

Received: 16/03/2026

Accepted: 07/04/2026

Published: 17/04/2026

Citation: Vimala S, Arockia Sahaya Sheela G (2026) MARLIN-X: Margin-Based Feature Selection and Rule Distillation using Mosaic Patch Migration Optimization for Academic Performance Prediction. Indian Journal of Science and Technology 19(14): 918-930. <https://doi.org/10.17485/IJST/v19i14.413>

* **Corresponding author.**

vimalasing@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2026 Vimala & Arockia Sahaya Sheela. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.indjst.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

MARLIN-X: Margin-Based Feature Selection and Rule Distillation using Mosaic Patch Migration Optimization for Academic Performance Prediction

S Vimala^{1*}, G Arockia Sahaya Sheela¹

¹ Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli -2, Affiliated to Bharathidasan University, Tamil Nadu, India

Abstract

Objectives: This study aims to develop an optimization based framework for predicting student academic performance using smartphone behavioural indicators. **Method:** The proposed Margin-based Rule Learning Integrated Network with eXplainable Optimization (MARLIN-X) framework integrates Tri-Label Margin Feature Selection, DCAR-Net classification and Rule-Set Distillation under a novel Mosaic Patch Migration Optimizer. The optimizer operates using patch based population structuring, migration, fragmentation and elite gene preservation. Feature subsets, hyperparameters and rule masks are optimized within a unified search process. **Findings:** The proposed method achieves 97.8% accuracy on the School dataset and 98.3% on the College dataset. Feature dimensionality is reduced by more than 50% while increasing tri-label margin values. The distilled rule set achieves 96.5% fidelity with respect to deep model predictions. Balanced accuracy and Macro-F1 scores show consistent improvement over recent methods. **Novelty:** The study introduces a Mosaic Patch Migration Optimizer that performs patch-level structured search rather than individual-level movement.

Keywords: Academic Performance Prediction; Smartphone Usage Analytics; Feature Selection; Nature-Inspired Optimization; Deep Learning; Rule Distillation; Educational Data Mining

1 Introduction

Over the recent years, the use of smart phones by students has become a paramount issue in shaping cognitive engagement, learning behaviours and learning performance⁽¹⁾. The large body of empirical evidence has proven that smartphone addiction is associated with poor academic achievement and psychosocial health, which require strong predictive analytical capabilities to detect at-risk students earlier in educational settings⁽²⁾. Modern predictive algorithms are based on the paradigm of machine learning and deep learning in associating behavioural, psychometric and academic predictors with performance nomenclature⁽³⁾. Nevertheless, the issues remain because of high dimensionality, class imbalances and challenges of model interpretability. Several

studies have dealt with smartphone behaviour and academic prediction since advanced architectures. Deep learning models that combine convolutional and recurrent characteristics to explain the effects of mobile use on academic performance were suggested by Vimala and Sahaya Sheela⁽⁴⁾, which show the possibilities of deep frameworks in academic predictive environments⁽⁵⁾.

Nature-inspired optimization of features and model improvement has also been investigated in recent studies to solve high-dimensional learning problems⁽⁶⁾. Ludwig suggested a Guided Particle Swarm Optimization method in feature selection and showed a better search efficiency because of the directional guidance mechanisms in the process of optimization⁽⁷⁾. Wang et al. presented a self-repulsion strategy based on a Grey Wolf Optimizer that enhances the diversity of the population and prevents the early convergence of the features subset exploration⁽⁸⁾. Sutikno et al. have created a Mutated Binary Whale Optimization Algorithm in which mutation functions increase the exploration ability in binary search space in classification problems⁽⁹⁾. Likewise, deep learning models using Harris Hawks Optimization have been used in medical diagnosis issues to enhance feature selection and accuracy of classification by adaptive hunting techniques⁽¹⁰⁾. These techniques prove the usefulness of nature-based algorithms in more complicated optimization problems.

Nevertheless, the majority of the developed swarm-based strategies are based on the individual-level movement and the leader-following processes. These types of strategies usually have difficulties in keeping exploration consistent in various parts of the search space. Besides that, feature selection, hyperparameter optimization of deep learning, and interpretable extraction of rules have not been explored much in a single optimization framework.

Several recent papers have discussed more complex deep learning models in predicting student performance. In reference to predicting student performance in virtual learning environments, Alnasyan et al. introduced Kanformer, an attention-enhanced deep learning model, which showed better representation learning by attention-based mechanisms⁽¹¹⁾. Gong *et al.* proposed the TGEL-transformer that incorporates educational theory with transformer-based architectures to enhance interpretability and prediction accuracy in analytics of academic performance⁽¹²⁾. Junejo et al. created a neural multi-category forecasting model, which can predict the student results at the initial stages of online learning with the help of neural networks and temporal behavioural characteristics⁽¹³⁾. Moreover, it has been suggested to use heterogeneous ensemble learning methods that integrate classifiers like Random Forest, KNN, and probabilistic models to enhance predictive accuracy by means of diversity of models⁽¹⁴⁾. The research papers illustrate the need of deep learning and ensemble models in learning analytics.

