution, while peak forefoot pressure exceeding 120% of body weight signals metatarsal stress risk. The computer vision system performs action pattern recognition through 30fps video streams, automatically identifying dangerous knee valgus angles exceeding 10 degrees during long jump landings using deep learning algorithms.

2. Data cleaning and standardization process

The raw data stream undergoes initial outlier filtering using the Tukey rule to automatically identify and remove outliers exceeding 1.5 times the interquartile range[4]. Motion artifact elimination employs wavelet transform technology, reconstructing effective components from accelerometer signals through five-level db4 wavelet decomposition. Sensor data interpolation during disconnection periods utilizes a state space model (SSM) to predict missing electromyographic signal values via Kalman filtering. Multi-device time calibration achieves synchronization of timing data across different sampling rates through dynamic time warping (DTW) algorithm.

3. Application of feature engineering and dimension reduction technology

Temporal feature extraction includes 12 fundamental statistical parameters such as mean, variance, and zero-crossing rate. The knee joint angle acceleration curve requires additional calculation of Jerk (acceleration rate of change). Frequency domain analysis employs Fast Fourier Transform (FFT) to extract main frequency band energy, where the proportion of energy in the 6-8Hz frequency band during gait cycles reflects gait stability. Nonlinear dynamic features encompass sample entropy (SampEn) and Lyapunov index. For long-distance runners, a SampEn value exceeding 0.25 in foot pressure sequences indicates fatigue. Spatial features are compressed through Principal Component Analysis (PCA) of high-dimensional motion trajectory data, retaining no more than 8 principal components that explain 95% of variance after dimensionality reduction of three-dimensional joint coordinates.

(3) Model construction method and verification

1. Machine learning algorithm selection

The random forest algorithm was selected as the benchmark model due to its capability in handling high-dimensional features. The ensemble structure of 500 decision trees effectively prevents overfitting, with feature selection optimized through Gini coefficient-based split nodes. The Support Vector Machine (SVM) employs a radial basis function (RBF) to construct nonlinear classification boundaries, with the regularization parameter C determined through grid search to 0.8. The Gradient Boosting Tree (XGBoost) employs a weak classifier sequence with a maximum depth of 6 and a learning rate of 0.05, terminating training when the validation set loss stagnates for five consecutive rounds using early stopping. For time-series data characteristics, the Long Short-Term Memory Network (LSTM) is configured with 128 hidden units to process continuous 30-frame motion sequences, while a dropout layer maintains a neuron inactivation rate of 0.3 to prevent overfitting.

2. Cross-validation and model optimization strategies

The hierarchical K-fold cross-validation (K=10) ensures each fold maintains the original damage ratio, with each validation process involving random partitioning of 80% training set and 20% validation set. The hyperparameter optimization employs Bayesian optimization algorithm instead of traditional grid search, identifying the optimal parameter combination for random forest within 50 iterations: maximum depth of 15 and feature subset ratio of 0.7. The category imbalance problem is addressed through SMOTE-Tomek joint sampling, with a minority class synthesis ratio set to 300% while removing boundary noise samples defined by Tomeklinks. Model regularization incorporates ElasticNet constraints, with L1 and L2 regularization coefficients set at 0.2 and 0.8 respectively to balance sparsity and stability.

3. Evaluation index system (accuracy, recall rate, F1 value)

The model performance evaluation employs a multi-dimensional metric system. Accuracy reflects overall prediction reliability but is influenced by sample distribution. Recall rate, as the core indicator, focuses on identifying high-risk individuals, requiring a recall rate of no less than 85% for individuals at risk level II or higher. Precision ensures effective early warning while avoiding excessive interference with normal training. The F1 value harmonizes precision and recall rates, achieving a macro F1 score of 0.91 after model integration. The area under the receiver's operating characteristic (ROC) curve (AUC) evaluates overall discriminative power, with the optimal model showing an AUC value of 0.94(95%CI:0.92-0.96. Specificity is maintained above 80%, reducing training interruptions caused by false positives. Differentiated evaluation criteria apply to risk levels: high-risk (Level IV) requires a recall rate>90%, while low-risk (Level I) prioritizes precision>85%.

