Machine Learning-Based Pattern Recognition of Risk Factors for Low Back Pain among Adolescent Cricket Players in Dhaka City

Authors

Marzana Afrooj Ria[1], Tasrima Trisha Ratna[2], Shudeshna Chakraborttye Purba[3], Rubal Kar[4], Dr. Mohoshina Karim[5], Md Osman Ali[6], Erfat Jaren Chaity[7], Shahadath Hossen[6], Joynal Abedin Imran[1]*, Shahriar Hasan[8]

[1]Department of Physiotherapy, National Institute of Traumatology & Orthopaedic Rehabilitation (NITOR), Dhaka-1207, Bangladesh; Email: marzanaafrooz@gmail.com (MAR)

[2]Shahabuddin Medical College Hospital, Dhaka, Bangladesh; Email: tasrimatrisha@gmail.com (TTR)

[3]Department of gastroenterology, Bangladesh medical college, Dhaka, Bangladesh; Email: dr.sudeshnapurba@gmail.com (SCP)

[4]Department of Physiotherapy, Gono Bishwabidyalay, Savar, Bangladesh; Email: physio.rubal@gmail.com (RK)

[5]Institute of Nutrition & Food Science, University of Dhaka, Bangladesh; Email: kekakarim@yahoo.com (MK)

[6]Department of Electrical and Electronic Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh; Email: osman.eee@nstu.edu.bd (MOA), shahadathhossen001@gmail.com (SH)

[7]Department of Pharmacy, East West University, Dhaka, Bangladesh; Email: erfatjaren2023@gmail.com

[8]Department of Public Health, North South University, Dhaka, Bangladesh; Email: shahriar372@gmail.com (SH)

Corresponding author: * 4abedin@gmail.com (JAI); https://orcid.org/0009-0002-7237-7715

Abstract

Low back pain (LBP) is a frequent occurrence in adolescent cricketers and is most predominantly brought about by repetitive lumbar region stress. The purpose of this study is to investigate LBP in adolescent cricketers in Dhaka City in Bangladesh with machine learning through the Support Vector Machine (SVM) model and regression analysis to assist in classifying LBP severity and factors contributing to it. Data was analyzed for 315 adolescent cricketers in this study and includes sociodemographic factors, game-related activities, preventive measures, and history of LBP-related factors. The severity of LBP was classified into four: no pain, mild, moderate, and severe. The sigmoid kernel exhibited best classification performance with a precision value of 78.9%, recall value of 79.7%, and F1 value of 79.2%, specifically performing well in separating between no-pain and mild-pain cases. Significant statistical correlations were evident between LBP and variables such as age, educational background, family income, duration of practice, warm-up and cool-down protocol, and past history of low back pain. These findings emphasize the importance of early preventive measures in lessening LBP risk among young cricket players. Generally, this study demonstrates effectiveness in machine learning and regression model in detecting injury pattern risks in sports environments and supporting data-driven prevention and management of sports injuries and providing a baseline model for future studies in the area.

Keywords: Low Back Pain, Adolescent Cricketers, Machine Learning, Support Vector Machine, Injury Prevention, Bangladesh

Introduction

Low back pain in adolescent sports players, particularly in cricket when repetitive stress to the spine increases the risk factor, is a cause of growing concern. Adolescents are most susceptible since they are in the development phase and are also involved in sports [1]. Repeated spinal movements in a game such as cricket is responsible for LBP in this case, with a rate of up to 30% of adolescents affected [2]. Women have a greater incidence of LBP compared to men [3]. In spite of its prevalence, effective prevention is limited and LBP's complicated nature, resulting from both biomechanical and psychosocial etiologies, hampers detection and intervention at onset [4].

In cricket, there is severe mechanical stress on the lumbar spine through repeated hyperextension, rotation, and lateral flexion, significantly raising the risk of LBP [5,6]. Immature spines in junior bowlers are especially susceptible [6]. Repeated acts of bowling have resulted in structural spine injuries such as spondylolysis through cyclic extension loading [7], and therefore cricket is also a high-risk sport for LBP. This stresses the urgent necessity of prevention and monitoring of biomechanics in cases of injury. LBP in young athletes can limit physical functioning, hinder return to sport, and have a detrimental effect on emotional well-being and identity. Adolescents normalize pain and this leads to underreporting and undertreatment with resulting damaging effects. LBP also strains peer and coach relationships [2]. Recurrent LBP in development years can contribute to adulthood risk of chronic pain and sick leave from school [1]. Thus, early and effective treatment is necessary to preserve both sports competence and future health [8].

Early detection of young athletes with potential LBP risk factors is essential to guarantee effective prevention and interventions [1]. Early detection of signs of LBP allows coaches

and medical professionals to establish personalized training and rehabilitation programs to reduce injury [4]. Isokinetic testing and electromyography are good screening instruments in identifying at-risk individuals [4]. Interventions made in a timely manner can avoid long-term consequences such as persistent pain and impaired performance [2].

More so, traditional manual protocols to detect injury-risk factors in athletes are limited in capacity and cannot detect complex patterns [9]. Contrary to this, machine learning (ML) can utilize a data-driven approach to identify hidden and risk factors through making sense of large volumes of big data information such as training history, biomechanics, and psychological factors that cannot be easily accumulated through manual methodologies [10]. This methodology is best applied to instances like adolescent cricket players of Dhaka City in which there is a lack of studies on lower back pain (LBP) associated factors [6]. The cricket biomechanics in general and fast bowling in particular account for increased LBP risk [11]. Risk factors are stratified through the use of ML models so prediction of the probability of LBP is possible and health professionals are able to forecast risks and individualize prevention and rehabilitation strategies [9,10].

There are no known predictive models with ML to be used in screening LBP in adolescent cricketers in Dhaka. The aim of this study is to fill this gap with ML and regression analysis to define important risk factors in the setting of Bangladesh. The findings will inform health professionals, coaches, and sports managers about best prevention and management strategies for LBP in young cricketers. This study can also serve to be a model for future studies in sports injury by ML applications in other regions or other sports.

Materials & Methods

Dataset Preparation

The dataset size was estimated based on previous studies on lower back pain (LBP) studies, where analyses were conducted on the same dataset comprising 310 patient observations involving various physical spine-related features [12,13,14]. The dataset included binary class labels, with 100 normal and 210 abnormal cases, and 12 medical attributes recorded for each patient. These attributes included pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, degree spondylolisthesis, pelvic slope, direct tilt, thoracic slope, cervical tilt, sacrum angle and scoliosis slope.

Study population

For this study, data were collected from 315 adolescent cricket players across various cricket clubs (BKSP, Lt. Sheikh Jamal cricket Academy, Kola Bagan Krira chakra, Khelaghar Cricket Academy and City club) in Dhaka City, with the primary focus on identifying the risk factors associated with the development of LBP. The data collection was conducted from 25 February 2024 to 28 November 2024.

