International Journal of Physical Education, Sports and Health 2025; 12(3): 564-570

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (RJIF): 5.38
IJPESH 2025; 12(3): 564-570

https://www.kheljournal.com
Received: 24-04-2025
Accepted: 27-05-2025

Anbil Mahesh Poyyamozhi
Research Scholar, National College, Tiruchirappali, Tamil Nadu, India

Dr. D Prasanna Balaji
Associate Professor and Head, PG and Research, Department of Physical Education and Sports Sciences, National College, Tiruchirappali, Tamil Nadu, India

Dr. A Mahaboobjan
Professor and Secretary (Sports), Department of Physical Education and Yoga, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India

Corresponding Author:
Anbil Mahesh Poyyamozhi
Research Scholar, National College, Tiruchirappali, Tamil Nadu, India

Analyzing the impact of sports participation on academic growth using machine learning techniques

Anbil Mahesh Poyyamozhi, D Prasanna Balaji and A Mahaboobjan

DOI: https://www.doi.org/10.22271/kheljournal.2025.v12.i3h.3855

Abstract
In recent years, physical activity has become one of the most prominent areas of study, emphasizing how it contributes to well-rounded student development. By analyzing and predicting performance outcomes, this study analyzes and examines the impact of sports participation on academic growth. In order to develop and evaluate machine learning models, we used a dataset consisting of academic records, frequency and type of sports involvement, as well as other demographic attributes related to the students. Academic improvement is positively correlated with regular engagement in sports activities, and key predictors include physical fitness indicators, duration of participation, and type of sport. The findings indicate the potential of data-driven approaches for uncovering hidden patterns and supporting policy development regarding the integration of physical education into academics. As sports play an increasingly important role in academic environments, this study highlights the role of machine learning in educational analytics.

Keywords: Physical activity, academic performance, curriculum design, machine leaning models

Introduction
Educating individuals and societal groups serves as a cornerstone of life, but its quality and impact vary significantly among different types of schools, including public, private, international, and government-sponsored institutions. Different school types have distinct pedagogical approaches, infrastructures, curriculum designs, and socioeconomic influences that influence students' academic journeys. In rural and government schools, while advanced resources and extracurricular activities are often available, insufficient infrastructure and shortages of teachers make learning difficult.

Although school types differ, there is a common challenge faced by students across them: maintaining academic progress while coping with growing cognitive and emotional pressures. Several factors contribute to a lack of academic achievement, including exam anxiety, a lack of personalized learning, socio-economic limitations, and sedentary lifestyles. In terms of academic and personal development, participation in sports activities contributes to the development of a wide range of essential life skills. As students participate in sports consistently, they enhance their communication skills by engaging in interactions with teammates and coaches and develop leadership skills by leading peers and taking initiative. By collaborating toward a common goal, students learn to tolerate and respect differences in viewpoints and backgrounds through collaborative learning. Furthermore, sports activities provide students with a platform for developing interpersonal skills, which facilitates the development of positive relationships, as well as an opportunity to develop self-reflection, discipline, and emotional regulation. Learning to handle high-pressure situations and solve problems under pressure develops learners' problem-solving skills, enabling them to engage in effective crisis management and develop their resilience. A sense of accomplishment and mastery gained through athletics enhances students' confidence, helping them develop a lifelong learning mindset and foster academic success.

Sport and physical activity are crucial components of fostering academic growth and holistic student development in this context. It has been proven through many studies that regular physical activity improves cognitive function, memory retention, classroom behavior, and emotional well-being. Besides enhancing physical health, sports also promote discipline, teamwork, time management, and resilience, all of which have positive effects on academic performance. Despite this, many schools, especially those that emphasize examination-oriented curricula, undervalue or fail to implement sports often.

As the importance of evidence-based education and student wellness grows, this study is analyzing and predicting academic progress based on students' participation in sports. It is intended to uncover hidden patterns and quantify the academic benefits of sports engagement by utilizing data-driven models.