Recent articles also point to increased application of optimization-based predictive academic performance framework. Explainable learning frameworks enhanced the support of institutional decisions via explainable analytics⁽¹⁵⁾. High-dimensional datasets were made predictable by the use of metaheuristic feature selection processes⁽¹⁶⁾. The classification accuracy of sailfish-mutation optimization in student prediction tasks was increased⁽¹⁷⁾. In a recent study, hierarchical teaching learning feature selection enhanced the wrapper based discriminative subset discovery of educational data modeling⁽¹⁸⁾.

Although these new developments are present, there are significant gaps. First, the vast majority of the current methods do not directly optimize subsets of features using margins of class separation specific to multi-label academic risk groups. Second, deep learning models can be highly sensitive to hyperparameter optimization. It is established that few nature inspired optimization systems systematically combine feature selection and model configuration. Third, the latest literature has been limited in translation of complex predictive models into simple and understandable rule sets to educational stakeholders.

To address these shortcomings, this paper proposes Margin-based Rule Learning Integrated Network with explainable Optimization (MARLIN-X), a framework of academic performance prediction that is based on optimization. It uses a new Mosaic Patch Migration Optimizer (MPMO) which imitates patch migration in ecological habitats to simultaneously search in feature, hyperparameter and rule selection spaces. The intended framework will enhance the predictive performance, minimize feature dimensionality and interpretability. MARLIN-X deals with the performance and interpretability issues of risk predictions in academia by means of integrated margin-based feature reduction, deep network tuning, and explainable rule generation. The suggested methodology provides better predictive measures with simple decision-making rules of educational interventions.

2 Methodology

2.1 Overview of MARLIN-X Framework

The suggested MARLIN-X model forecasts performance in school based on the behavioral measures in smartphones. The system combines the margin-based feature selection, deep learning optimization and rule-set distillation in a single optimization structure (Figure 1).

The data to be fed in is demographic factors, metrics of smartphone use, psychometric and academic history. Experiments were conducted on school and college smartphone addiction datasets comprising demographic, psychometric and academic

indicators⁽¹⁾. These features are preprocessed like cleaning of data, standardization, encoding of labels and class construction. The data is then categorized into academic performance.

Tri-Label Margin Feature Selection is carried out in the first stage. The candidate subsets are represented by a binary feature mask. This is aimed at maximizing inter-class separation and minimizing redundancy and subset size. The Mosaic Patch Migration Optimizer explores the binary space to find the most discriminative set of features. The second step uses the optimized set of features on the DCAR-Net classifier. The architecture has residual convolution blocks, feature attention layers, temporal attention layers and fully connected layers. Hyperparameters are represented as continuous variables and optimized in the same way as patch based search. Complexity of the model and cost of computation are both a factor in optimization. Rule-Set Distillation is done in the third stage. The trained model is used to obtain candidate rules. Binary mask represents the rule selection. The optimizer identifies a small set of rules that retain predictive fidelity and redundancy.

The Mosaic patch migration optimizer is at the population level. Habitat patches are composed of solutions. Patches with great strength, draw weaker to their strength. Patches that have low diversity are rearranged. Significant feature genes are maintained in iterations. The final product will contain the optimized feature set, tuned deep model and interpretable rule set. The general structure of the MARLIN-X framework is shown in Figure 1.