Ethical consideration

This study adhered to the ethical principles of the Declaration of Helsinki. This study was also approved by the IRB of the Institutional Review Board (IRB) of the National Institute of Traumatology and Orthopaedic Rehabilitation (NITOR/PT/93/lRB/2024/05). The study's purpose, methodology, and procedures were fully known to the participants. The participants were entitled to withdraw the interview completely at any time. Confidentiality and anonymity were ensured. For this study, no physical specimens were taken. Data were collected only through the standardized questionnaire.

Outcome variables

To facilitate the data collection and classification process, LBP severity was categorized into four distinct classes, such as Class 0: No pain (No LBP), Class 1: Mild pain, Class 2: Moderate pain, Class 3: Severe pain. Notably, there are 49 instances were recorded for Class 2 (Moderate pain) and 9 instances for Class 3 (Severe pain), which indicated a class imbalance, with the majority of participants reporting either no pain or mild pain. Specifically, 112 instances were categorized as Class 0 (No pain), while the remaining 203 instances were categorized as Class 1 (Mild pain).

Explanatory variables

The dataset for includes a variety of features that captured the demographic, physical, and training-related characteristics of adolescent cricket players. Data were gathered through structured surveys and physical assessments, and were organized into the following categories:

- Sociodemographic factors: Age, gender, education, monthly family income, weight, and body mass index.
- Games related Factors: Playing position, playing experience, and duration of practice per week.
- Preventive measures factors: Warmup before sports activity, duration of warmup, cooldown after sports activity, and duration of cool down.
- Low back Pain-related Features: Information on LBP occurrence (present/absent), pain severity (mild, moderate, severe), and previous LBP occurrences.

Each of these variables captures distinctive features that offer a full insight into what is involved. These variables are detailed and described at length in this next section with emphasis on their individual contributions to this study.

Player age was between 11 years and 19 years. As a demographical factor of significance, age is a distinguishing point in identifying players' physical development phase since younger players might have less musculoskeletal development and would be at greater risk of injuries such as LBP. Height was recorded between 120 cm and 180.3 cm and is useful in identifying players’ height, their physical stature affecting their biomechanics in cricketing activities. The same goes for their weight, between 20 kg and 80 kg, with weight also a critical point in identifying strain on musculoskeletal systems in activities of high intensities like cricket.

The players' Body Mass Index (BMI) varied between 8.53 and 29.81. This measure was split into three categories: underweight (BMI < 18.5), normal (BMI 18.5 - 24.9), and overweight (BMI ≥ 25), and it had implications for the overall risk of injury to the player. The educational level was divided into four categories: J.S.C./8th grade, S.S.C./O-levels, H.S.C./A-levels, and undergraduate. This variable is important because education might indicate underlying socio-economic factors affecting players' overall health and physical activity.

Position of play was another factor of significance, following categories such as right-hand batsman, left-hand batsman, spin hand bowler, pace hand bowler, wicket-keeper, and all-rounder. These roles defined the type of biomechanic stress faced by the players, with bowlers and especially pace bowlers at a greater risk of developing LBP due to their repetitive movements. The players were classified in terms of their playing experience among three categories, 1-3 years, 4-6 years, and 7-10 years. The players' practice duration per week (hour) was classified in three categories viz. <10 hours, 10-20 hours, and >20 hours.

Training habits are captured through the doing and duration of warmup and cooldown. The players were doing the warmup before the sports activity and cooldown after the sports activity was categorized into three categories which were always, often and sometimes. The warmup duration ranges from <10 minutes to >15 minutes, while cooldown time is similarly categorized into <10 minutes, 10-15 minutes, and >15 minutes. These factors are critical as appropriate warmup and cooldown routines are known to reduce the risk of injury and improve muscle recovery.

Lastly, past history of low back pain is documented either “Present” or “Absent” because it is known that players with a history of LBP are at greater risk of recurrences of this type of pain and are thus an important factor in this study.

In brief, the machine learning model for LBP severity classification was established based on an integrated group of attributes such as sociodemographic factors, game factors, and preventive and LBP pain related factors. The work also sought to establish major contributing factors to LBP in adolescent cricket players in Dhaka City. The performance of the model was critically tested through a 10-fold cross-validation.

Support Vector Machine

The most popular classification models are machine learning-based K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM), and deep learning-based Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN). The machine learning-based model was chosen in this study because it is good at processing structured data in contexts with a small number of samples. The deep learning models are in need of a large volume of data to automatically extract and learn intricate feature representation efficiently [16].

Of the above-mentioned machine learning algorithms, KNN is a non-parametric model whereas RF and SVM have good classification properties and are suitable for non-linear decision boundary modelling. SVM was chosen for this study because it is a human-interpretable model and has good generalizing capacity on small data samples and is less susceptible to overfitting in comparison to ensembled-based models such as RF [17,18,19].

SVM is a supervised learning machine learning algorithm with excellent performance in multiclass classification even when it has been trained on relatively small data sets. SVM generates one or more optimal separating decision boundaries in high-dimensional feature space called hyperplanes to classify data into their individual classes. The hyperplanes are such that they are maximally positioned between classes. In other words, the hyperplanes are put in such a way that they are maximally distant from data points of any class. These data points are called support vectors. Maximizing this distance between classes is what aims to help improve the model's performance in generalizing new data [17].

Several mathematical functions known as kernels are employed to deal with non-linearly separable data like linear kernel, polynomial kernel, radial basis function (RBF), and sigmoid. These enable us to map the original data to a feature space of a higher dimension so that it is linearly separable. Mathematical representation of these kernels is provided in Table 1 and a representation of SVM-based decision boundaries for these kernels is shown in Fig.1.

Table 1. SVM kernel functions and their mathematical expressions.

Kernel Type	Expression
Linear	$x_i^T x_j$
Radial basis function (RBF)	$exp\left(\frac{\|\|x_i - x_j\|\|^2}{2\sigma^2}\right)$
Polynomial	$\left(x_i^T x_j + c\right)^p$
Sigmoid	$\tanh\left(x_i^T x_j + c\right)$

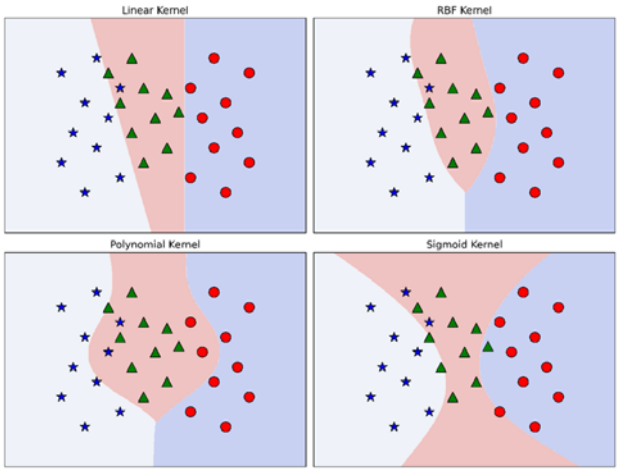


Fig. 1. An illustration of SVM decision boundaries using different kernel functions.

