SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(4), 625.
https://doi.org/10.62617/mcb625

Article

Research on real-time collection and analysis of student health and physical fitness data using biosensors

Yuliang Zhou

PE Department, Wuxi Institute of Technology, Wuxi 214121, China; 17766343741@163.com

CITATION

Zhou Y. Research on real-time collection and analysis of student health and physical fitness data using biosensors. Molecular & Cellular Biomechanics. 2024; 21(4): 625. https://doi.org/10.62617/mcb625

ARTICLE INFO

Received: 25 October 2024
Accepted: 20 November 2024
Available online: 11 December 2024

Abstract: Biosensors have emerged as efficient devices for monitoring personal fitness levels and health profiles as an important part of this technological development. With growing concern about students' health and bodily fitness, educational and health experts as well as lawmakers have increasingly emphasized their importance. The goal of the study is to explore a real-time system for collecting and analyzing data on students' physical fitness and health utilizing biosensors and advanced algorithms. The study proposed a novel Efficient Osprey Optimized Adjustable Random Forest (EOO-ARF) to predict the student health and physical fitness level. The student health and physical fitness data was gathered from a Kaggle source. To gather information using wearable biosensors to constantly monitor crucial health parameters such as blood oxygen levels, body temperature, heart rate, and physical activity. The data was pre-processed using the *Z*-score normalization to enhance the quality of the data. The Principal Component Analysis (PCA) was used to extract the features from pre-processed data. This model takes the indices of students' physical health as the input parameters and produces an overall health score. EOO is used for optimization, and the process aims at selecting the most appropriate features to identify the health metrics most relevant to influencing students' general fitness levels. ARF is applied to predict the health and fitness levels of students. The performance of the suggested approach is evaluated in terms of F1-score (98.13%), recall (98.2%), and accuracy (98.44%). The integration of biosensors with innovative analytic methods could transform the monitoring and improvement of the physical fitness and health of students take place in real-time.

Keywords: student health; physical fitness; biosensors; principal component analysis (PCA); efficient osprey optimized adjustable random forest (EOO-ARF)

1. Introduction

Physical fitness and health status among students are essential indicators that show the level of health and performance of students. The necessity of increasing the physical component when considering a person's learning process has drawn attention recently, which led to the interest in methods of monitoring student health [1]. Since fitness-related problems can be addressed in educational settings, there is a growing interest in studying student's levels of fitness and health prediction. The health histories of students enable teachers, policymakers, and even health practitioners to offer more targeted support, ensuring that learners achieve academic excellence while exercising a healthy lifestyle [2].

Fitness in the physical aspects embraces range of motion, muscle strength, stamina, and the proportion of the body, which are enhancements to the general health of an individual. Importantly, health is not only the absence of disease but also encompasses physical, mental, and social well-being [3]. Students' health and fitness depend on various aspects, including diet, habits of living, regular exercise as well as, and the socio-economic status of the learner. Routine physical checkups and traditional health assessments are useful but cannot capture real-time information or long-term patterns [4]. Advances in big data analytics and ML have completely revolutionized the health and fitness predictions made in this digital era. The prediction algorithms are also capable of considering behaviors such as sleep, food, and physical activity, and associating them with fitness-related outcomes [5].

The important aspect of predicting the health and physical condition levels of students is that it extends to their well-being [6]. Academic achievement is directly correlated with physical fitness because physically active students perform better in thinking, memory, and concentration. The prediction models can be used for the early detection of health risks that can be useful in controlling life conditions such as obesity diseases, heart diseases, and mental health diseases, among others that are on the rise in younger generations [7].

To forecast the health and degree of physical fitness of students, the research suggested a unique Efficient Osprey Optimized Adjustable Random Forest (EOO-ARF) approach.

The study is separated into the following sections: related works, methodology, results, and conclusion.

2. Related works

The BDNN-CSMHPM was proposed in the investigation [8], to assess the psychological strain of college students during study tours. The BDNN-CSMHPM design fared higher than other approaches as per mental health prediction ratio 98.9%, accuracy 96.4%, emotion recognition 95.3%, Pearson correlation coefficient 97.2%, and psychological monitoring 94.3%. An outline for using IoT for the tracking of student health was discussed in the research [9]. They applied smart healthcare technology that enabled them to monitor students' health conditions continually and capture changes in physiological and behaviors. The results showed that the proposed model made it possible to achieve the specified requirements for the speed and competitiveness of the model in terms of the students' health identification. A new IoT architecture for monitoring Sports health was introduced and established in the investigation [10] which used CNN and Big data analytics models for prediction. According to the results obtained, it was evident that the suggested strategy outperformed the other assessment measures of accuracy, specificity, sensitivity, MCC, and F1 score. Using the wearable smart bracelet technology, an investigation [11] suggested 1D-CNN with an LSTM-based assessment model to evaluate teenagers' physical fitness levels. The outcomes suggested the viability of running PPG records to forecast young people's levels of physical fitness.

The use of DT and the correlation analysis method to analyze the physical condition of college students was examined in the research [12]. The findings demonstrated that an excellent categorization accuracy rate could be attained by training and the accuracy rate could be achieved at 85.033% by improving the depth. To increase college students' knowledge of their physical health and help them to establish objectives for their daily activity, an investigation [13] examined how students' vital abilities, weight, height, gender, and other characteristics affected their fitness levels. The outcomes demonstrated that the BP neural network forecasting algorithm was capable of accurately predicting students' performance. Using modern medical technology, an IoT-based approach to student health management was proposed in research [14] to continuously track students' health indicators and detect biological and behavioral changes. The performance of the SVM increased as a result of having the recommended model evaluated, which was a very good result for the objectives.