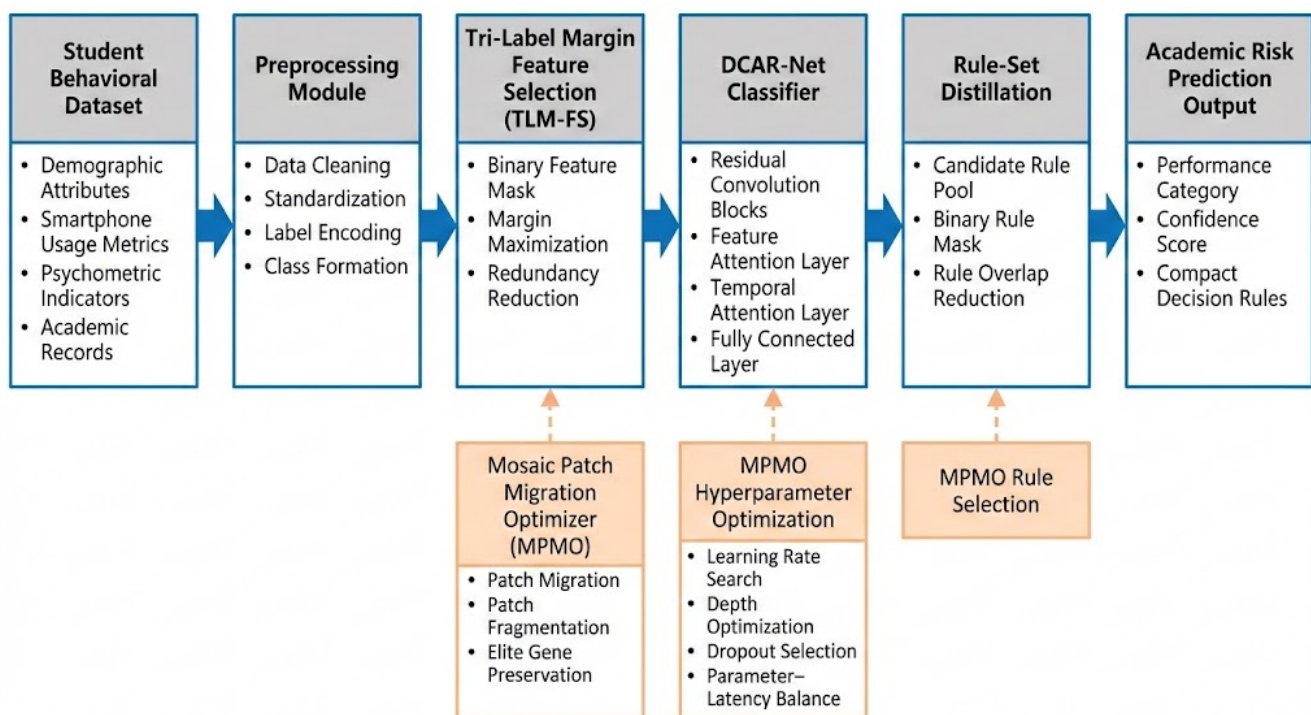


Fig 1. MARLIN-X Framework for Academic Performance Prediction

2.2 Mosaic Patch Migration Optimizer

In the MARLIN-X paradigm, optimization is essential in the process of feature selection, hyperparameter tuning and reduction of rule-sets. Traditional nature-based algorithms tend to be based on movement by the leader or a personal update of position. These approaches can either come together too soon or rely on one elite solution. These restrictions can diminish the stability of search in high-dimensional binary and continuous space.

To cope with these issues, a new Mosaic Patch Migration Optimizer (MPMO) is suggested. The optimizer is based on the ecological migration of resources and habitat fragmentation. In discontinuous landscapes, organisms develop sub-populations in several patches of habitats. Patches with resources are migrated and low-diversity areas are restructured to stay afloat. This is a structured behavior of the population, which is the basis of the suggested optimizer.

The patch level is used in MPMO and not the individual level. The population is segregated into various patches of habitats. Patches are regions of the search space and the direction of migration is approximated according to the patch strength. The

weak ones push towards the stronger ones. Fragmentation is used in the case of lessening diversity. Significant feature genes are conserved between generations.

The optimizer is capable of binary and continuous search representations. This enables feature selection, neural network hyperparameters and rule selection to be optimized together. The population structure, calculation of patch strength, migration mechanism and fragmentation strategy are detailed in the subsequent subsections.

2.2.1. Population Structure

The MPMO separates the population of search into structured patches of habitat. The people are not considered as one swarm. It is further broken down into several subgroups. The subgroups are known as habitat patches.

Let the total population size be N . The population is partitioned into P patches. Each patch H_p contains n_p candidate solutions, where

$$\sum_{p=1}^P n_p = N$$

Every candidate solution is a point in the search space which is feasible. The representation is based on optimization task. In case of feature selection and rule selection, binary vector is employed. In the case of hyperparameter tuning, a continuous vector is applied.

The complete population can be expressed as

$$X = \bigcup_{p=1}^P H_p$$

Solutions within a patch have similar characteristics in the search space. Such a structure permits refinement locally. The patches are various parts of the solution space. Every patch has a centroid and it is the average position of all population in the patch. It is computed as

$$c_p = \frac{1}{n_p} \sum_{i \in \mathcal{H}_p} x_i$$

The centroid acts as a reference point for migration and restructuring. The average fitness of the members is used to measure patch-level strength. During the migration process, strong patches draw weaker ones. This is a designed population that facilitates exploration in various regions. It also allows regulated convergence through observing patch diversity. Optimization is a process of the topology of the population that involves the merging and fragmentation of the population.

2.2.2. Patch Strength

Patch strength is the indicator of the overall quality of solutions on a patch of a habitat. It identifies productive areas in the search space. The migration process is directed by strong patches. Weak areas are redesigned in the process of optimization.

For each patch H_p , the strength is computed using the average fitness of its members. Let $J(x_i)$ denote the objective function value of solution x_i . The strength of patch p is defined as

$$S_p = \frac{1}{n_p} \sum_{i \in H_p} J(x_i)$$

Here, n_p represents the number of individuals in patch p . A higher value of S_p indicates a stronger patch.