Result

The research result was analyzed by two different types of analytical method, among them one was machine learning based classification result which included SVM model. Another analytical model was binary logistic regression model.

Machine learning-based classification

An SVM model was trained to classify the LBP levels into four categories, where Class 0 indicates a no-pain state and Class 3 presents a severe pain state. The model's performance was evaluated in the Visual Studio Code environment with Python version 3.9.13. The classifier's effectiveness was assessed using commonly used evaluation metrics, including accuracy, precision, recall, and F-1 score, which is also necessary to comprehensively analyze the classifier's performance in the presence of class imbalance. These metrics are calculated based on the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions, as defined below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

To optimize model performance, hyper-parameter tuning was carried out, and the best results were achieved using a sigmoid kernel with a regularization parameter of C = 1. This configuration achieved a precision of 78.9%, recall of 79.7%, and F1-score of 79.2%. The next best performance was observed with the RBF kernel, which achieved the highest

precision of 82.6%. A performance comparison of the best results achieved across different kernel functions is presented in Table. 2.

Table 2. SVM classification performance across kernels using 10-fold cross-validation.

Kernels	**Precision**	**Recall**	**F1-Score**
RBF	**0.826**	0.755	0.788
Linear	0.688	0.705	0.695
Sigmoid	0.789	**0.797**	**0.792**
Poly	0.640	0.514	0.567

The heat map of the confusion matrix for the best-performing SVM model is shown in Fig. 2. The result indicated that the model performs well in identifying no-pain and mild-pain classes. Specifically, all 11 instances of *Class 0* (no pain) were correctly classified, while 16 out of 17 instances of *Class 1* (mild pain) were correctly identified, with only one misclassified. In contrast, the model shows comparatively limited effectiveness in categorizing moderate and sever pain levels. This performance variation might be attributed to class imbalance and overlapping feature distributions among higher pain levels.

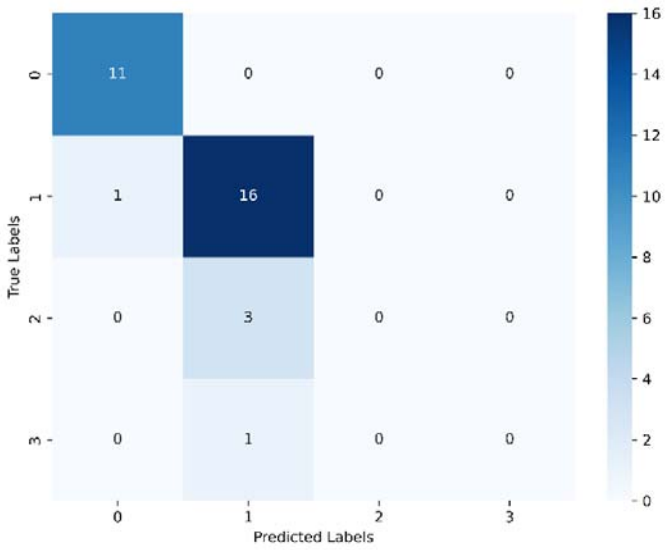


Fig. 2. Heatmap of confusion matrix of the SVM classifier using the sigmoid kernel (C=1) for LBP classification.

Fig. 3. ROC curve of the SVM classifier using the sigmoid kernel (C=1) for LBP classification.

The ROC curve of the SVM classifier is presented in Fig. 3, reflecting its performance in distinguishing between four classes of low back pain: No pain, Mild pain, Moderate pain, and

Severe pain. The curve initially rises steeply, indicating high sensitivity with minimal false positives, which is ideal for early and accurate detection. As the False Positive Rate increases, the curve becomes more gradual and eventually flattens, suggesting the classifier faces difficulty distinguishing closely related pain levels, likely due to overlapping features. Despite this, the curve remains well above the random guess line, achieving a macro-average AUC of 0.829, which indicates solid overall classification performance. This level of accuracy suggests the SVM model can effectively support risk pattern recognition in adolescent cricket players. While some false positives may occur in later stages, the system remains reliable for early screening and can contribute meaningfully to preventive healthcare strategies in sports.

Sociodemographic, games and preventive measures related characteristics of the participants

Table 3 demonstrated the significant associations between low back pain (LBP) and sociodemographic characteristics among adolescent cricket players. A significant relationship was observed between age and LBP (P = 0.018), with older adolescents (17 to 19 years) showed a higher prevalence of LBP (72.8%) compared to younger participants (51.1%). Education level, however, exhibited a strong and significant association with LBP (P<0.001). The participants having higher educational levels (Undergraduate) showing a greater prevalence of LBP (37.4%) compared to those with lower educational attainment. Additionally, monthly family income was significantly associated with LBP (P<0.001), with the income group (20,000-40,000 BDT) showed the highest proportion of LBP (44.3%).

Table 3. Distribution of sociodemographic characteristics of the participants (n=315)

Characteristics	**Low back pain**		**Chi-square**	***P*-value**
	Present (%) (*n* = 203)	**Absent (%) (*n* = 112)**		
Age				
11 to 13 years	24 (51.1)	23 (48.9)	8.008	0.018[a]
14 to 16 years	88 (61.5)	55 (38.5)		
17 to 19 years	91 (72.8)	34 (27.2)		
Gender				
Male	140 (64.2)	78 (35.8)	0.016	0.901[b]
Female	63 (64.9)	34 (35.1)		

Characteristics	Low back pain		Chi-square	*P*-value
	Present (%) (*n* = 203)	Absent (%) (*n* = 112)		
Education				
J.S.C/8th grade	17 (8.4)	31 (27.7)	37.780	0.000[a]
S.S.C/O-levels	58 (28.6)	44 (39.3)		
H.S.C/A-levels	52 (25.6)	24 (21.4)		
Under graduation	76 (37.4)	13 (11.6)		
Monthly family income (BDT)[c]				
Below 20000 BDT	20 (9.9)	39 (34.8)	33.91	0.000[b]
20000-40000 BDT	90 (44.3)	47 (42.0)		
41000-60000 BDT	63 (31.0)	17 (15.2)		
>60000 BDT	30 (14.8)	9 (8)		
Weight				
Below 40 kg	4 (2.0)	6 (5.4)	4.40	0.111[b]
40-59 kg	169 (83.3)	83 (74.8)		
60 kg and above	30 (14.8)	22 (19.8)		
Body Mass Index				
Underweight	36 (17.9)	29 (26.1)	3.43	0.180[b]