Table 2: Comparison of risk prediction model performance (mean value of 10-fold cross-validation)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  algorithm  |  precision  |  recall  | Accuracy | F1 value | AUC |
|  random forest  | 0.89 | 0.86 | 0.82 | 0.84 | 0.92 |
| XGBoost | 0.91 | 0.88 | 0.85 | 0.86 | 0.93 |
| SVM(RBF) | 0.85 | 0.81 | 0.79 | 0.80 | 0.87 |
| LSTM | 0.90 | 0.87 | 0.84 | 0.85 | 0.91 |
| Integrated model | 0.92 | 0.90 | 0.88 | 0.91 | 0.94 |

2. Architecture design of intelligent prevention system

(1) Overall system architecture design

1. Edge computing layer

The edge layer deploys a lightweight IoT gateway powered by an ARMCortex-A72 processor for local computation offloading. Motion sensor data is transmitted via BLE5.0 protocol to edge nodes with latency controlled within 20 milliseconds. The local preprocessing module performs critical feature extraction, where raw IMU data (50Hz) undergoes Butterworth low-pass filtering and features are compressed to 15% of original dimensions. A real-time rule engine executes initial risk assessment at the edge, triggering yellow alerts when heart rate exceeds 90% of the maximum age-based heart rate (220-age). A data caching mechanism ensures continuous monitoring during network outages, while local flash memory stores 72 hours of uninterrupted data.

2. Cloud analysis layer

The cloud-based system employs a microservices architecture, with the risk prediction service providing asynchronous API interfaces through RESTful protocols. The model inference cluster utilizes NVIDIAT4GPU acceleration, achieving under 800ms per prediction response. InfluxDB is employed for sensor data streams in the time-series database, delivering 500,000 data points per second. Structured archival data is stored in PostgreSQL relational databases, supported by sharded clusters enabling concurrent access for thousands of users. The batch processing system performs daily reanalysis at midnight, leveraging nighttime computing resources to recalculate model parameters.

3. User interaction layer

The ReactNative cross-platform solution for mobile app development features a centralized display of risk heat maps and alert lists on the coach's app homepage. Athletes can access personalized training logs with corrective animation guides (e.g., standard squat posture demonstrations). The PC console, developed using the Vue.js framework, employs a multi-window layout enabling real-time comparison of biomechanical parameters and environmental monitoring data. The video playback module integrates keyframe annotation functionality to automatically pinpoint technical movement anomalies. The push notification system delivers tiered alerts: regular reminders are sent every 30 minutes, while emergency alerts (such as concussion risks) trigger instant vibration notifications.

(2) Development of core functional modules

1. Real-time risk monitoring and early warning module

The dynamic monitoring system processes real-time data streams of hundreds of athletes per second, utilizing a sliding time window mechanism to detect abrupt short-term risk changes. Its multi-level early-warning system includes: instantaneous threshold alerts (e.g., knee valgus angle>15°), trend accumulation alerts (five consecutive instances of insufficient landing cushioning), and combined condition alerts (fatigue index>70% + ground hardness>90). The warning notification mechanism employs a tiered response strategy: yellow alerts trigger pop-up notifications on the APP, orange alerts notify coaches via SMS, and red alerts activate on-site audio-visual alarms. A false alarm suppression mechanism introduces a confirmation delay, requiring continuous detection of transient anomalies for over three seconds before triggering formal alerts.

2. Personalized training program generation module

The solution engine recommends training content based on knowledge graph analysis. When core strength deficiencies are detected, it automatically links to the "Advanced Plank Training Library". The load calculation model follows Banister's Training Response Theory (TRIMP), dynamically adjusting daily training intensity: <70 during adaptation phase, 70-120 during intensification phase, and <140 during peak phase. The modular exercise design includes a standardized structure: warm-up (dynamic stretching), main training (technical/physical conditioning), and recovery (PNF stretching). Age-specific adaptations are implemented, with prepubescent athletes avoiding strength training exceeding 50% of body weight. The solution outputs structured JSON documents containing action demonstration video links, set range parameters, rest intervals, and other essential elements.

