Machine Learning-based Prediction of Running Injuries: Pooled Analysis of MotionMetrix Data and Screening Tests

Miss LI Mei Hung, BSocSc (Hons) in Sports and Recreation Management,
Department of Sport and Recreation
Supervisor: Mr Indy HO, Assistant Professor and Dr Jim LUK , Associate Professor

BACKGROUND

Running-related injuries (RRI), particularly overuse injuries, are a significant concern for runners. Relying solely on traditional measures like running volume oversimplifies the assessment of training stress, often neglecting important factors such as ground reaction force and foot-strike pattern.

Instead of the linear and unidirectional causality view of sports injury etiology, the complex system perspective proposed a multifactorial nature of injuries, emphasizing unknown interactions and varying weights among determinants.

PURPOSES

The current study integrated survey data, screening tests, and joint kinematics and kinetics results from the novel markerless running assessment system, MotionMetrix:

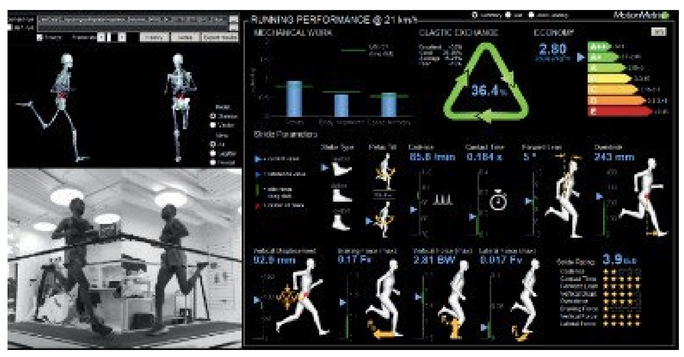
1. To identify the high-importance features, and
2. To develop machine learning-based predictive model for assessing running injury risk (yes or no).

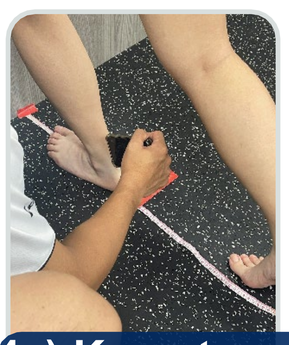
METHODS

A Quantitative Research Design with A Retrospective Approach

59 trained long-distance runners (44 males and 15 females) were recruited using purposive and convenience sampling methods.

1. **Survey**
2. **Warm-up**
3. **MotionMetrix Running Assessment**
   - 1-minute
   - 12 km/h speed, 0% inclination
   - Joint kinematics and kinetics data



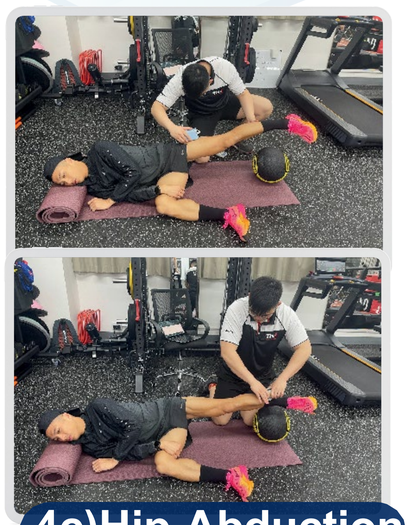


4a) Knee-to-wall Ankle Dorsiflexion Test

- Record distance with tape
- iLevel app measures angles
- Average values of 3 trials

4b) Single Leg Squat Test

- Identifies movement deviation
- Video recording of 3 trials
- Scoring system



4c)Hip Abduction Strength Test

- 5s isometric maximum voluntary contraction of hip abduction
- Hand-held dynamometer record the highest value of 3 trials

4. **Screening Tests**
5. **Development of Machine Learning Algorithms**

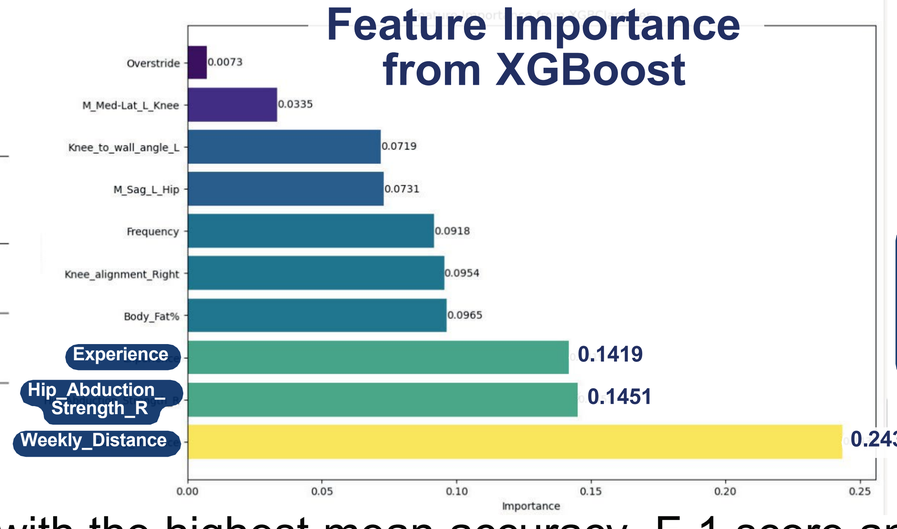
RESULTS

Comparison of Performance of Three Prediction Models

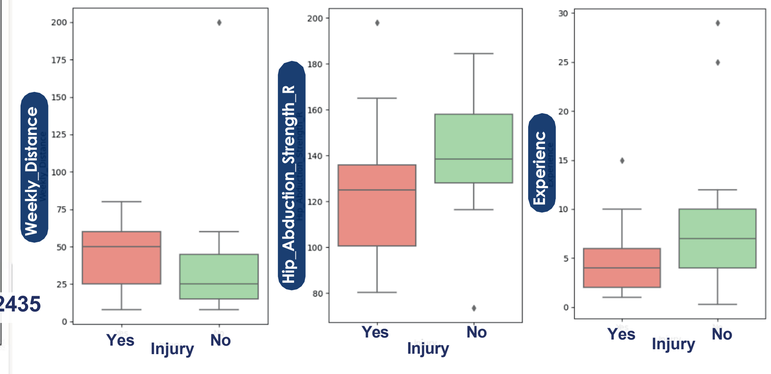
Classifier	Features Selected	Mean Accuracy	Mean F1-score	Mean AUC
Random Forest	14	0.667	0.758	0.620
SVM	48	0.617	0.757	0.500
XGBoost	10	0.717	0.759	0.735

Note. AUC: Area under the ROC curve.

Feature Importance from XGBoost



Distribution of Features on Injury Status



XGBoost model outperformed others with the highest mean accuracy, F-1 score and AUC.

DISCUSSION

- The best predictive performance does not necessarily reveal causality; it requires an integrated comprehensive understanding of underlying mechanism with different features.

Important Features

- Weekly Running Distance is the most important feature. As running distance increased and exposure time lengthens, the accumulated load progressively raises the risk of injury overtime.
- Weak Hip Abduction Strength is linked to altered running mechanics and increased risk of RRI, especially iliotibial band syndrome.
- Runner with high years of running Experience enjoy a protective effect, may experience a lower injury rate due to adaptations, better load management, and injury awareness.

PRACTICAL APPLICATION

- The developed prediction model allows for the assessment of RRI risk, providing insights into individual likelihood of injury.
- In our sample population, interventions such as training load control, running gait training, and hip abductor strengthening could effectively reduce injury risk.
- A valuable tool for coaches and athletes in making informed decisions regarding injury prevention.