The effectiveness of various ML approaches in forecasting university students' social, physical, and psychological conditions was evaluated in the research [15]. RF algorithm outperformed traditional ML methods with significantly enhanced recall, F1 score, accuracy, and precision. The POAANN was proposed in research [16] for forecasting student physical health assessments. The suggested POA-ANN obtained the values of f1-score (0.965), sensitivity (0.964), accuracy (0.973), and precision (0.961). To build a hyper parameter adaptive optimization-based framework for predicting physical fitness scores, research [17] employed the GWO to enhance the GRU neural network's variables. The outcomes demonstrated that the GWO-GRU model-based forecasting methodology was more reliable and accurate as compared to traditional methodologies and formed a reliable tool for the teaching of physical education and monitoring college students' health.

3. Methodology

The student health and physical health data was collected from Kaggle for this purpose. The data was normalized using the *Z*-score normalization technique. PCA is used to obtain features from the analyzed data that was pre-processed. For the purpose of the given study, the new novel Efficient Osprey Optimized Adjustable Random Forest (EOO-ARF) approach was introduced. An overview of the methodology is depicted in **Figure 1** below.

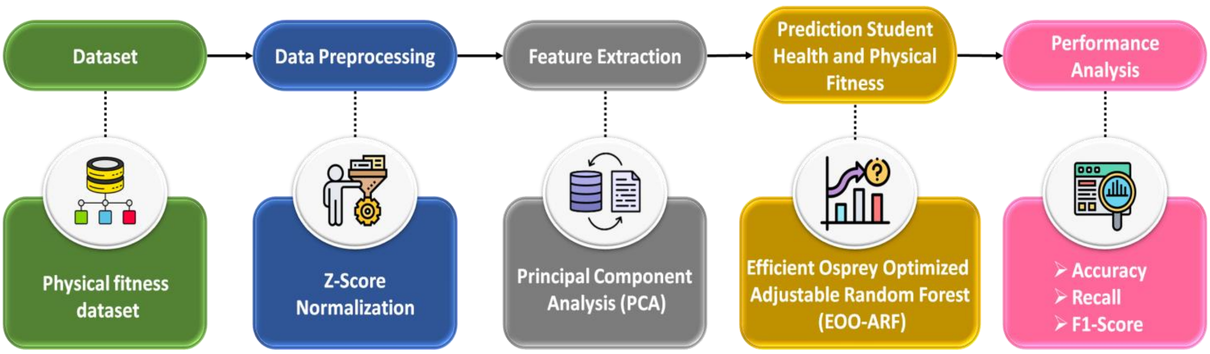

Figure 1. Overview of methodology.

3.1. Data collection

The student health and physical fitness data was gathered for the Kaggle source [18]. The data collection process for the Student Health and Physical Fitness Dataset involves the use of wearable biosensors to track the real-time health metrics of students. These sensors are worn by the students throughout the day, capturing various physiological indicators such as heart rate, blood oxygen levels, body temperature, and physical activity. Data is collected continuously at regular intervals, and each entry is timestamped for temporal analysis. The data is supplemented by self-reported values such as stress levels, providing a more holistic view of the student's health. The dataset is designed to capture a comprehensive range of health metrics, which are then used for predictive analysis and machine-learning models aimed at improving student fitness and wellness. This dataset uses modern ML approaches to make it easier to analyse and forecast students' levels of fitness and health. **Table 1** gives the details about what are all the data collected from the students.

Table 1. Key features of the student health and physical fitness dataset.

Feature	Description
Student ID	Unique identifier for each student to maintain individual records.
Age	Age of the student (15 to 18 years).
Gender	Gender classification (Male, Female, Other).
Blood Oxygen Level	Blood oxygen saturation percentage, a key indicator of respiratory health
Body Temperature	Temperature reading in Celsius, monitoring the student's health status.
Heart Rate	Heart rate in beats per minute (bpm), reflecting cardiovascular health
Physical Activity Level	Intensity of physical activity (METs, metabolic equivalent tasks).
Overall Health Score	Computed score based on multiple health metrics, reflecting overall health
Date and Time	Timestamp for each data entry, enabling temporal analysis
Sleep Duration	Hours of sleep, affecting recovery and health.
Hydration Level	Daily water intake in liters, important for physical performance
Stress Level	Self-reported stress level on a scale of 1 to 10, indicating mental well-being.

3.2. *Z*-score normalization

A normalizing technique based on the data's SD and mean is called Z-score normalization. This strategy is highly beneficial when the actual maximum and minimum values of the information are not known. The following Equation (1) is employed.

$$W_{new} = \frac{W-\mu}{\sigma} = \frac{W - Mean(W)}{StdDev(W)} \tag{1}$$

where, μ—Population mean, W_{new}—New value,σ—SD value,W—Old value.

3.3. Principal component analysis (PCA)

PCA is a common feature extraction technique based on statistical methods. The basic idea is to use linear transformation to transfer the sample data from the HDS into the LDS while maintaining the most accurate representation of the original information. This allows for the extraction of the primary features of the original data and the removal of correlations between the features, or unnecessary data. The expansion serves as the foundation for the PCA.