Characteristics	Low back pain		Chi-square	*P*-value
	Present (%) **(*n* = 203)**	**Absent (%)** **(*n* = 112)**		
Normal	155 (77.1)	75 (67.6)		
Overweight	10 (5.0)	7 (6.3)		
[a]Pearson Chi-square; [b]Fisher's exact correction test used for having 20% or more expected frequencies less than 5; [c]1 US dollar = 121.85 BDT (in May 2025)				

Table 4 indicated several notable relationships between game-related characteristics and the presence of low back pain (LBP) among adolescent cricket players. Playing position did not show a significant association with LBP (P = 0.476), suggesting that different playing roles, including right-hand batsman, left-hand batsman, spin bowler, pace bowler, wicket keeper, and all-rounder, do not contribute differently to the likelihood of experiencing LBP. Similarly, playing experience also was not significantly associated with LBP (P = 0.072), although there was a trend suggesting that players with 1-3 years of experience had a higher prevalence of LBP (41.8%) compared to those with more experience. However, a significant relationship was observed between the duration of practice per week and LBP (P<0.001). Players who practiced > 20 hours per week had the highest prevalence of LBP (79.8%), while those who practiced <10 hours per week had the lowest prevalence (55.3%).

Table 4. Distribution of games related characteristics of the participants (n=315)

Characteristics	Low back pain	Chi-	*P*-

	Present (%) (*n* = 203)	Absent (%) (*n* = 112)	square	value
Playing position				
Right hand batsman	41 (30.8)	92 (69.2)	4.52	0.476[a]
Left hand batsman	12 (37.5)	20 (62.5)		
Spin hand bowler	15 (48.4)	16 (51.6)		
Pace hand bowler	7 (29.2)	17 (70.8)		
Wicket keeper	7 (36.8)	12 (63.2)		
All rounder	30 (39.5)	46 (60.5)		
Playing experience				
1-3 years	64 (41.8)	89 (58.2)	5.26	0.072[a]
4-6 years	38 (30.9)	85 (69.1)		
7-10 years	9 (25.7)	26 (74.3)		
Duration of practice per week				
< 10 hours	21 (55.3)	23 (20.2)	18.44	0.000[a]
10-20 hours	91 (55.8)	72 (44.2)		
> 20 hours	91 (79.8)	17 (44.7)		

[a]Pearson Chi-square; [b]Fisher's exact correction test used for having 20% or more expected frequencies less than 5

Table 5 highlighted the significant role of preventive measures in the occurrence of low back pain (LBP) among adolescent cricket players. A strong association was found between warming up before sports activity and LBP (P<0.001), with those who always warm up having a lower prevalence of LBP (56.6%) compared to those who warm up often (92.9%) or "sometimes" (96.0%). A similarly strong association was observed between cooldown after sports activity and LBP (P<0.001), with those who always perform a cooldown having a lower prevalence of LBP (54.8%) compared to those who often (87.5%) or sometimes (85.4%) engage in cooldown exercises.

Table 5. Distribution of preventive measure related characteristics of the participants (n=315)

Characteristics	**Low back pain**		**Chi-square**	***P*-value**
	Present (%) (***n*** **= 203)**	**Absent (%)** (***n*** **= 112)**		
Warmup before sports activity				
Always	142 (56.6)	109 (43.4)	33.44	0.000[a]
Often	13 (92.9)	1 (7.1)		
Sometimes	48 (96.0)	2 (4)		
Duration of warmup				
<10 min	6 (66.7)	3 (33.3)	5.54	0.063[a]
10-15 min	135 (69.2)	60 (30.8)		
>15 min	62 (55.9)	49 (44.1)		
Cooldown after sports activity				

Characteristics	Low back pain		Chi-square	*P*-value
	Present (%) (*n* = 203)	Absent (%) (*n* = 112)		
Always	119 (54.8)	98 (45.2)	28.11	0.000[a]
Often	14 (87.5)	2 (12.5)		
Sometimes	70 (85.4)	12 (14.6)		
Duration of cooldown				
<10 min	21 (84)	4 (16)	0.64	0.725[a]
10-15 min	151 (66.5)	76 (33.5)		
>15 min	31 (49.2)	32 (50.8)		

[a]Pearson Chi-square; [b]Fisher's exact correction test used for having 20% or more expected frequencies less than 5

Logistic regression to predict risk factors of low back pain

Table 6 presented the logistic regression analysis and this table revealed several key risk factors for low back pain (LBP) among adolescent cricketers. In the regression model, age and education level were significant predictors of LBP, with younger age groups (11-13 and 14-16 years) showing lower odds (OR=0.59 and 0.38) of experiencing LBP compared to the 17-19 years age group. Education level remained a significant factor, with participants with higher education (Undergraduate) exhibiting substantially higher odds of LBP (OR = 11.51, 95% CI: 2.92-45.36, P = 0.000). In terms of practice duration, while those practicing 10-20 hours per week showed lower odds of LBP in the unadjusted model (OR = 0.31, P = 0.004).

Warm-up practices before the sports activity were found to be a significant predictor of LBP, with players who “sometimes” warmed up showed higher odds (OR = 12.72, 95% CI= 2.37-68.20, P = 0.003). Similarly, cooldown after the sports activity also showed significant associations with LBP, with “sometimes” performing a cooldown associated with higher odds of LBP in the (OR = 4.80, CI= 2.46-9.37, P = 0.000). Participants who performed a cooldown for less than 10 minutes had significantly higher odds of experiencing LBP (OR = 5.41, 95% CI: 1.66-17.59, P = 0.005) compared to those who cooled down for more than 15 minutes. Previous history of LBP was a consistent and strong predictor of current LBP, with participants who had a prior history of LBP showing significantly higher odds in both unadjusted (OR = 7.28, 95% CI=4.33-12.24, P = 0.000) and adjusted models (OR = 6.62, 95% CI= 3.58-12.22, P = 0.000).

Table 6. Logistic regression for sociodemographic, games and preventive measures related risk factors of low back pain among adolescent cricketers (n=315).