Background of the study

In today's educational environment, academic excellence is not the only indicator of success. A growing number of institutions are realizing that participation in co-curricular and extracurricular activities, particularly sports, contributes to students' cognitive, social, and emotional development (Bailey, 2006) [2]. By engaging in sports activities, students can develop important life skills like discipline, resilience, time management, and teamwork, which directly affects classroom performance and motivation (Shephard, 1997) [11].

Physical activity and academic performance are positively correlated in numerous studies. As a result of regular physical activity, memory, attention, and processing speed are significantly enhanced, which are all attributes that significantly enhance learning capabilities (Donnelly *et al*., 2016; Martinez-Bueno *et al*., 2017) [4, 1]. Furthermore, sports participation tends to increase school attendance and classroom behavior, factors that are related to academic success in the long run (Rasberry *et al*., 2011) [10]. It has been demonstrated that physical education should be integrated into school curricula as a fundamental component.

Even though this is known, schools still prioritize academics over physical education, particularly in developing countries. The exam-centric, high-pressure environment found in many institutions prevents many institutions from utilizing sports programs as development tools (Hardman & Green, 2011) [5]. Therefore, many students lose out on physical activity's cognitive and psychosocial benefits, especially in schools with limited resources (Singh *et al*., 2012) [12].

Data-driven educational models are enabling new ways to evaluate and optimize student development, while artificial intelligence advances are offering new opportunities. Bakker & Inventado (2014) [3] describe how machine learning can extract predictive insights from complex and high-dimensional educational datasets. The application of machine learning to sports and academic data allows the discovery of latent patterns, identification of key growth parameters, and prediction of academic trajectories based on activity involvement (Lu *et al*., 2018; Peña & Larrañaga, 2020) [7, 9].

It is important to note that research has largely focused on academic indicators such as grades and test scores without considering qualitative factors such as communication, leadership, emotional intelligence, and self-confidence. Nowadays, learning environments are dynamic and collaborative, so these soft skills are essential (Heckman & Kautz, 2012) [6]. In addition to providing a natural context for fostering these attributes, sports can also facilitate the measurement of those attributes and provide actionable educational strategies for students (McKenzie & Lounsbery, 2013) [8].

The present study incorporates machine learning techniques with a multidimensional evaluation of students' academic and personal growth through sports participation in order to fill this gap. In addition to academic outcomes, nine essential developmental parameters are examined in this study communication, leadership, collaborative learning, tolerance, interpersonal skills, intra-personal skills, problem-solving, crisis management, sports skill, and confidence drawn from a variety of school contexts. In this way, the research aims to inform inclusive, balanced, and effective educational policies by providing empirical evidence.

Methodology

- In order to assess the impact of sports on academic growth, nine key developmental parameters were used: Communication, Leadership, Collaborative Learning, Tolerance, Interpersonal Skills, Intra-personal Skills, Problem Solving and Crisis Management, Sporting Skill, and Confidence. Students' development was assessed both cognitively and non-cognitorily through these parameters.
- A variety of sporting events and physical activities took place over a fixed academic term for students of different school types. An evaluation of the development impact was conducted using a pre- and post-test survey design. In order to ensure the comparative analysis of student progress, data were collected before and after the intervention (sports program).
- We carefully designed the survey instrument to collect both quantitative and qualitative information, incorporating multiple-choice questions, open-ended responses, and Likert ratings. A mixed-methods approach enabled the study to assess nine targeted parameters of student perception, behavioral change, and skill development.
- In order to identify patterns and correlations between responses, statistical techniques and machine learning algorithms were used to code, clean, and analyze the responses. In addition to ensuring objective assessment, the methodology also captured nuanced implications of sports participation for academic and personal development.

Results and Discussion

Comparison of learning skills with different school level

In order to assess students' learning abilities and IQ levels, a questionnaire was initially distributed. As part of the pretest, students were assessed to determine their cognitive abilities, interpersonal skills, and learning capacity. A detailed description of the questionnaires used for the assessment can be found in the annexure section. The data collected revealed a wide range of skill levels across school types offering valuable insight into specific areas that need to be addressed.