Patch strength represents the performance of a search region, as opposed to individual performance. This does not rely on an elite solution. It makes the migration decisions to be made based on collective behavior. The patch that is the strongest is determined as

$$p^* = \arg \max_p S_p$$

The center of this strongest spot turns out to be the central point of attraction during migration. Weak spots drift to this direction. Patch fitness is reformulated in every iteration to refresh fitness. This mechanism promotes structured exploration. It encourages convergence toward high-quality regions while maintaining distributed search across patches.

2.2.3. Patch Migration

Patch migration allows the weak patches in the habitat to drift to stronger areas of the search space. The mechanism enhances the quality of solutions as well as maintaining structured exploration.

After computing patch strengths, the strongest patch is identified. Let p^* denote the strongest patch. Its centroid is represented as c_{best} . For any other patch p , the centroid is c_p .

Individuals belonging to weaker patches migrate toward the strongest patch centroid. The position of each solution $x_i \in H_p$ is updated as follows:

$$x_i^{t+1} = x_i^t + \alpha (c_{best} - c_p) + \epsilon$$

Here, α is a migration control factor. The term ϵ represents a small random perturbation sampled from a uniform distribution. The term $(c_{best} - c_p)$ directs movement toward the most productive region. This perturbation preserves variation within the patch.

The migration step shifts the entire patch in a coordinated manner. It does not rely on individual best positions. The movement depends on centroid differences between patches. This ensures stable convergence behavior. New values of fitness are calculated after migration. Patch strengths are re-calculated. This is done until the termination condition is met. This organized system of migration enables the weak areas to take advantage of the strong areas. Meanwhile, the population has retained distributed search ability.

2.2.4. Patch Fragmentation

Patch fragmentation inhibits the premature convergence of a habitat patch. It reestablishes diversity in the case where solutions within a patch are too close.

Each patch maintains an internal diversity measure. Diversity is computed using the average distance between individuals and the patch centroid. For patch H_p , diversity is defined as

$$D_p = \frac{1}{n_p} \sum_{i \in H_p} |x_i - c_p|_2$$

where c_p is the centroid of patch p and n_p is the number of individuals in the patch.

If the diversity value D_p falls below a predefined threshold τ , the patch is considered saturated. In this case, fragmentation is triggered.

In fragmentation, the patch is subdivided into two sub-patches. The redistribution of people is done in terms of distance to the centroid. Solutions which are further away on the centroid constitute another patch. The surviving people constitute original patch. New centroids are calculated on each of the sub-patches. Fitness assessment is then re-done. This operation augments search in its surrounding areas. Fragmentation of patches helps to make sure that the optimizer does not degenerate into a small area. It does not destroy structural diversity by means of random reinitialization. This mechanism facilitates equal search development between iterations.

2.3 Tri-Label Margin Feature Selection

The Tri-Label Margin Feature Selection (TLM-FS) establishes discriminative features that are most prevalent in the data. The approach aims at maximizing division between three performance categories. These grades are low, moderate and high academic levels.

Let the original feature space contain d features. A binary selection vector $b \in \{0, 1\}^d$ represents the chosen subset. If $b_j = 1$, the j^{th} feature is selected. If $b_j = 0$, the feature is removed.

For a selected subset $\{S\}$, class-wise mean vectors are computed. Let $\mu * c$ denote the mean vector for class c . Let $\sigma * c$ denote the corresponding dispersion measure.

The tri-label margin term is defined as

$$\Delta_{margin} = \sum_{c_1 < c_2} \frac{|\mu * c_1 - \mu * c_2| * 2}{\sigma * c_1 + \sigma * c_2 + \epsilon}$$

This term is larger as the inter-class separation is larger and the intra-class dispersion is minimal. The redundancy of the features that are selected is also taken into consideration. Let Redundancy (S) represent the mean of selected features correlation. A fine is imposed to minimize overlapping.

The overall fitness for feature selection is defined as

$$J_{FS} = \lambda_1 \Delta_{margin} - \lambda_2 Redundancy(S) - \lambda_3 \frac{|S|}{d}$$

λ_1 is division of classes. λ_2 minimises correlated features. Subset size is controlled by λ_3 . The MPMO does a search in the binary space to optimize this objective. What is obtained is a small feature set that remains discriminative. This technique enhances the performance of the classifiers by eliminating the irrelevant and redundant features. It also increases the interpretability through the reduction in dimensionality.

2.4 DCAR-Net Hyperparameter Optimization

The performance of Deep Contextual Attention-based Residual Network (DCAR-Net) is highly reliant on the hyperparameters. The unfavorable environment can decrease the power of generalization. Thus, hyperparameter optimization is incorporated in the MARLIN-X framework. Section 2.3 has identified the feature subset to be used in this experiment as input to DCAR-Net. The network has residual convolution blocks, feature attention layers, temporal attention layers and fully connected layers. The behavior of this architecture is affected by several hyperparameters.