A popular orthogonal transform used to emphasize differences and decreasing correlation is the transform. w is represented by the weighted average of m orthogonal basis vectors, assuming that W is an m-dimensional arbitrary variable.

$$w = \sum_{j=1}^{m} \alpha_j \varphi_j \tag{2}$$

α_j-Weighting coefficient, φ_j-Orthogonal basis vector and $\varphi_j$provides,

$$\varphi_j^S \varphi_i = \begin{cases} 1\ j = i \\ 0\ j \neq i \end{cases} \tag{3}$$

The matrix for Equation (2) is represented as,

$$w = (\varphi_1, \varphi_2, \dots.., \varphi_m) \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} = \Phi\alpha \tag{4}$$

The $\alpha = (\alpha_1, \alpha_2, \dots..., \alpha_m)^S, \Phi = (\varphi_1, \varphi_2, \dots, \varphi_m)$ in the calculation are orthogonal matrices, corresponding to $\Phi^S\Phi = 1$. The subsequent equation may be constructed by pre-multiplying Φ^S by two sides of Equation (4), where Φ is an orthogonal matrix,

$$\alpha = \Phi^S \times w \tag{5}$$

The entire autocorrelation matrix for w is assumed to be,

$$Q = F[w \times w^S] \tag{6}$$

The following Equation (7) is produced when Equation (4) is submitted into Equation (6),

$$Q = F\left[\Phi\alpha\alpha^S \times \Phi^S\right] = \Phi Q[\alpha\alpha^S]\Phi^S \tag{7}$$

The subsequent calculation should be met if there is no relationship between the elements of the essential vector α.

$$F\left[\alpha_i\alpha_l^S\right] = \begin{cases} \lambda_i & i = l \\ 0 & i \neq l \end{cases} \tag{8}$$

Equation (8) is expressed as a rectangular form.

$$FF\{\alpha\alpha^S\} = \begin{bmatrix} \lambda_1 & & 0 \\ & \cdots & \\ 0 & \lambda_m & \end{bmatrix} = \Lambda \tag{9}$$

The following calculation may be derived by substituting Equation (9) into Equation (7),

$$Q = \Phi\Lambda\Phi^S \tag{10}$$

The two sides of the equation above are post-multiplied by Φ. The subsequent equation may be generated since Φ is an orthogonal matrix.

$$Q\Phi = \Phi\Lambda\Phi^S\Phi = \Phi\Lambda \tag{11}$$

That is:

$$Q\varphi_i = \lambda_i\varphi_i (i = 1,2, \dots.., m) \tag{12}$$

According to Equation (12), λ_i is the eigenvalue of the autocorrelation matrix Q, and φ_i is the associated eigenvector. Several eigenvalues are orthogonal to the associated eigenvectors where Q is a real pair matrix.

3.4. Efficient osprey optimized adjustable random forest (EOO-ARF)

An innovative approach called EOO-ARF was developed to be able to better predict student levels of physical fitness and health. This model includes the optimization layer inspired by the osprey hunting behavior and utilizes the strength of the chosen random forest approach that is widely utilized for regression tasks. Within this structure, an optimization technique mimics the effective foraging of ospreys and is employed to optimize the random forest's adjustable parameters, including the number of trees, and the maximum depth. This flexibility allows the algorithm to the constantly varying datasets, leading to improved relevancy of the model on the forecast of the students' levels of fitness and health.

EOO-ARF can help to determine the general health condition of students by combining various variables like biometric characteristics, food habits, and the level of exercise. Ensemble learning using random forest avoids overfitting and ensures the strength of the decision-making process. The reliability and effectiveness of the model provide a basis for application by educational organizations in promoting improved levels of health, including student's quality of life as well as academic performance.

3.4.1. Adjustable random forest

The RF method improves the prediction accuracy of the decision tree node splitting technique by optimizing an adaptive parameter selection procedure. Different decision trees will result from choosing different node-splitting strategies for the same data set due to the various features. The result indicates that random forest prediction demonstrates varying levels of efficiency. To create a new splitting rule for selecting and partitioning node characteristics, it is recommended that, after the decision tree is constructed, the best feature for dividing the nodes be identified and the node-splitting method be expressed as a linear combination.

The $Gini$ index and information gain, derived by dividing the sample set C by features b, are displayed using the node splitting equation.

$$Gain(C,b) = Ent(C) - \sum_{u=1}^{U} \frac{|C^u|}{|C|} Ent(C^u) \tag{13}$$

$$Gini(C,b) = \sum_{u=1}^{U} \frac{|C^u|}{|C|} Gini(C^u) \tag{14}$$

where C^u denotes all the samples C that have a b^u value on the feature b and are located in the u branch node.

$$Ent(C) = -\sum_{l=1}^{|z|} o_l log_2 o_l \tag{15}$$

$$Gini(C) = -\sum_{l=1 l' \neq l}^{|z|} olol' = 1 - \sum_{l=1}^{|z|} ol^2 \tag{16}$$

The following Equation (17) is the combination of the node splitting equation and the adaptive parameter-choosing procedure, which should be used to target the greater purity of the data set following separation.

$$G = \min_{\alpha,\beta \in Q} E\{C,b\} = \alpha Gini(C,b) - \beta Gain(C,b)$$
$$\text{s.}t.\begin{cases} \alpha + \beta = 1 \\ 0 \leq \alpha, \beta \leq 1 \end{cases} \tag{17}$$

The weight coefficient of feature splitting is represented as α, β. In this stage, G has a very low value. The best combination of parameters is obtained by using the adaptive parameter-choosing technique. This indicates the best node division conditions to enhance the prediction.