A structured program incorporating sports-based activities was then introduced, which focused on nine key aspects of learning skills. In addition to enhancing academic learning, this innovative approach positively affected students' mental health. Schoolchildren showed significant improvements in leadership, communication, collaboration, and confidence following the intervention, which confirmed that it facilitated holistic development.

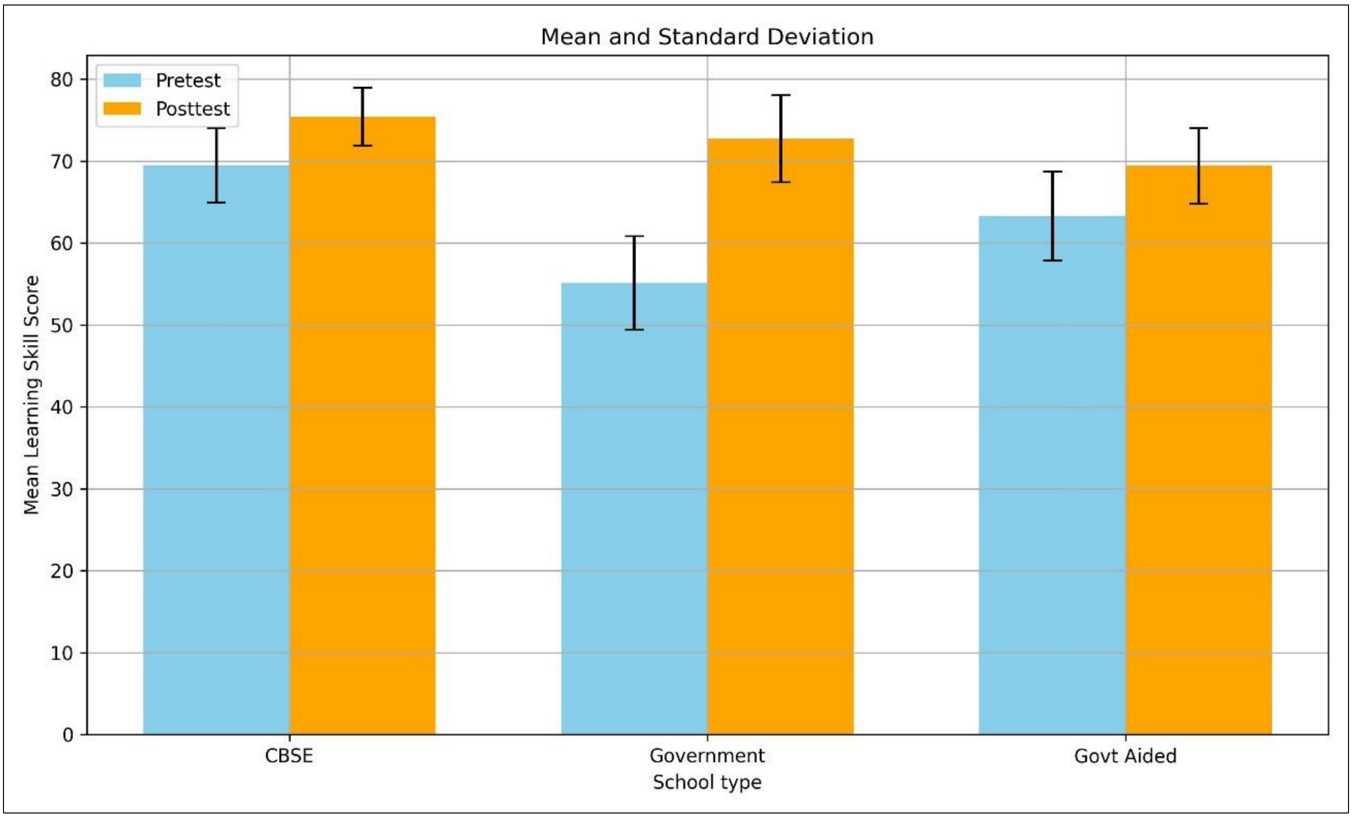

Fig 1: Mean and standard deviation of learning skill score by school type

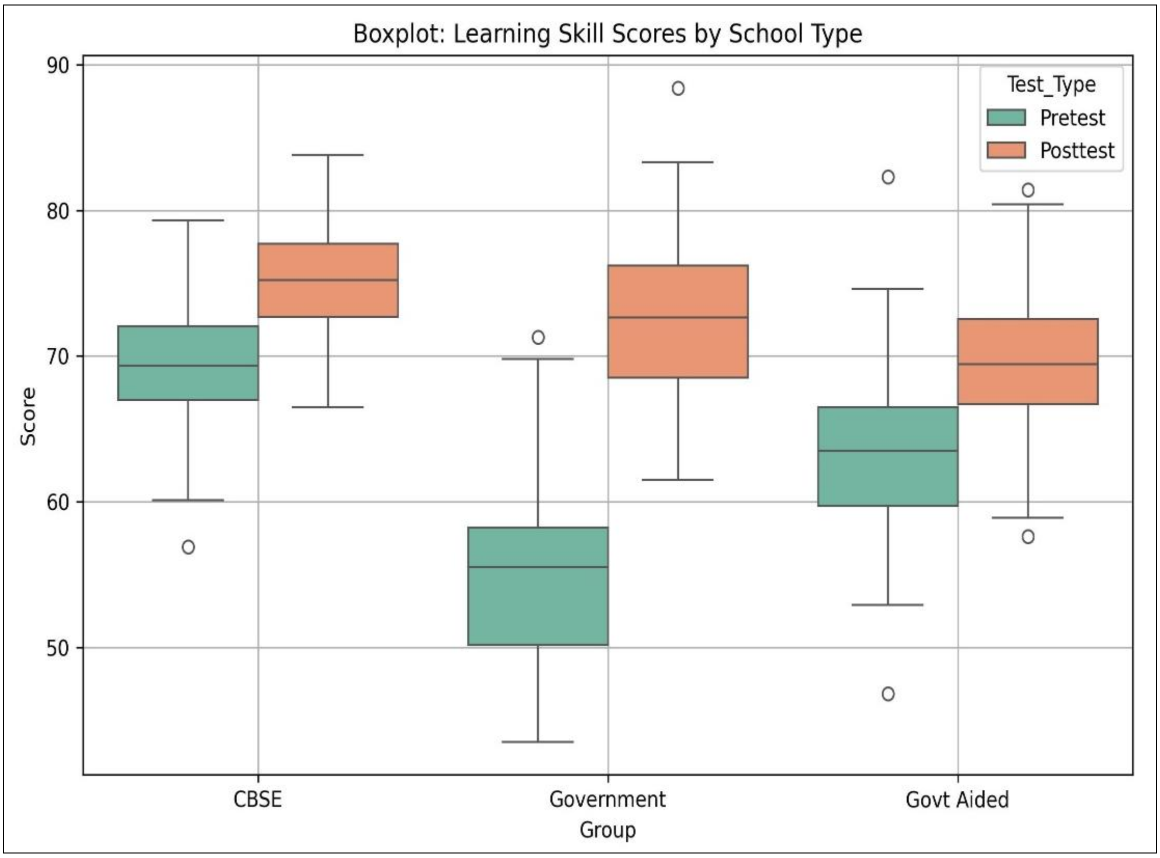

Fig 2: Median learning skill score by school type

Figure 1 and Figure 2 show how the sports-integrated learning skills intervention is affecting students across different types of schools. The results confirm this trend by showing clear increases in post-test scores for all groups compared with pretest scores. There was a significant gain for Government schools at the end of the study, moving from the lowest pretest median to the highest post-test median, whereas Government Aided schools showed moderate improvement with more variation in scores. By implementing learning skills through sports activities, students improved their leadership, communication, and confidence performance, thus demonstrating the effectiveness of the intervention in improving both learning outcomes and mental well-being among students.