3. Injury emergency response guidance module

The intelligent diagnostic engine performs rapid preliminary screening based on injury mechanisms and symptoms. For instance, a diagnosis of "ankle varus injury with localized swelling" indicates a 76% probability of anterior talofibular ligament injury. The emergency protocol library integrates the POLICE principle (Protect/Apply Appropriate Load/Ice/Compression/Elevate), providing animated guidance for using tourniquets in cases of severe injuries like open fractures. The positioning system coordinates emergency resources, with a one-click call function transmitting the victim's location (accuracy ±3 meters) and preliminary diagnosis to medical centers. The emergency supplies management module monitors first-aid kit status, automatically generating procurement lists when consumables like bandages drop below 20% of inventory.

(III) System integration and interface design

1. Intelligent hardware docking for sports equipment

The Hardware Abstraction Layer (HAL) manages multi-brand device connectivity through unified management, supporting seven communication protocols including ANT+, Bluetooth, and WiFi[5]. Smart running shoe sensors transmit real-time pressure distribution data via customized APIs at 100Hz sampling frequency. Protective gear sensors monitor impact forces, while rugby helmets equipped with built-in accelerometers automatically record collision videos when subjected to impacts exceeding 98g. The sports camera interface supports real-time streaming of GoProHERO series footage through RTMP protocol, transmitting 1080p training videos. The hardware health monitoring system issues alerts for device anomalies, such as calibration reminders triggered by detached electromyography sensor electrodes.

2. Access to third-party health data platforms

The medical data interface complies with the HL7FHIRR4 standard, enabling access to critical data from hospital electronic health records such as growth curves and bone age reports. The school physical examination system utilizes customized middleware to automatically collect annual fitness test results (including vital capacity and sit-and-reach tests). Parent-end health diaries are imported through OAuth2.0 authorization, recording daily data such as sleep quality and nutritional supplementation. The open platform architecture provides standard API suites to support partners in developing value-added applications (e.g., rehabilitation facility booking modules). A data mapping engine resolves heterogeneity in data standards, converting BMI percentile scores from different systems into WHO-standardized values.

3. Multi-party coordination interface between coach, parent and medical staff

The role-based permission system enables refined management. Coaches can view group analytics but have restricted access to medical details, while team doctors possess full biomechanical data permissions. The collaborative workbench supports multi-party online consultations, with video conferencing integrated for 3D motion data sharing. The information release center employs a tiered notification strategy: real-time updates on training plan changes notify coaches immediately, while weekly report summaries are sent to parents every Friday. An emergency contact linkage mechanism automatically alerts three preset contacts (coach, parent, team doctor) during red alert status. Compliance audits track all data access records, with GDPR-compliant data operation logs retained for two years.

 sum up ：

The youth sports injury risk model developed in this study integrates a multidimensional indicator system and employs machine learning algorithms to achieve precise injury risk prediction. The intelligent prevention system adopts a three-tier architecture design with four core functional modules, enabling real-time risk monitoring, personalized interventions, and multi-party collaborative protection. The system application has significantly reduced training injury rates, propelling youth sports training into the era of intelligent management. Future developments will focus on deepening explainable artificial intelligence technology, expanding performance optimization modules, and enhancing system adaptability to special sports scenarios, thereby providing technological support for the sustainable development of young athletes.

 reference documentation ：

[1] Li Shangjin. Research on sports Injury and Prevention Strategies in Adolescent Sports Training [J]. Boxing and Fighting, 2025, (11):116-118.

[2] Lu Xiaohong, Lu Haotian. Strategies for the Prevention of Sports Injuries in Adolescent Physical Training [J]. Boxing and Fighting, 2025(09):113-115.

[3] Yu Xiang, Yang Yun, Wei Gongbo, et al. Exploring the Value of AI in Sports Teaching and Training: A Study Based on Sports Injury Modeling [J]. Sports World, 2025(03):15-19+23.

[4] Chen Baoqiang. Research on sports injury prevention in youth sports training [J]. Boxing and Fighting, 2024, (21):102-104.

[5] Sun Huafei. Early-warning model for joint injuries in youth basketball training [J]. Journal of Changchun University, 2024,34(04):32-38.