Performance is evaluated using the accuracy rate and the prediction error rate. Equation (18) defines the prediction error rate for sample C.

$$F(e;C) = \frac{1}{n}\sum_{j=1}^{n} JJ\left(e\left(w_j\right) \neq z_j\right) \tag{18}$$

Equation (19) defines the accuracy rate,

$$acc(e;C)\frac{1}{n}\sum_{j=1}^{n} JJ\left(e\left(w_j\right) = z_j\right) = 1 - F(e;C) \tag{19}$$

3.4.2. Efficient osprey optimization algorithm

To enhance accuracy in predicting models regarding students' health and fitness, a new innovative approach called the EOO algorithm was developed. Based on the hunting tactics of ospreys, the EOOA can identify optimal parameters and patterns, thereby improving precision in health assessments and providing personalized workout recommendations for students.

Osprey optimization algorithm

The Osprey Optimization Algorithm (OOA) approach may be used to emulate Osprey behavior. This hunting strategy assists the osprey in locating its prey, hunting it, and then putting the prey where it is to be eaten. A sample osprey behavior is used to exemplify the proposed OOA approach which is subdivided into two processes namely exploitation and exploration.

The OOA approach is a population-based technique that uses a repetition-based process to provide suitable solutions based on the population members' ability to explore the problem-solving area. The significance of the problem variable is determined by every osprey in the OOA population based on its position in the SS. The problem is formally expressed utilizing vectors, and each osprey represents a possible solution. All ospreys comprise the OOA population, which may be simulated using Equation (20). At the beginning of the OOA application, the starting point in the SS is arbitrarily initialized using Equation (21).

$$W = \begin{bmatrix} W_1 \\ \vdots \\ W_j \\ \vdots \\ W_M \end{bmatrix}_{M\times n} = \begin{bmatrix} W_{1,1} & \cdots & W_{1,i} & \cdots & W_{1,n} \\ \vdots & \ddots & \vdots & ⋰ & \vdots \\ W_{j,1} & \cdots & W_{j,i} & \cdots & W_{j,n} \\ \vdots & ⋰ & \vdots & \ddots & \vdots \\ W_{M,1} & \cdots & W_{M,i} & \cdots & W_{M,n} \end{bmatrix}_{M\times n} \quad (20)$$

$$W_{j,i} = ka_i + q_{j,i}.(va_i - ka_i) \quad (21)$$

where,M—Number of ospreys, $W_{j,i}$-i^{th} size (Problem-related variables), W—Osprey population matrix, n—Number of variable problems, and W_j-j^{th}osprey.ka_i—Lower bound,$q_{j,i}$—Random value between 0 and 1, and va_i—Upper bound.

According to Equation (22), a vector may be used to indicate the assessed values for the problem's OF.

$$E = \begin{bmatrix} E_1 \\ \vdots \\ E_j \\ \vdots \\ E_M \end{bmatrix}_{M\times n} = \begin{bmatrix} E(W_1) \\ \vdots \\ E(W_j) \\ \vdots \\ E(W_M) \end{bmatrix}_{M\times k} \quad (22)$$

The OF value for the j^{th} osprey is represented by E_j, whereas E is a vector of OF values. The assessed value for the OF is the primary factor used to measure the quality of possible solutions. As a result, the greatest value obtained for the OF represents the ideal candidate solution or the ideal member, and the worst value discovered for the OF represents the ideal candidate solution or the worst member. The ideal solution candidate should be modified in conjunction with the current locations in the SS, which are altered with each iteration.

Exploration stage

Ospreys can find fish underwater due to their outstanding hunting skills and sharp vision. They find the fish and then go into the water to seek and attack it. The first stage of osprey population regeneration in the OOA has been examined using models of osprey behavior in nature. When osprey locations are greatly modified in the SS by simulating attacks on fish, the OOA's power exploration in identifying optimum locations and escapes from optimal locations is enhanced. The OOA model describes each underwater fish as an additional osprey's location in the SS, with every osprey having a greater OF value. The fish composition for every osprey was determined using Equation (23).

$$FP_j = \{W_j | l \in \{1,2, \ldots \ldots, M\} \wedge E_l < E_j \cup \{W_{best}\} \quad (23)$$

$$W_{j,i}^{OJ} = w_{j,i} + q_{j,i}.\left(SF_{j,i} - J_{j,i}.w_{j,i}\right) \quad (24)$$

$$W_{j,i}^{OJ} = \begin{cases} W_{j,i}^{OJ}, ka_i \leq W_{j,i}^{OJ} \leq va_i \\ ka_i, \quad W_{j,i}^{OJ} < ka_i \\ va_i, \quad W_{j,i}^{OJ} > ka_i \end{cases} \quad (25)$$

$$W_j = \begin{cases} W_j^{OJ}, E_j^{OJ} < E_j \\ W_j, else \end{cases} \tag{26}$$

where, FP_j—Collection of fish locations for the j^{th} osprey, and W_{best}—Best osprey solution.

When the osprey randomly finds one of these fish, it attacks. The relevant osprey's modified position is determined using Equation (24) by simulating its approach to the fish. This new location replaces the osprey's initial one if it increases the value of the OF, based on Equation (26). Based on the original OOA stage, it W_j^{OJ}represents the prey's new position.$q_{j,i}$. Represents a random integer in the range [0, 1], $J_{j,i}$ represents an arbitrary number from the set {1, 2}, and $W_{j,i}^{OJ}$ represents its i size. The fish chosen for the j^{th} prey is SF_j, the OF value is E_j^{OJ}, and the i size is $SF_{j,i}$.