ANOVA test

ANOVA (Analysis of Variance) is used to determine whether there are statistically significant differences between the means of three or more groups.

In your case, the experiment comparing the Learning Score across three school types:

- CBSE
- Government
- Govt Aided

ANOVA is calculated using following equations

Between group sum of square (SSB) = $\sum n_i(\bar{x}_g - \bar{x}_o)$ (1)

In Equation 1, where $\bar{x}_g$ is the group mean, $\bar{x}_o$ is the overall mean and n_i number of observations in each group.

With group sum of Square (SSW) = $\sum\sum(x_{ij} - \bar{x}_g)$ (2)

In Equation 2 the x_{ij} is number of data point in each group g.

Degree of freedom - Between groups $df_g = k - 1$

Degree of freedom - Within groups $df_w = n - k$

$$MSB = \frac{SSB}{df_g}$$

$$MSW = \frac{SSW}{df_w}$$

$$F = \frac{MSB}{MSW} \quad (3)$$

The results obtained from ANOVA is

F-statistic: 118.9845

P-value: 0.0000

According to the analysis of variance (ANOVA) results, the F-statistic values of the two groups are 118.9845, indicating that the variance between the two groups is significantly greater than that within them. There is a strong indication that the group means differ significantly because of the high F-value. Moreover, the p-value of 0.000 indicates that there is very little chance that such an extreme F-value can occur by chance. In this case, we reject the null hypothesis as the p-value was below the threshold of 0.05, thereby validating the statistical significance of the differences between the groups.

Null hypothesis (H_0): It is one of the most fundamental concepts in statistics that the null hypothesis (often written as H0) represents. Assumptions about a population or conditions associated with an experiment are based on a default or baseline assumption. According to the null hypothesis, there are no significant differences or relationships between groups.

In the experiments, three types of schools (CBSE, Government, Government Aided) were compared for Learning Scores.

Null Hypothesis (H_0)

- There is no difference in the average Learning Scores across the three school types.

Alternative hypothesis (H_1)

- At least one school type has a significantly different average Learning Score.

If the p-value > 0.05, we fail to reject the null hypothesis.

→ This means we don't have strong evidence that group means differ.

If the p-value ≤ 0.05, we reject the null hypothesis.

→ This suggests that at least one group's mean is significantly different.

The test result concluded that the average learning skills are not the same across the school types.

Machine leaning models

Statistically speaking, regression analyses can be used to examine the relationship between one or more dependent variables and one or more independent variables. In the context of education, regression is useful for quantifying the contribution of different factors to differences in student performance, such as sports participation, school type, or parent education. In order to predict learning scores, a regression model is used that identifies which variables have the most significant effects. It is especially useful when it comes to making data-driven decisions about education policy as well as the design of educational programs.

Students' academic performance is improved by participating in sports to what extent can regression analysis be used to investigate this relationship? In this regression model, learning scores are used as the dependent variable, and sports activity is treated as an independent variable (alongside other variables like teacher qualification and economic status). Sports participation correlates positively with better learning outcomes when the coefficient is positive and statistically significant. Students' achievement can be enhanced through targeted interventions that incorporate physical activities.

A clear definition of the research problem was made at the beginning of the study. It was to understand the impact of sports activity, teacher education, and school type on student learning outcomes. To measure the impact of sports activity on multiple dimensions, nine variables were used. School data was collected from a variety of schools, including public and private institutions, capturing information such as sports participation, teacher qualification, and academic performance scores of students. In order to understand why learning is so different, this data set provided a base for exploring relationships and patterns that could be related to learning. During the course of the study, samples of data were collected, as shown in Table1.