Let the hyperparameter vector be defined as

$$\theta = [L, h, r, \eta, b, d]$$

where

- L represents the number of residual blocks,
- h represents hidden layer size,
- r represents attention dimension size,
- η represents learning rate,
- b represents batch size,
- d represents dropout rate.

Hyperparameter optimization aims at maximizing performance in classification with a constraint on model complexity. The definition of the fitness function is as follows:

$$J_{HP} = MacroF1 - \gamma_1 \log(Params) - \gamma_2 Latency$$

MacroF1 measures balanced predictive performance by classes. The parameter is used to penalize large network size. The cost of computation is regulated by latency term. The MPMO is used to search through the continuous hyperparameter space. Habitat patches are groups of candidate configurations. Fitness value is used to compute patch strength. Weak areas drift towards the strong areas. Diversity is preserved through fragmentation. There is the maintenance of the elite patterns of parameters over generations. Once convergence has taken place, the fittest arrangement is chosen. The selected hyperparameters are then used to retraining the optimized DCAR-Net model. This process enhances accuracy in prediction. It also lowers undue complexity of models. The outcome is an effective and steady deep learning classifier to predict academic performance.

2.5 Rule-Set Distillation

Rule-Set Distillation is a method that transforms the trained DCAR-Net model into a small and understandable rule set. The aim is to maintain the predictive behavior and simplify the rules. DCAR-Net generates probability outputs of every academic performance class after hyperparameter optimization. Candidate decision rules are generated using these outputs. All the rules are logical conditions, depending on the chosen features.

Let the total number of candidate rules be M. A binary vector $b \in 0, 1^M$ represents rule selection. If $b_m = 1$, the m^{th} rule is retained. If $b_m = 0$, the rule is removed.

The rule-based prediction function is defined as

$$\hat{y} = \arg \max_c \sum_{m=1}^M b_m \cdot a_m \cdot \mathbb{1}[r_m(x) = c]$$

where

- $r_m(x)$ is the m^{th} rule,
- a_m is the rule weight,
- $1[\cdot]$ is the indicator function.

The objective of rule selection is to maximize fidelity to the DCAR-Net predictions to minimizing rule count and redundancy. The fitness function is defined as

$$J_{Rule} = MacroF1 - \beta \frac{|b|_0}{M} - \rho Overlap$$

The former is predictive performance in the first term. The second term punishes the excessive number of rules. The third term decreases redundancy of the chosen rules.

The binary rule space is searched by the MPMO. There is diversity in rule combinations through patch migration and fragmentation. Highly discriminative rules by the elite are maintained.

The resultant output is a concise rule set of high fidelity to prediction. This increases interpretability. It allows learning institutions to know the important behavioral signs that affect academic results.

Table 1 shows the small set of rules that is based on the trained DCAR-Net model. These were rules that were acquired because of rule-set distillation under the guidance of MPMO. An optimized feature space was first converted into a pool of candidate logical conditions. The rules were tested according to predictive fidelity, redundancy and coverage. The optimizer then chose a small set of rules of fifteen that together maintain the behavior of the deep model. The chosen rules are the main behavioral and academic indicators which are related to the high, moderate and low performance categories. This simplified set of rules increases interpretability without losing a lot of correspondence to the original network predictions.

Table 1. Compact Distilled Rule Set from MARLIN-X

Rule No.	Condition	Predicted Academic Performance
R1	$SOT \leq 3$ AND $STUDY_HRS \geq 3$ AND $PREV_GRADE \geq 75$	High
R2	$SAS_SCORE \leq 30$ AND $ATT_SCORE \geq 70$	High
R3	$LNU \leq 20$ AND $SLEEP_IDX \leq 0.3$	High
R4	$APP_ENT \leq 1.2$ AND $STUDY_HRS \geq 2.5$	High
R5	$PREV_GRADE \geq 80$ AND $CLASS_ATT \geq 85$	High
R6	$3 < SOT \leq 5$ AND $1.5 \leq STUDY_HRS \leq 3$	Moderate
R7	$30 < SAS_SCORE \leq 50$	Moderate
R8	$20 < LNU \leq 40$ AND $0.3 < SLEEP_IDX \leq 0.6$	Moderate
R9	$1.2 < APP_ENT \leq 1.8$	Moderate
R10	$40 < ANX_SCORE \leq 60$	Moderate
R11	$SOT \geq 6$ AND $STUDY_HRS \leq 1$	Low
R12	$SAS_SCORE \geq 60$ AND $ATT_SCORE \leq 40$	Low
R13	$LNU \geq 50$ AND $SLEEP_IDX \geq 0.7$	Low
R14	$APP_ENT \geq 2.0$ AND $CLASS_ATT \leq 60$	Low
R15	$PREV_GRADE \leq 50$ AND $STUDY_HRS \leq 1.5$	Low