Exploitation stage

The initial step in simulating osprey behavior in OOA is to utilize equation (27) to determine a new random location for each population member suitable for fish eating. The relevant osprey's prior location is replaced by the new location, if the OF value remains constant at this new location, based on Equation (28).

$$W_{j,i}^{O2} = w_{j,i} + \frac{ka_i + q.(va_i - ka_i)}{s}; j = 1,2, \ldots\ldots, M; i = 1,2, \ldots.., n; s = 1,2, \ldots.., S \tag{27}$$

$$W_{j,i}^{O2} = \begin{cases} W_{j,i}^{O2}, ka_i \le W_{j,i}^{O2} \le va_i \\ ka_i, \quad W_{j,i}^{O2} < ka_i \\ va_i, \quad W_{j,i}^{O2} > ka_i \end{cases} \tag{28}$$

$$W_j = \begin{cases} W_{j,i}^{O2}, \quad E_j^{O2} < E_j \\ W_j, \quad else \end{cases} \tag{29}$$

where W_j^{O2} represents the prey's new location according to the subsequent OOA stage. E_j—Fish was chosen for the j^{th} prey, $W_{j,i}^{O2}$ - i^{th} size, q—Random integer between 0 and 1,E_j^{O2}—Value of the OF,S—Total number of iterations, and s—Method's iteration counter.

Lévy flight optimization

A random walk is a random event where waves or particles follow arbitrary paths. Random walks were initially used to explain how particles moved through fluids (Brownian motion). Lévy flight is a specific type of general random walk where a heavy-tailed probability distribution describes the stride length throughout the walk. They can characterize all scale-invariant random procedures.

$$K(W_i) \approx |W_i|^{1-\alpha} \tag{30}$$

where, $1 < \alpha \le 2$—Exponential power, and W_i—Flight length.

Equation (31) defines the probability density of the Lévy stable procedure in integral form.

$$e_K(w;\alpha,\gamma) = \frac{1}{\pi}\int_0^{\infty} \exp(-\gamma r^{\alpha}) \cos(rw)\, dr \tag{31}$$

Where γ chooses the scale units and α is the distribution index, which regulates the process's scale characteristics. Especially in certain instances may integrals in equation (30) be solved analytically. A Gaussian distribution is represented when $\alpha = 2$, while a Cauchy distribution is represented when $\alpha = 1$. When extremely large values, in Equation (32), the series expansion approach often needs to be used to solve the integral in Equation (30),

$$e_K(w;\alpha,\gamma) = \frac{\gamma\Gamma(1+\alpha)\sin\left(\frac{\alpha\pi}{2}\right)}{\pi W^{(1+\alpha)}}, w \to \infty \tag{32}$$

Where, Γ—Gamma function.

To produce stable Lévy procedures for actual values of the index distribution (α) between 0.3 and 1.99, a precise and rapid approach was suggested. The Lévy distribution in Equation (33) serves as the foundation for the random number generation technique.

$$Levy(\alpha) = 0.05 \times \frac{w}{|z|^{1/\alpha}} \tag{33}$$

$$w = Normal\left(0, \sigma_w^2\right) \tag{34}$$

$$z = Normal\left(0, \sigma_z^2\right) \tag{35}$$

$$\sigma_w = \left[\frac{\Gamma(1+\alpha)\sin\left(\frac{\alpha\pi}{2}\right)}{\Gamma\left(\frac{(1+\alpha)}{2}\right)\alpha 2^{\frac{(\alpha-1)}{2}}}\right]^{1/\alpha} \quad and \ \ \sigma_w = 1 \quad \alpha = 1.5 \tag{36}$$

When two commonly distributed variables, $w \ and \ z$, have SDs of $\sigma_w \ and \ \sigma_z$, respectively.

Efficient osprey optimization (EOO)

The OOA is an efficient solution for unimodal and low-dimensional optimization issues. However, OOA's results are not very excellent when it comes to multimodal and high-dimensional optimization challenges. The EOO method is suggested to enhance exploration, OOA convergence, exploitation, and local optimal avoidance. The suggested approach integrates Lévy flight optimization with the OOA. By optimizing Lévy flight, search agent variability could be maximized, ensuring that the technique can effectively travel the search area and minimize local avoidance. The Lévy flying algorithm can find the global optimum solution in a significant, complicated SS. Its ability to integrate extensive research with the exploitation of appealing tracts is the basis for development. Moreover, ranging from discrete optimization to continuous optimization problems with complexity parameters can be solved using the Lévy flight method. Lévy flying paths are relevant to the optimization of the transition from the OOA exploration phase to the exploitation phase. Furthermore, the Lévy flight approach's adaptation to OOA improves the capacity to resolve difficult parameter issues. Consequently, following the position change, the Osprey's location is updated using the Lévy flying path. By updating Equation (27) and adding Equation (33) to Equation (27), the suggested EOO technique is a variant of the OOA approach, as shown in the following Equation (37),

$$W_{j,i}^{O2} = w_{j,i}\frac{Levy(\alpha)}{s}; j = 1,2, \ldots, M; i = 1,2, \ldots, n; s = 1,2, \ldots, S \quad (37)$$

4. Result

The suggested approach was tested on the Windows 11 laptop that has an Intel i5 9th Gen CPU, 16 GB of RAM, and an environment configured for Python 3.10.1. Key libraries used included TensorFlow and Keras for machine learning (ML), NumPy and Pandas for data processing, and Matplotlib for visualization. scikit-learn was utilized for model evaluation, while OpenCV supported any necessary computer vision tasks. The setup ensured efficient model training and real-time predictions.