Data pre-processing

The data was pre-processed after collection in order to handle missing values, encode categorical variables, and normalize numerical features. In order to identify trends and anomalies, Exploratory Data Analysis (EDA) was carried out using visualizations and summary statistics. Sports activities are associated with greater improvements in learning outcomes in government schools, for example. Further detailed analyses were undertaken to quantify these effects rigorously as a result of this initial insight.

Table 1: Sample data set

Parent education	Parent occupation	Economic status	Location	School type	Teacher qualification	School facility	Sports activity	Learning score
postgraduate	engineer	middle	semi-urban	government	M.Ed	good	yes	91.48
postgraduate	engineer	low	rural	government	M.Ed	poor	no	72.31
graduate	engineer	high	semi-urban	private	PhD	good	no	78.24
postgraduate	business	middle	rural	private	M.Ed	average	no	78.62
graduate	teacher	high	semi-urban	private	M.Ed	good	yes	75.83
secondary	engineer	low	rural	private	B.Ed	poor	no	65.83
secondary	engineer	low	urban	private	B.Ed	poor	no	74.9
postgraduate	labor	middle	semi-urban	private	M.Ed	average	yes	76.84
none	business	middle	semi-urban	government	M.Sc	average	yes	78.65
postgraduate	engineer	low	rural	private	M.Ed	average	no	73.71
none	business	low	semi-urban	private	B.Ed	average	no	60.68
primary	labor	high	semi-urban	government	M.Ed	good	yes	80.67
none	business	middle	urban	private	B.Ed	good	no	70.21
graduate	labor	low	semi-urban	government	B.Ed	average	yes	74.43

Information gain

Information Gain is a crucial concept in machine learning, particularly in decision tree algorithms such as ID3 and C4.5, where it helps determine which attribute should be used to split the data at each decision node. It measures the reduction in entropy or uncertainty after the dataset is divided based on a specific attribute.

$$IG(D, A) = Entropy(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} \cdot Entropy(D_v)$$

Where:

- D is the dataset,
- A is an attribute,
- D_v is the subset of data where attribute A has value v,
- $Entropy(D)$ is a measure of the disorder or impurity in the dataset.

According to Table 2, information gain is reflected in the following values. A feature's separation ability is determined by how well it distinguishes data from each other. In other words, a higher Information Gain indicates that the feature provides a better split that contributes to identifying different outcome types more effectively. This technique is particularly useful in classification problems since it leads to purer subsets with more homogeneous class labels. According to Information Gain, we can determine which of the nine sports activity variables help classify students into improved and not improved learning categories based on their performance in sports. In the example above, if the variable "frequency of sports participation" yields the highest Information Gain, it would result in more clearly defined learning outcome classes when data are split based on this variable. Thus, by using attributes with high Information Gain, the classification model will perform better because it will be able to discriminate classes better.

Table 2: Information ratio values

Features	Information gain ratio
Sports activity	0.1285
Teacher qualification	0.1066
School_type	0.090
Parent_education	0.0433
location	0.0090
Economic_status	0.0082