Variables	**Category**	**Low back pain**			
		Absent vs Present			
		Unadjusted model		**§Adjusted model**	
		OR 95% CI	**p-value**	**OR 95% CI**	**p-value**
Age	17 to 19 years	Reference			
	14 to 16 years	0.38 (0.19-0.78)	0.008	1.21 (0.56-2.59)	0.617
	11 to 13 years	0.59 (0.35-1.00)	0.052	2.52 (0.76-8.31)	0.127
Education	J.S.C	Reference			

Variables	Category	Low back pain			
		Absent vs Present			
		Unadjusted model		§Adjusted model	
		OR 95% CI	p-value	OR 95% CI	p-value
	S.S.C/O-levels	2.40 (1.18-4.88)	0.015	2.65 (0.97-7.19)	0.055
	H.S.C/A-levels	3.95 (1.84-8.48)	0.000	4.45 (1.42-13.94)	0.010
	Under graduation	10.66 (4.62-24.55)	0.000	11.51 (2.92-45.36)	0.000
Practice duration per week (hour)	> 20 hours	Reference			
	10-20 hours	0.31 (0.18-0.55)	0.000	0.78 (0.38-1.61)	0.514
	< 10 hours	0.31 (0.14-0.68)	0.004	1.05 (0.39-2.84)	0.912
Warmup before the sports activity	Always	Reference			
	Often	9.97 (1.28-77.45)	0.028	5.61 (0.55-56.39)	0.143
	Sometimes	18.42 (4.38-77.47)	0.000	12.72 (2.37-68.20)	0.003
Warmup duration	>15 mins	Reference			
	10-15 mins	1.77 (1.09-2.88)	0.019	0.69 (0.36-1.31)	0.264

Variables	Category	Low back pain			
		Absent vs Present			
		Unadjusted model		§Adjusted model	
		OR 95% CI	p-value	OR 95% CI	p-value
	<10 mins	1.58 (0.37-6.64)	0.532	1.17 (0.17-7.84)	0.867
Cooldown after the sports activity	Always	Reference			
	Often	5.76 (1.27-25.97)	0.023	3.49 (0.60-20.15)	0.161
	Sometimes	4.80 (2.46-9.37)	0.000	1.40 (0.57-3.44)	0.457
Cooldown duration	>15 mins	Reference			
	10-15 mins	2.05 (1.16-3.61)	0.013	1.20 (0.58-2.46)	0.609
	<10 mins	5.41 (1.66-17.59)	0.005	1.25 (0.24-6.27)	0.786
Pervious history of low back pain	Absent	Reference			
	Present	7.28 (4.33-12.24)	0.000	6.62 (3.58-12.22)	0.000

Discussion

This research offers valuable findings regarding risk factors of low back pain (LBP) in adolescent cricket players of Dhaka City, applying machine learning to predictive modeling.

The results highlighted the multi-factorial cause of LBP, stressing interaction between sociodemographic, game-related, and prevention-related factors.

In this research it was found that the prevalence of LBP increases with age, with adolescents aged 17-19 years reporting higher rates compared to those aged 11–13 years. In a systematic review and meta-analysis encompassing 80 studies, the pooled prevalence of LBP in adolescent athletes was found to be 42% over the past 12 months, with older adolescents exhibiting higher prevalence rates.

Similarly, in a study utilizing inertial measurement units to investigate biomechanical risk factors of low back pain in cricket fast bowlers, researchers found that 38% of junior bowlers (mean age 16.9 years) reported a history of low back pain, compared to 57% of senior bowlers (mean age 24.1 years) [6]. This higher prevalence among older players suggests that low back pain risk may increase with age, potentially due to cumulative load or changes in bowling biomechanics.

In this study, education level showed a strong and significant association with LBP ($P<0.001$). The participants having higher educational levels (Undergraduate) showing a greater prevalence of LBP (37.4%) compared to those with lower educational attainment. But a Brazilian study of 1,455 adolescents found that those who did not attend physical education classes had 2.26 times higher odds of experiencing LBP compared to participants [20].

In this study it was found that players who practiced > 20 hours per week had the highest prevalence of LBP (79.8%), while those who practiced <10 hours per week had the lowest prevalence (55.3%).

Similarly, in a research, adolescent basketball players practicing over 6.6 hours/week exhibited 43.8% LBP prevalence [21] and youth weightlifters training long-term

weightlifting developed lumbar disc degeneration within 4 years, progressing to herniation in 33% by year 5 [22].

A strong association was found between warming up before sports activity and LBP ($P<0.001$), with those who always warm up having a lower prevalence of LBP (56.6%). Similarly, a meta-analysis of Warm-up Intervention Programs (WIPs) found a 36% reduction in sports injuries (pooled injury rate ratio of 0.64, 95% CI = 0.54–0.75) when accounting for hours of risk exposure [23].

A strong association was observed between cooldown after sports activity and LBP ($P<0.001$), with those who always perform a cooldown having a lower prevalence of LBP (54.8%). Similarly, a randomized crossover study compared aqua- and land-based cooldown exercises found both have similar recovery effects on muscle soreness and performance-based parameters [24].

Previous history of LBP was a consistent and strong predictor of current LBP, with participants who had a prior history of LBP showing a higher odd of presenting low back pain (OR = 7.28). A comprehensive systematic review investigating the risk of LBP recurrence found that "a history of previous episodes of LBP prior to the most recent episode was the only factor that consistently predicted recurrence of LBP" [25]. Another systematic review also discovered intrinsic factors related to LBP in cricket fast bowlers to include previous injury as a factor known to influence load tolerance and LBP injury [26].

The application of machine learning here is a useful sports medicine innovation. Methods of traditional risk factor analysis cannot account for such complex, non-linear interdependencies amongst a large number of variables such as sociodemographic, game and prevention-of-injury factors and LBP factors.

Conclusion

The aim of this study was to perform a risk assessment and classification of LBP in adolescent cricketers in Bangladesh. This was in an effort to fill this knowledge gap through machine learning techniques to derive important risk factors in this population. A regression model was suggested to predict LBP risk factors and a support vector machine (SVM) model to classify. Many variables were considered including sociodemographic, game-related, preventive measure, and LBP-related variables to evaluate model performance. Individually, the findings were useful to health professionals, coaches, and sport organizations in designing effective interventions to prevent and treat LBP in young players. This study also gave a starting point for future studies conducted to investigate sports-related injuries with machine learning algorithms and regression analysis in other regions and sports disciplines.

Availability of data and materials

All supporting data are available on request from the authors.

Acknowledgement

The authors would like to thank coaches and players of BKSP, Lt. Sheikh Jamal cricket Academy, Kola Bagan Krira chakra, Khelaghar Cricket Academy, and City club for all assistance and cooperation throughout the study.

Supporting information

S1 Appendix. Informed consent form. The written consent form was administered to the participants.

S2 Appendix. Questionnaire. The included questions comprised the questionnaire administered to the participants.

S3 Appendix. Data. This file contains research data.

Author Contribution

Conceptualization: Marzana Afrooj Ria, Tasrima Trisha Ratna, Shudeshna Chakraborttye Purba, Rubal Kar, Dr. Mohoshina Karim, Md Osman Ali, Erfat Jaren Chaity, Shahadath Hossen, Joynal Abedin Imran, Shahriar Hasan

Data Curation: Erfat Jaren Chaity, Shahriar Hasan

Formal Analysis: Tasrima Trisha Ratna, Rubal Kar

Investigation: Marzana Afrooj Ria, Shahriar Hasan

Methodology: Marzana Afrooj Ria, Md Osman Ali, Shudeshna Chakraborttye Purba

Project Administration: Dr. Mohoshina Karim, Joynal Abedin Imran

Resources: Rubal Kar, Erfat Jaren Chaity, Dr. Mohoshina Karim

Software: Md Osman Ali, Shahadath Hossen, Joynal Abedin Imran

Supervision: Marzana Afrooj Ria, Shahriar Hasan, Shudeshna Chakraborttye Purba

Validation: Tasrima Trisha Ratna, Shudeshna Chakraborttye Purba

Visualization: Md Osman Ali, Shahadath Hossen, Rubal Kar

Writing – Original Draft: Marzana Afrooj Ria, Tasrima Trisha Ratna

Writing – Review & Editing: Dr. Mohoshina Karim, Shahadath Hossen

Funding

No specific funding was allotted for this study.