The effectiveness of the suggested strategy is compared to conventional approaches, including Categorical Boosting (CatBoost) [19], and Crayfish Optimization-driven Adaptive-Weighted AdaBoost (CO-AWAdaBoost) [20].

Figure 2 presents a breakdown of student health and fitness metrics based on bio-sensor data, categorizing students into three groups: Healthy, At Risk, and Unhealthy. Most students fall under the healthy category across various parameters. For heart rate, 70% of students are healthy, while 20% are at risk and 10% are unhealthy. In terms of step count, 60% are healthy, 25% are at risk, and 15% are unhealthy. Regarding physical activity, 75% of students are healthy, with 15% at risk and 10% unhealthy. Sleep duration shows that 65% of students are healthy, 20% are at risk, and 15% are unhealthy. Overall, while the majority of students show good health and fitness, there is a notable percentage who fall into the at-risk or unhealthy categories, particularly in terms of step count, physical activity, and sleep duration. This highlights the need for targeted interventions to promote physical activity, improve sleep habits, and address cardiovascular health.

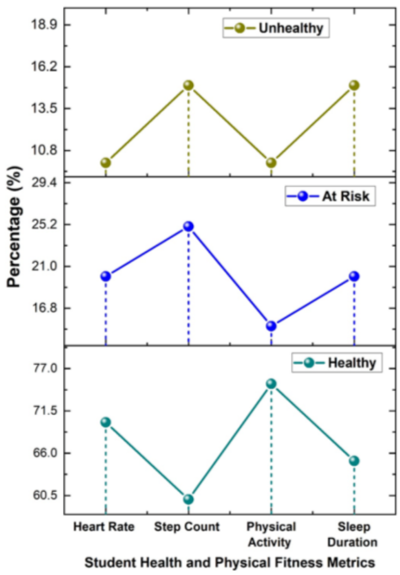

Figure 2. Output of student health and physical fitness metrics.

Figure 3 compares the F1 scores of three methods for predicting healthy students based on health and fitness data. The F1-score is an important metric that balances precision (accuracy of positive predictions) and recall (ability to identify all true positives), reflecting how well a model reduces false positives and false negatives. The F1-score of the CatBoost method reached 84.3% and shows satisfaction, as there is scope for improvement. The F1-score values of the CO-AWAdaBoost method increased significantly to 97.88%, which means that the precision and recall values are balanced with each other. F1-score values of EOO-ARF [Proposed] improved over the others and obtained the highest value as 98.13%. This outcome demonstrated that the EOO-ARF method was indeed a good predictor of proper, reliable forecasts and hence the most effective model in determining student health in this study.

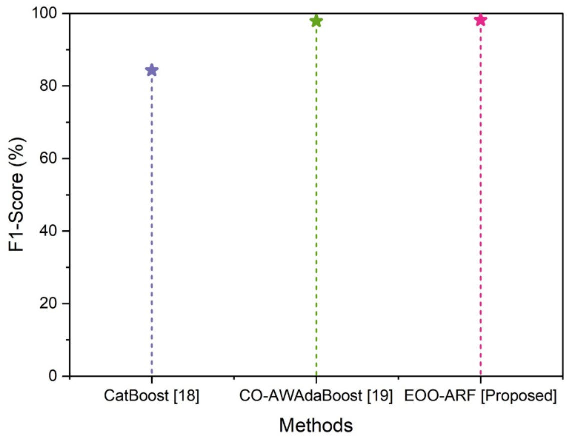

Figure 3. F1-score comparison between the proposed method and existing methods.

The accuracy measures the percentage of the total number of forecasts with the correct predictions. It supports the identification of fitness patterns and accurate focused health action since it provides an assessment of the effectiveness of the model. The CatBoost model achieves an accuracy of 86.7%, while CO-AWAdaBoost shows a slightly higher accuracy of 98.25%. However, the model proposed in this article—EOO-ARF—has been an outperformer with an accuracy of 98.44%. The higher accuracy suggests that the EOO-ARF approach happens to be a better alternative than traditional models and would be able to capture some of the complex health data patterns. This makes it a good option for monitoring fitness and addressing targeted health interventions. Therefore, the EOO-ARF model presents an applicative potential in terms of a real-world application where exact health predictions are necessary to improve wellness outcomes (**Figure 4**).

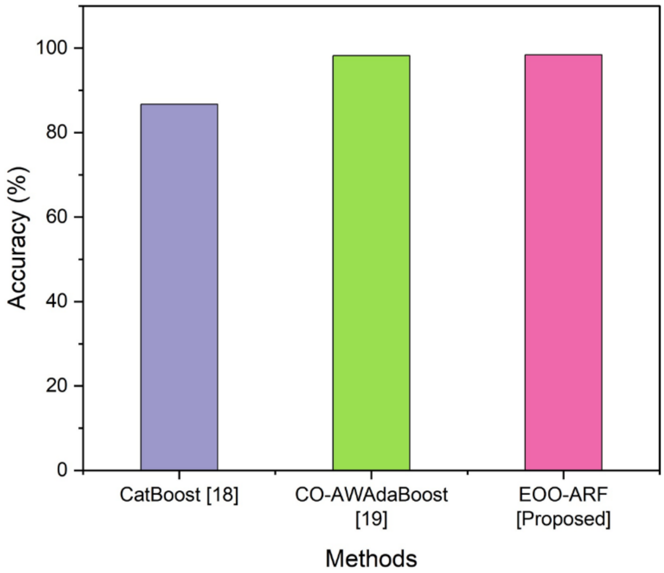

Figure 4. Accuracy performance in prediction comparison between existing and proposed methods.