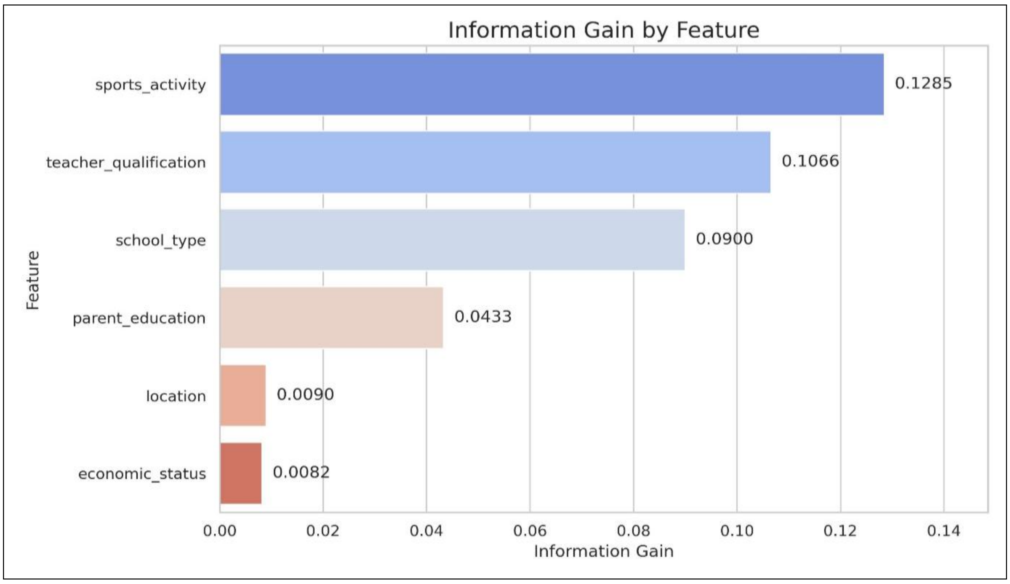

Fig 3: Information gain ratio

Transition of regression to classification model

Since the goal of this study was to categorize students' learning outcomes (such as improved versus not improved), continuous academic performance scores were transformed into categorical labels. In this case, the problem has changed into what resembles a classification task, as opposed to a regression problem. As an example, threshold scores were set so that scores above them would be labelled as 'Improved' and below them as 'Not Improved.' This allowed classification algorithms to be used to understand the factors that influenced categorical learning outcomes.

Model selection and training

A number of classification models, such as Decision Trees, Random Forests, and Support Vector Machines (SVMs), were trained to predict the learning outcome class using input features. Metrics such as accuracy, precision, recall, and F1-score were used for evaluating the models. True positives, false positives, false negatives, and true negatives were depicted using confusion matrices. The initial results showed an imbalance of instances between classes, showing that one class had more instances than the other, which could result in inaccurate predictions.

Addressing class imbalance

A number of techniques were used to correct the imbalance between classes, including Synthetic Minority Over-sampling Technique (SMOTE) and class weighting. SMOTE balances the dataset by creating synthetic samples of minorities, which increases the classifier's ability to detect minorities. In general, such modelling techniques resulted in better classification performance across all classes as well as a reduction of bias toward the majority class after application.

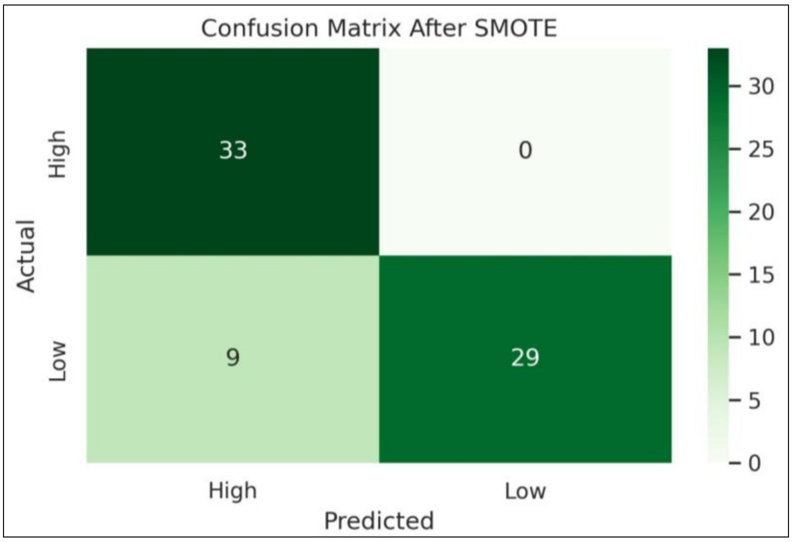

Fig 4: Confusion matrix

This confusion matrix illustrates how the classification model performs better when class imbalance is handled by applying the Synthetic Minority Oversampling Technique (SMOTE). It was noted that 33 cases of the "High" class were correctly predicted, and no false negatives were observed, indicating that the model correctly identified those cases. A total of 29 instances were classified correctly by the "Low" class, while nine were misclassified by the "High" class, indicating some residual overlap between the two classes. As a result of SMOTE, the model is much more capable of distinguishing between "High" and "Low" learning outcomes, in particular ensuring that the minority class does not remain underrepresented during training. It is evident from the confusion matrix that the model is more reliable and less biased because it predicts both classes accurately. This is especially true in educational datasets where identifying learning improvement levels (based on variables such as sports activity) is crucial.