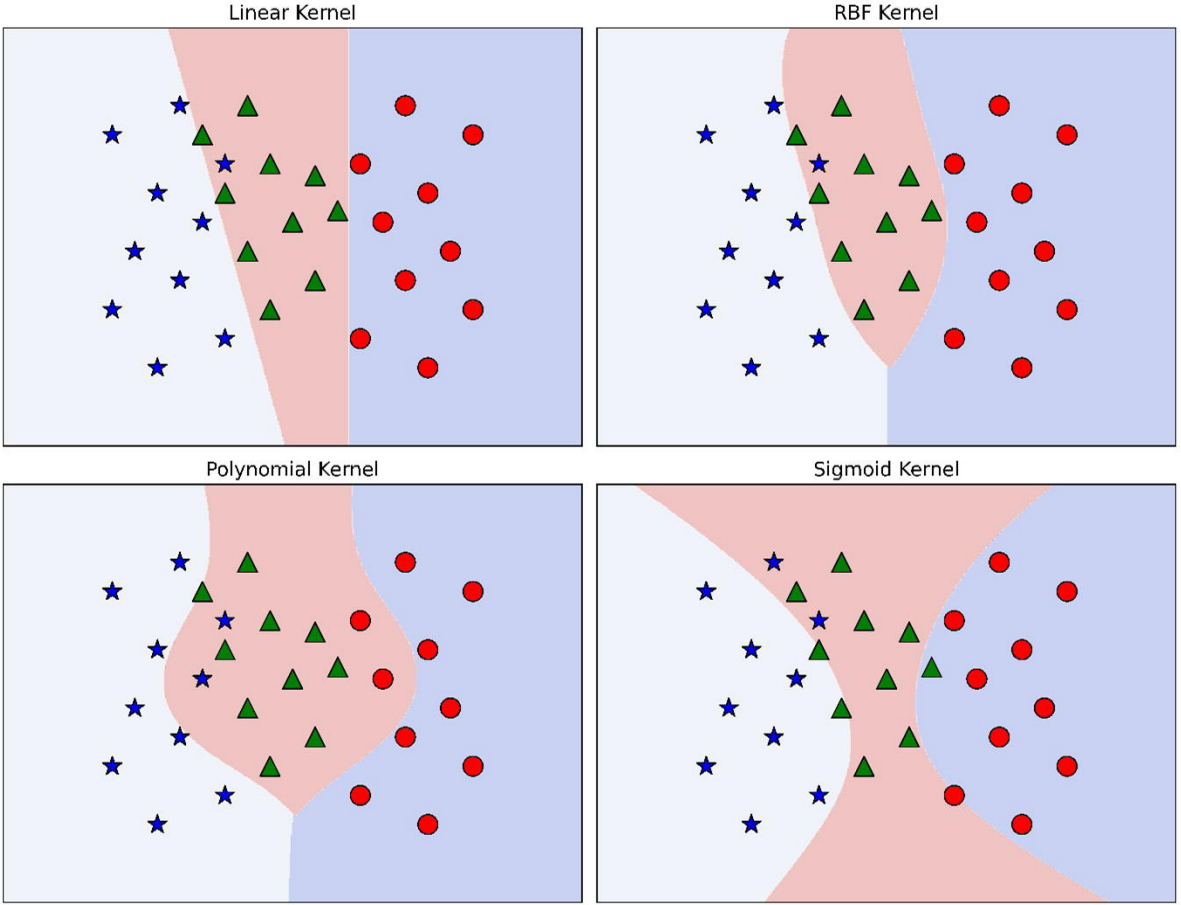
Reference

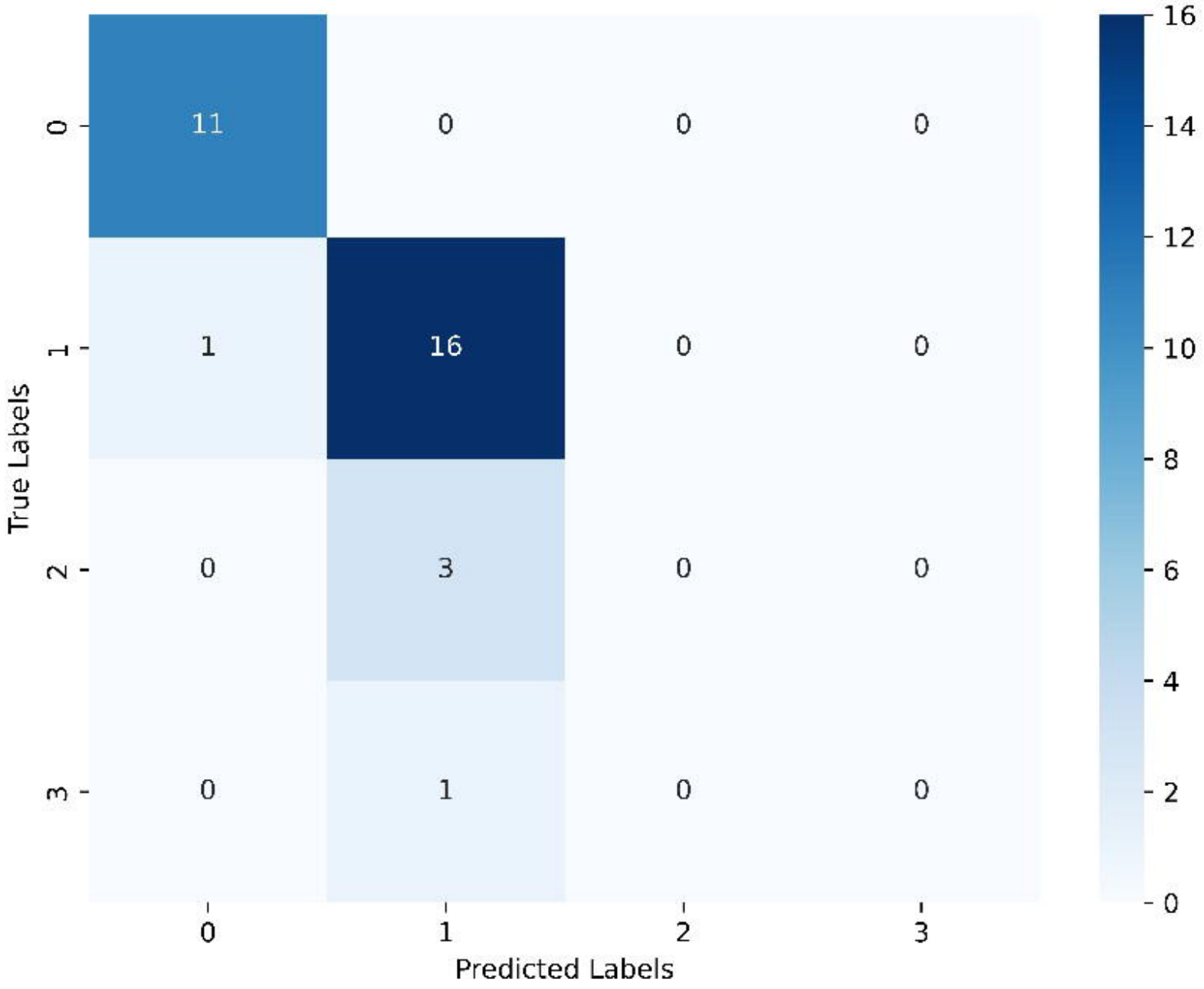
1. Gera A, Pereira PC, Eapen C. Low Back Pain in Adolescent Athletes; Evaluation and Rehabilitation. In: Journal of Exercise Science and Physiotherapy [Internet]. 2015 [cited 2025 Apr 2]. p. 76. https://doi.org/10.18376//2015/v11i2/67706
2. Wall J, McGowan E, Meehan W, Wilson F. “Back pain is part of sport … I’m just gonna have to live with it”: Exploring the lived experience of sport-related low back pain in adolescent athletes. Phys Ther Sport. 2023 Jul 1;62:71–8. https://doi.org/10.1016/j.ptsp.2023.05.005
3. Kikuchi R, Hirano T, Watanabe K, Sano A, Sato T, Ito T, et al. Gender differences in the prevalence of low back pain associated with sports activities in children and adolescents: a six-year annual survey of a birth cohort in Niigata City, Japan. BMC Musculoskelet Disord. 2019 Jul 13;20(1):327. https://doi.org/10.1186/s12891-019-2707-9
4. Vij N, Naron I, Tolson H, Rezayev A, Kaye AD, Viswanath O, et al. Back pain in adolescent athletes: a narrative review. Orthop Rev [Internet]. 2022 Aug 5 [cited 2025 Apr 2];14(3). https://doi.org/10.52965/001c.37097
5. Orchard J, Saw R, Kountouris A, Redrup D, Farhart P, Sims K. Management of lumbar bone stress injury in cricket fast bowlers and other athletes. South Afr J Sports Med [Internet]. 2023 Jun 5 [cited 2025 Apr 2];35(1). https://doi.org/10.17159/2078-516X/2023/v35i1a15172
6. Senington B, Lee RY, Williams JM. Biomechanical risk factors of lower back pain in cricket fast bowlers using inertial measurement units: a prospective and retrospective investigation. BMJ Open Sport — Exerc Med. 2020 Aug 13;6(1):e000818. https://doi.org/10.1136/bmjsem-2020-000818