As **Figure 5** shows, the recall rates of the three methods used in the study are presented: CatBoost, CO-AWAdaBoost, and the proposed EOO-ARF approach. Recall represents the ratio of correctly classified true-positive instances, which is a performance metric, and describes how well the model can detect students requiring fitness increase or health improvements. This figure clarifies that the EOO-ARF method proposed here is more efficient than the conventional methods by providing a higher value of recall, which is 98.2% compared to CatBoost of 82% and CO-AWAdaBoost with a value of 97.86%. Therefore, this means that the EOO-ARF method is more efficient in identifying students who need health interventions. The proposed method has a higher recall value, showing the number of relevant cases captured is much more, which in health-related prediction tasks is a valuable signal. It visually supports this improvement by showing the recall performance of the different models. **Table 2** shows that the Overall performance of EOO-ARF method proves the most effective in terms of accurately predicting students' health needs.

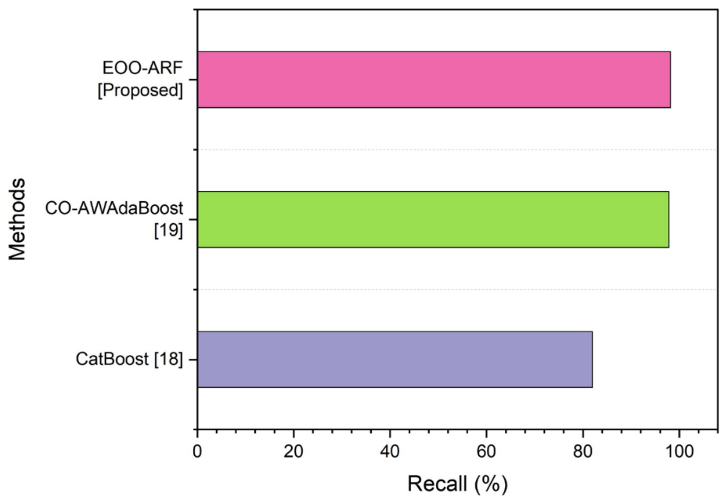

Figure 5. Recall the comparison within the proposed method and existing methods.

Table 2. Overall result performance.

Methods	Accuracy (%)	Recall (%)	F1-score (%)
CatBoost [19]	86.7%	82%	84.3%
CO-AWAdaBoost [20]	98.25%	97.86%	97.88%
EOO-ARF [Proposed]	98.44%	98.2%	98.13%

5. Discussion

The current algorithms, including CatBoost and CO-AWAdaBoost, have shortcomings that the EOO-ARF addresses. Though efficient, CatBoost finds difficulty with handling multidimensional health data and tends to have low recall, missing some at-risk students. Though CO-AWAdaBoost is better than CatBoost, it leaves room for improvement between precision and recall, sometimes misclassifying the health status. Moreover, both of the above methods are vulnerable to overfitting and not adaptive to fluctuations in real-time health data. The suggested EOO-ARF model reduces the limitations as it uses an improved selection process in features and a more sophisticated hybrid approach that improves accuracy and recall. This model adaptively learns to effectively capture the intricate patterns in the health and fitness data, making sure there are even more at-risk students identified and interventions that are more accurately targeted. The high F1-score of the proposed EOO-ARF method will also ensure that the balance between precision and recall is optimized, reducing errors and providing more reliable predictions than the existing methods. This makes the proposed model a better tool for real-time health monitoring and intervention planning.