As a final step, figures 3 and 4 illustrate the results in terms of their educational impact. A recent study found that participation in sports activities positively affects learning outcomes, particularly in government schools, suggesting they may be effective educational interventions. In addition to factors such as teacher education, school type, and parental education, students' learning skills are further enhanced when they actively participate in sports activities. It is based on these insights that policymakers and educators can prioritize sports and teacher training programs as means of boosting academic achievement.

Conclusions

A study conducted by these author shows that sports participation contributes significantly to students' academic development and overall personal development. In this study, machine learning techniques are incorporated into an assessment of cognitive and non-cognitive parameters, such as communication, leadership, problem-solving, and self-confidence, which demonstrates that physical activity is not merely a secondary aspect of education, but an integral part of the learning process.

A pre-test and post-test evaluation using a variety of data collection methods revealed that regular sports involvement significantly improved student academic performance and personal growth. Models like Decision Trees, Random Forests, and SVM are highly effective at capturing and predicting these outcomes, demonstrating the potential of data-driven interventions in education. Additionally, this study emphasizes the importance of structured physical education programs in all types of educational settings, providing valuable insights for educators, school administrators, and policymakers. The study supports the use of a well-balanced educational approach that incorporates academics and sports to foster a well-rounded student.

References

1. Álvarez-Bueno C, Pesce C, Cavero-Redondo I, Sánchez-López M, Garrido-Miguel M, Martínez-Vizcaíno V. Academic achievement and physical activity: A meta-analysis. Pediatrics. 2017;140(6):e20171498.
2. Bailey R. Physical education and sport in schools: A review of benefits and outcomes. J Sch Health. 2006;76(8):397-401.
3. Baker RS, Inventado PS. Educational data mining and learning analytics. In: Learning Analytics. New York: Springer, c2014. p. 61-75.
4. Donnelly JE, Hillman CH, Castelli D, Etnier JL, Lee S, Tomporowski P, *et al*. Physical activity, fitness, cognitive function, and academic achievement in children: A systematic review. Med Sci Sports Exerc. 2016;48(6):1197-222.
5. Hardman K, Green K. Contemporary issues in physical education: International perspectives. Aachen: Meyer & Meyer Sport, 2011.
6. Heckman JJ, Kautz T. Hard evidence on soft skills. Labour Econ. 2012;19(4):451-464.
7. Lu OHT, Huang AYQ, Huang JCH, Lin AJQ, Yang SJH. Applying learning analytics for the early prediction of students' academic performance in blended learning. Educ Technol Soc. 2018;21(2):220-232.
8. McKenzie TL, Lounsbery MAF. Physical education teacher effectiveness in a public health context. Res Q Exerc Sport. 2013;84(4):419-430.
9. Peña C, Larrañaga M. Using machine learning to predict academic performance and school dropout: A systematic literature review. Rev Iberoam Tecnol Aprendiz. 2020;15(3):109-117.
10. Rasberry CN, Lee SM, Robin L, Laris BA, Russell LA, Coyle KK, *et al*. The association between school-based physical activity, including physical education, and academic performance: A systematic review of the literature. Prev Med. 2011;52(Suppl):S10-S20.
11. Shephard RJ. Curricular physical activity and academic performance. Pediatr Exerc Sci. 1997;9(2):113-126.
12. Singh A, Uijtdewilligen L, Twisk JW, van Mechelen W, Chinapaw MJ. Physical activity and performance at school: A systematic review of the literature including a methodological quality assessment. Arch Pediatr Adolesc Med. 2012;166(1):49-55.