7. Desouza C, Jani C, Patil V. Non-Isthmic Spondylolysis Imaging Features: A Case Report. J Orthop Case Rep. 2022;12(02):42–4. https://doi.org/10.13107/jocr.2022.v12.i02.2658

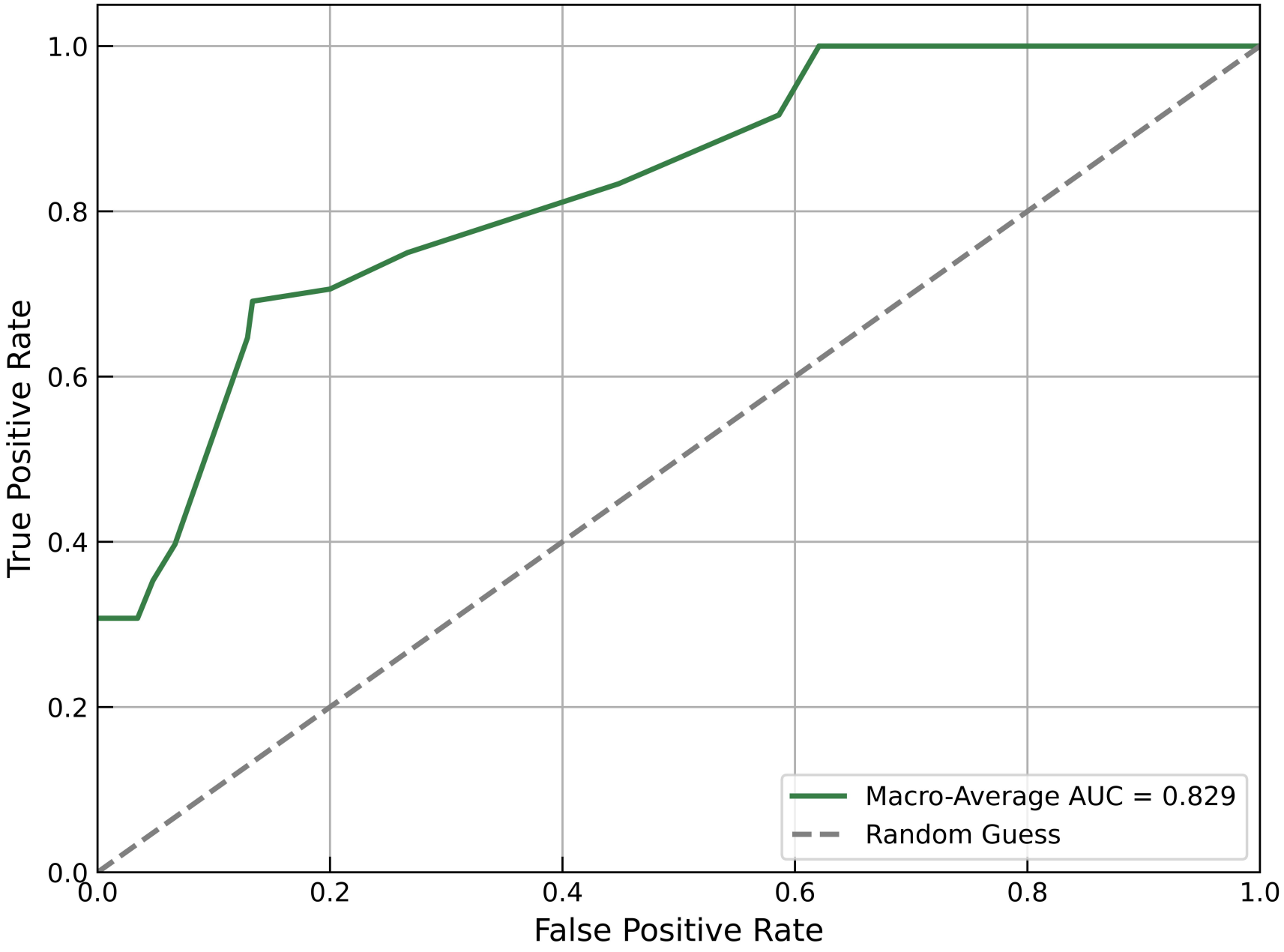
8. Selhorst M, Rodenberg R, Padgett N, Fischer A, Ravindran R, MacDonald J. An Alternative Model of Care for the Treatment of Adolescent Athletes with Extension-Based Low Back Pain: A Pilot Study. Int J Sports Phys Ther [Internet]. 2021 Feb 1 [cited 2025 Apr 2];16(1). https://doi.org/10.26603/001c.18715
9. Seas A, Zachem TJ, Valan B, Goertz C, Nischal S, Chen SF, et al. Machine learning in the diagnosis, management, and care of patients with low back pain: a scoping review of the literature and future directions. Spine J. 2025 Jan;25(1):18–31. https://doi.org/10.1016/j.spinee.2024.09.010
10. Abdollahi M, Ashouri S, Abedi M, Azadeh-Fard N, Parnianpour M, Khalaf K, et al. Using a Motion Sensor to Categorize Nonspecific Low Back Pain Patients: A Machine Learning Approach. Sensors. 2020 Jun 26;20(12):3600. https://doi.org/10.3390/s20123600
11. Bayne H, Elliott B, Campbell A, Alderson J. Lumbar load in adolescent fast bowlers: A prospective injury study. J Sci Med Sport. 2016 Feb 1;19(2):117–22. https://doi.org/10.1016/j.jsams.2015.02.011
12. Abdullah AA, Yaakob A, Ibrahim Z. Prediction of Spinal Abnormalities Using Machine Learning Techniques. In: 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA) [Internet]. 2018 [cited 2025 May 1]. p. 1–6. https://doi.org/10.1109/ICASSDA.2018.8477622
13. Al Imran A, Rifatul Islam Rifat Md, Mohammad R. Enhancing the Classification Performance of Lower Back Pain Symptoms Using Genetic Algorithm-Based Feature Selection. In: Uddin MS, Bansal JC, editors. Proceedings of International Joint Conference on Computational Intelligence. Singapore: Springer Nature; 2020. p. 455–69. https://doi.org/10.1007/978-981-13-7564-4_39