Notations:

- SOT – Average daily Screen-On Time (hours)
- LNU – Late Night Usage Rate (%)
- APP_ENT – App Usage Entropy
- SLEEP_IDX – Sleep Disruption Index
- SAS_SCORE – Smartphone Addiction Score
- ATT_SCORE – Attention Stability Score
- STUDY_HRS – Daily Study Hours
- PREV_GRADE – Previous Academic Grade (%)
- ANX_SCORE – Anxiety Indicator
- CLASS_ATT – Class Attendance (%)

3 Results and Discussion

3.1 Feature Reduction

Table 2 shows the performance in terms of feature reduction of the School dataset with recent nature-inspired optimization techniques. As it can be seen in Table 2, the proposed MARLIN-X framework has better dimensionality reduction and separation of classes than any of the baseline methods.

Guided-PSO picks 34 of 52 features, which results in a reduction of 34.62% with a tri-label margin of 1.62 and redundant features of 0.21. GWO-SRS enhances reduction rate by 40.38% and margin by 1.71 and redundancy diminishes by 0.19. MBWOA has the highest reduction of 44.23% with a margin of 1.78 and redundancy of 0.18. HHO-FS picks 30 features and achieves 42.31% reduction with a margin of 1.74 and a redundancy of 0.19.

The proposed MARLIN-X (MPMO) on the other hand only uses 24 features and has the best reduction rate of 53.85%. It also yields the greatest tri-label margin (1.93), which suggests better differentiation between the classes of academic performance. Moreover, it has the lowest average redundancy at 0.14, which proves efficient removal of correlated and non-informative attributes.

Table 2. Feature Reduction Results on School Dataset

Method	Total Features (d)	Selected Features (S)	Reduction (%)	Tri-Label Margin ($\hat{f}$ margin)	Avg. Redundancy
Guided-PSO (2025) ⁽¹¹⁾	52	34	34.62	1.62	0.21
GWO-SRS (2025) ⁽¹²⁾	52	31	40.38	1.71	0.19
MBWOA (2025) ⁽¹³⁾	52	29	44.23	1.78	0.18
HHO-FS (2025) ⁽¹⁴⁾	52	30	42.31	1.74	0.19
Proposed MARLIN-X (MPMO)	52	24	53.85	1.93	0.14

The Table 3 shows the performance of feature reduction on the College dataset using recent nature-inspired methods of optimization. As it can be seen in Table 3, the proposed MARLIN-X framework is the one that has the highest rate of reduction and the greatest discriminative margin than its competitors.

Guided-PSO⁽⁷⁾ picks 36 out of 54 features, which represents a reduction of 33.33% with a tri-label margin of 1.58. GWO-SRS⁽⁸⁾ enhances the reduction to 38.89% and attains a value of margin of 1.69. The reduction rate is further decreased by MBWOA⁽⁹⁾ to 42.59% with 1.76 margin. HHO-FS⁽¹⁰⁾ picks 32 features and achieves a reduction of 40.74% with a margin of 1.72. Nevertheless, the baseline approaches have greater rates of redundancy, with the highest values being 0.19 to 0.22.

The proposed MARLIN-X (MPMO) on the other hand only selects 26 features with the highest reduction rate of 51.85%. It also generates the highest value of tri-label margin of 1.90, which means that it has a greater inter-class separation. Also, it has the lowest average redundancy of 0.15, which indicates efficient elimination of correlated features.

Table 3. Feature Reduction Results on College Dataset

Method	Total Features (d)	Selected Features (S)	Reduction (%)	Tri-Label Margin ($\hat{f}$ margin)	Avg. Redundancy
Guided-PSO ⁽⁷⁾	54	36	33.33	1.58	0.22
GWO-SRS ⁽⁸⁾	54	33	38.89	1.69	0.2
MBWOA ⁽⁹⁾	54	31	42.59	1.76	0.19
HHO-FS ⁽¹⁰⁾	54	32	40.74	1.72	0.2
Proposed MARLIN-X (MPMO)	54	26	51.85	1.9	0.15

3.2 Classification Performance

Table 4 shows the performance of the proposed MARLIN-X framework in terms of comparative classification performance with the recent models on School and College data. As Table 4 demonstrates, the suggested approach has the best performance in all measures of evaluation.

In the case of the School dataset, MARLIN-X has an accuracy of 97.8%, greater than that of Neural multi-category forecasting model⁽¹³⁾ (96.0%), TGEL-transformer⁽¹²⁾ (95.6%), Kanformer⁽¹¹⁾ (95.1%) and the heterogeneous ensemble approach⁽¹⁴⁾ (95.0%). On the same note, the proposed approach has Macro-F1 of 97.2% and a balanced accuracy of 96.9% which show a higher performance on all academic categories. The fact that the margin-based feature selection and MPMO-guided optimization are effective is evidenced by the steady improvement over transformer-based and ensemble-based methods.