6. Conclusion

In the last few decades, growing awareness about students' health and physical fitness has drawn the attention of educators, health professionals, and governments toward it. In this study, a new EOO-ARF approach is proposed for the prediction of students' health and their level of physical activity. Student health and physical fitness data have been collected from the Kaggle source. The efficiency of the proposed technique is validated through F1-score (98.13%), accuracy (98.44%), and recall (98.2%). These experiments clearly show that the EOO-ARF approach does better in the case of health and fitness prediction than the traditional approaches such as CatBoost and CO-AWAdaBoost. Further, it can be noticed that a higher F1 score was associated with better precision and recall, while accuracy and recall rates portrayed the potential of the model to well identify those students who need health interventions. Overall, the EOO-ARF model proves to be a reliable tool for monitoring student health, with significant potential for real-world applications in fitness tracking and targeted health interventions. There are several limitations in the self-reported information, lack of continuous monitoring, and possible sample biases minimized generalization due to geographical differences between samples, and possible external influences on students' health and fitness levels. Th description of abbreviation as shown in **Table A1** (Appendix). Future studies ought to include self-reporting biases, longitudinal research, and other sample methods using different samples while improving on generalizability across areas, or even controlling the external factors that may intervene in the health and fitness results.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Zhamardiy, V., Griban, G., Shkola, O., Fomenko, O., Khrystenko, D., Dikhtiarenko, Z., Yeromenko, E., Lytvynenko, A., Terentieva, N., Otravenko, O. and Samokish, I., 2020. Methodical system of using fitness technologies in physical education of students.International Journal of Applied Exercise Physiology, (9 (5)), pp.27-34.,
2. Zhang, K., Cadenas-Sanchez, C., Fraser, B. and Lang, J.J., 2024. Health-Related Physical Fitness Assessment in School Settings. InPromotion of Physical Activity and Health in the School Setting(pp. 107-132). Cham: Springer Nature Switzerland.
3. Rabaya, R.R., Mejarito, C., Esmael, N. and Eligue, J., 2024. Physical Fitness Exercise: Student’s Attitude and Engagement.Psychology and Education: A Multidisciplinary Journal,20(6), pp.766-786.
4. Suyati, E.S., Sonedi, S., Bulkani, B., Fatchurahman, M., Nurbudiyani, I. and Setiawan, M.A., 2022. The relationship between physical fitness and socioeconomic status and students’ learning achievement. Retos: nuevastendenciaseneducaciónfísica, deporte y recreación, (46), pp.494-500.
5. Lee, H.S. and Lee, J., 2021. Applying artificial intelligence in physical education and future perspectives.Sustainability,13(1), p.351.
6. Zhang, Z., He, Z. and Chen, W., 2022. The relationship between physical activity intensity and subjective well-being in college students.Journal of American College Health,70(4), pp.1241-1246.
7. deBruijn, A.G., de Greeff, J.W., Temlali, T.Y., Oosterlaan, J., Smith, J. and Hartman, E., 2023. Objectively measured physical activity during primary school physical education predicts intrinsic motivation independently of academic achievement level.British Journal of educational psychology,93, pp.90-112.
8. Li, K., 2024. Using biosensors and machine learning algorithms to analyse the influencing factors of study tours on students’ mental health.Molecular & Cellular Biomechanics,21(1), pp.328-328.
9. Souri, A., Ghafour, M.Y., Ahmed, A.M., Safara, F., Yamini, A. and Hoseyninezhad, M., 2020. A new machine learning-based healthcare monitoring model for student’s condition diagnosis in Internet of Things environment.Soft Computing,24(22), pp.17111-17121.
10. Chao, Z., Yi, L., Min, L. and Long, Y.Y., 2024. IoT-Enabled Prediction Model for Health Monitoring of College Students in Sports Using Big Data Analytics and Convolutional Neural Network.Mobile Networks and Applications, pp.1-18.
11. Guo, J., Wan, B., Zheng, S., Song, A. and Huang, W., 2022. A Teenager Physical Fitness Evaluation Model Based on 1D-CNN with LSTM and Wearable Running PPG Recordings.Biosensors,12(4), p.202.
12. Fan, J., Yang, Y. and Liu, J., 2024. Research on the Application of Decision Tree and Correlation Analysis Algorithm in College Students’ Physical Fitness Analysis.International Journal of High Speed Electronics and Systems, p.2440019.
13. Min, X., Zeyu, L., Zheng, S. and Xunzi, L., 2022, July. Prediction of College Students’ Physical Test Effect Based on BP Neural Network. In2022 5th International Conference on Data Science and Information Technology (DSIT)(pp. 1-7). IEEE.
14. Pradeepa, M., Jamberi, K., Sajith, S., Bai, M.R. and Prakash, A., 2022, October. Student health detection using a machine learning approach and IoT. In2022 IEEE 2nd Mysore sub section International Conference (MysuruCon)(pp. 1-5). IEEE.
15. Tyulepberdinova, G., Mansurova, M., Sarsembayeva, T., Issabayeva, S. and Issabayeva, D., 2024. The physical, social, and mental conditions of machine learning in student health evaluation.Journal of Computer Assisted Learning,40(5), pp.2020-2030.
16. Wang, C., 2024, August. Student Physical Health Assessment Based on Puzzle Optimization Algorithm with Artificial Neural Network. In2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)(pp. 1-4). IEEE.
17. Yang, X. and Zeng, H., 2024, May. An Effective Prediction Method of Physical Fitness for College Students Based on GWO-GRU Model. In2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE)(pp. 983-987). IEEE.

18. Kaggle. Biosensor-Student Health Fitness Data. Available online: https://www.kaggle.com/datasets/ziya07/biosensor-studenthealthfitnessdata (accessed on 5 July 2024).
19. Liu, Y., 2020, September. College students' physical fitness test data analysis, visualization and prediction using data mining techniques. InJournal of Physics: Conference Series(Vol. 1631, No. 1, p. 012121). IOP Publishing.
20. Liao, X., Lei, X. and Sun, P., 2024. A Systematic Study of Physical Fitness Assistance Training for Adolescents Based on Kinect Motion Capture.Journal of Electrical Systems,20(9s), pp.1454-1463.

Appendix

Table A1. Description of abbreviations.

1D	**One dimensional**	**IoT**	**Internet of Things**
OF	Objective function	BP	Backpropagation
BDNN-CSMHPM	Biosensor-based and deep neural network-based College Student Mental Health Prediction Model	POAANN	Puzzle Optimization Algorithm with Artificial Neural Network
GWO	Grey Wolf optimistic	HDS	High-dimensional space
CNN	Convolutional neural networks	DT	Decision Tree
PPG	Photoplethysmography	SS	Search space
SVM	Support vector machine	MCC	Matthews correlation coefficient
LDS	Low-dimensional space	LSTM	Long Short-Term Memory
ML	Machine learning	GRU	Gated Recurrent Unit
SD	Standard deviation		