14. Hasan KA, Hasan MdAM. Identification and Analysis of Risk Factors of Lower Back Pain Using Machine Learning Approaches. In: 2020 11th International Conference on Electrical and Computer Engineering (ICECE) [Internet]. 2020 [cited 2025 May 1]. p. 129–32. https://doi.org/10.1109/ICECE51571.2020.9393098
15. Sandag GA, Tedry NE, Lolong S. Classification of Lower Back Pain Using K-Nearest Neighbor Algorithm. In: 2018 6th International Conference on Cyber and IT Service Management (CITSM) [Internet]. 2018 [cited 2025 May 1]. p. 1–5. https://doi.org/10.1109/CITSM.2018.8674361
16. Hossen S, Ali MO, Rashid MA, Higa H. Development of proning pose classification system for selfcare of COVID-19 patients. Prog Eng Sci. 2025;2(2):100064. https://doi.org/10.1016/j.pes.2025.100064
17. Abe S. Support Vector Machines for Pattern Classification [Internet]. London: Springer London; 2010 [cited 2025 May 1]. (Advances in Pattern Recognition). https://doi.org/10.1007/978-1-84996-098-4
18. Mohsen S, Elkaseer A, Scholz SG. Human Activity Recognition Using K-Nearest Neighbor Machine Learning Algorithm. In: Smart Innovation, Systems and Technologies [Internet]. Singapore: Springer Singapore; 2022 [cited 2025 May 1]. p. 304–13. https://doi.org/10.1007/978-981-16-6128-0_29
19. Xu L, Yang W, Cao Y, Li Q. Human activity recognition based on random forests. In: 2017 13th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD) [Internet]. IEEE; 2017 [cited 2025 May 1]. p. 548–53. https://doi.org/10.1109/FSKD.2017.8393329
20. Bergmann GG, Pinheiro E dos S, Mello JB, Graup S. Associação entre diferentes domínios da atividade física e a dor lombar inespecífica em adolescentes. Motricidade. 2020 Sep 30;16(3):245–54. https://doi.org/10.6063/motricidade.18141

21. Arend M, Toomsalu L, Kaasik P. High prevalence of ankle, knee and low back problems in highly trained adolescent basketball players at the beginning of their competitive season. Pap Anthropol. 2024 Nov 22;33(1):72–84. https://doi.org/10.12697/poa.2024.33.1.04
22. Yoshimizu R, Nakase J, Yoshioka K, Shimozaki K, Asai K, Kimura M, et al. Incidence and temporal changes in lumbar degeneration and low back pain in child and adolescent weightlifters: A prospective 5-year cohort study. PLoS ONE. 2022 Jun 29;17(6):e0270046. https://doi.org/10.1371/journal.pone.0270046
23. Ding L, Luo J, Smith DM, Mackey M, Fu H, Davis M, et al. Effectiveness of Warm-Up Intervention Programs to Prevent Sports Injuries among Children and Adolescents: A Systematic Review and Meta-Analysis. Int J Environ Res Public Health. 2022 May 23;19(10):6336. https://doi.org/10.3390/ijerph19106336
24. Chin ECY, Chung-Nam Lai S, Tsang SF, Ho-Ngai Chung S, Wong YL, Sran N, et al. Comparing the Effects of Aqua- and Land-Based Active Cooldown Exercises on Muscle Soreness and Sport Performance: A Randomized Crossover Study. Int J Sports Physiol Perform. 2024 Dec 1;19(12):1381–90. https://doi.org/10.1123/ijspp.2024-0020
25. da Silva T, Mills K, Brown BT, Herbert RD, Maher CG, Hancock MJ. Risk of Recurrence of Low Back Pain: A Systematic Review. J Orthop Sports Phys Ther. 2017 May;47(5):305–13. https://doi.org/10.2519/jospt.2017.7415
26. Farhart P, Beakley D, Diwan A, Duffield R, Rodriguez EP, Chamoli U, et al. Intrinsic variables associated with low back pain and lumbar spine injury in fast bowlers in cricket: a systematic review. BMC Sports Sci Med Rehabil. 2023 Sep 20;15(1):114. https://doi.org/10.1186/s13102-023-00732-1


Linear Kernel
RBF Kernel
Polynomial Kernel
Sigmoid Kernel


True Labels
Predicted Labels
0
1
2
3
11
0
0
0
1
16
0
0
0
3
0
0
0
1
0
0
16
14
12
10
8
6
4
2
0


1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
Macro-Average AUC = 0.829
Random Guess