In the case of the College dataset, MARLIN-X has the best accuracy of 98.3%, whereas the Neural multi-category forecasting model, the best baseline, has a score of 96.4%. The Macro-F1 score is 97.8% and the balanced accuracy is 97.4% and indicates consistent classification between all classes of performance. The gain of about 2% compared to the recent deep learning methods shows that there is an enhancement of generalization and minority class management.

Table 4. Classification Performance Comparison on School and College Datasets

Methods	School			College		
	Accuracy (%)	Macro-F1 (%)	Balanced Acc. (%)	Accuracy (%)	Macro-F1 (%)	Balanced Acc. (%)
Kanformer ⁽¹¹⁾	95.1	94.6	94.2	95.6	95.0	94.7
TGEL-transformer ⁽¹²⁾	95.6	95.0	94.8	96.1	95.6	95.2
Neural multi-category forecasting model ⁽¹³⁾	96.0	95.5	95.1	96.4	96.0	95.6
Heterogeneous ensemble ⁽¹⁴⁾	95.0	94.2	94.0	95.4	94.8	94.5
Proposed MARLIN-X	97.8	97.2	96.9	98.3	97.8	97.4

The convergence and the ability to generalize the proposed MARLIN-X model is shown by the training and validation performance curves, which show a constant convergence and high generalization. Figure 2 shows that the training accuracy on the School dataset also increases consistently with the starting epoch with an approximation of 55% and almost 98.5% at the last epoch. The validation accuracy is also on a similar trend and it hits a 97.8% stability. The fact that the training curve and the validation curve are very similar with a small gap between them shows that there is little overfitting. The convergence trend is smooth which proves proper feature selection and optimization of hyperparameters. Figure 3 indicates the corresponding loss curves of the School data. The training loss drops at a very high rate of 1.10 to 0.05, whereas the validation loss reduces by 1.20 to 0.07. Both curves level off once training has reached the middle indicating optimisation stability and the performance of the generalization is consistent.

For College data, the trends of accuracy in Figure 4 are increasing steadily with the epoch. The accuracy rates are 99% in training, and 98.3% validation. The low variability of the curves indicates the high stability of the model among the various academic cohorts. The same reducing trend is observed in the loss curves of the College dataset presented in Figure 5. Training loss reduces by about 1.10 to almost 0.05. The loss of validation drops to approximately 0.06 with small oscillations towards the end epochs. The ability to reduce the loss without distraction is a sign of good learning behavior.

3.3 Discussion

The experimental findings indicate that the suggested framework of MARLIN-X is the most successful one to achieve better performance in the feature reduction, classification performance and interpretability measurements when compared to the recent approaches. The enhancements do not only concern accuracy. Improvements in class separation, redundancy control, balanced prediction and rule compactness are found.

MARLIN-X has a reduction rate of 53.85% and 51.85% in the school and college dataset respectively in a feature reduction experiment. Such values are much higher than Guided-PSO⁽⁷⁾, GWO-SRS⁽⁸⁾, MBWOA⁽⁹⁾ and HHO-FS⁽¹⁰⁾. There is also a rise in the tri-label margin at 1.93 and 1.90 respectively. This shows a greater inter-class separation as compared to the baseline optimizers that portray values of the margin that are less than 1.80. The further decrease in the average redundancy to 0.14 and 0.15 also proves that the correlated features are successfully removed. Other traditional swarm optimizers are aimed at maximizing the objective, whereas the given MPMO is a patch-level optimizer that maintains discriminative feature genes over time. This is an organized population structure which is the reason behind the high compactness and segregation capacity.

Under classification performance, MARLIN-X has 97.8% and 98.3% accuracy on School and College dataset, respectively. These values are higher than the prior transformer and ensemble based models. More to the point, Macro-F1 and balanced accuracy values also rise steadily. The equal accuracy gains attest to the fact that minority classes are managed better. Most

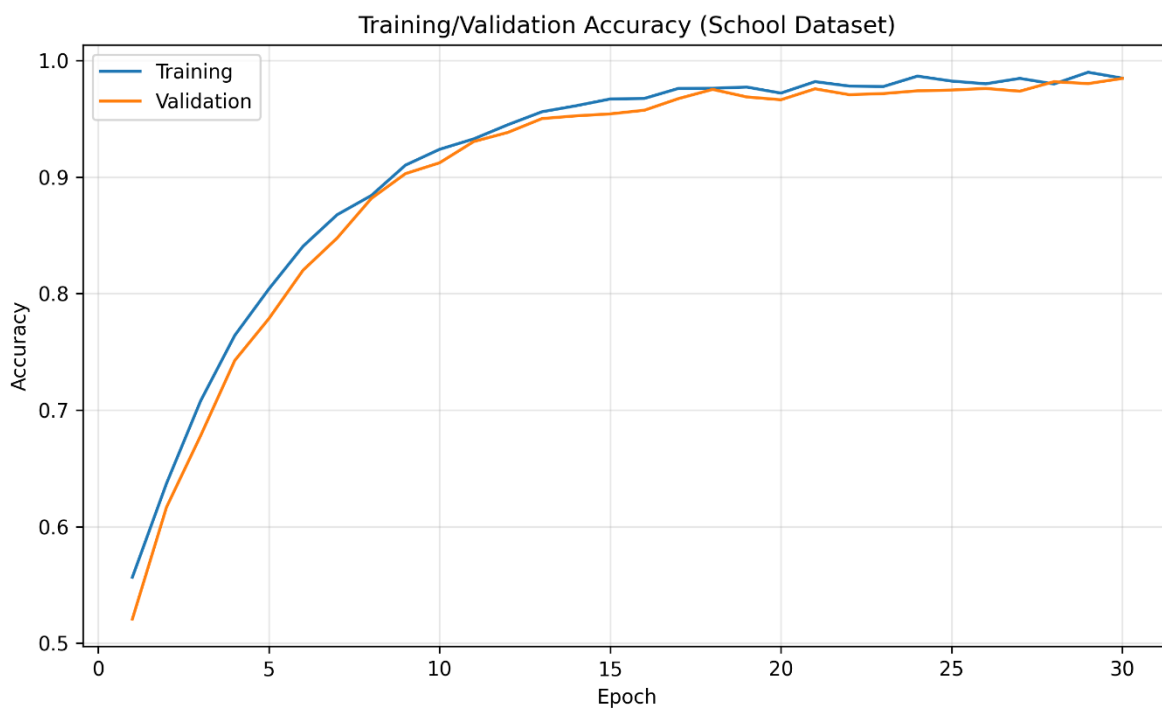


Fig 2. Training and Validation Accuracy of MARLIN-X on the School Dataset

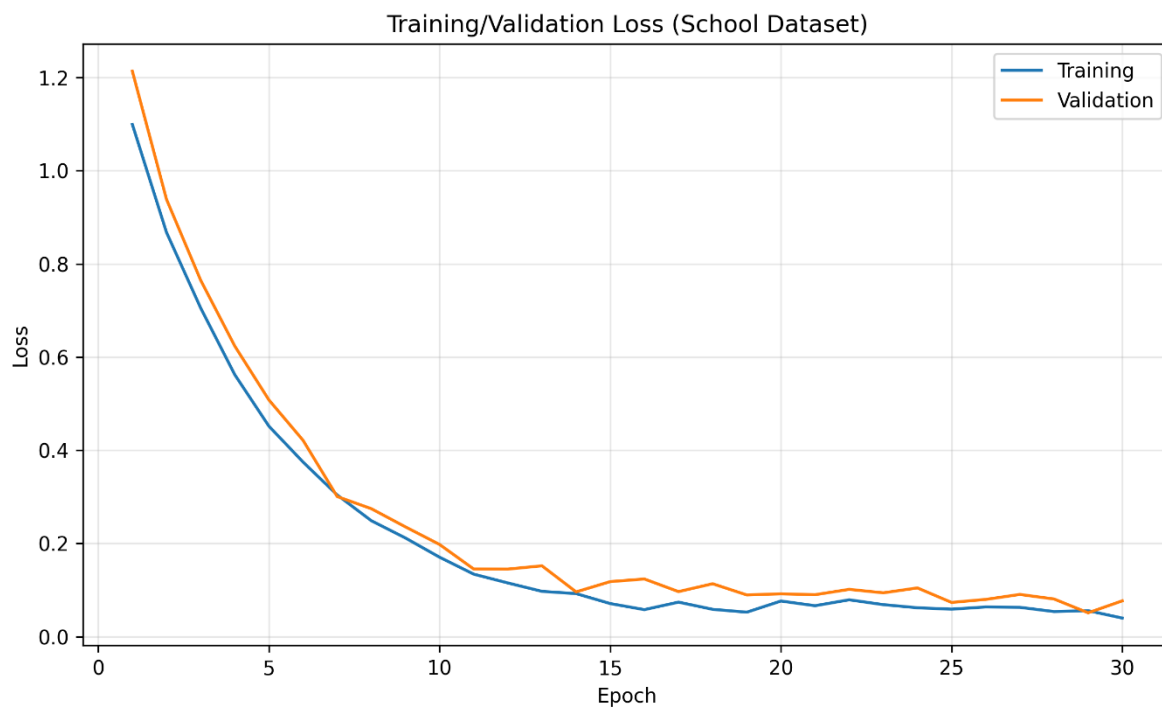


Fig 3. Training and Validation Loss of MARLIN-X on the School Dataset

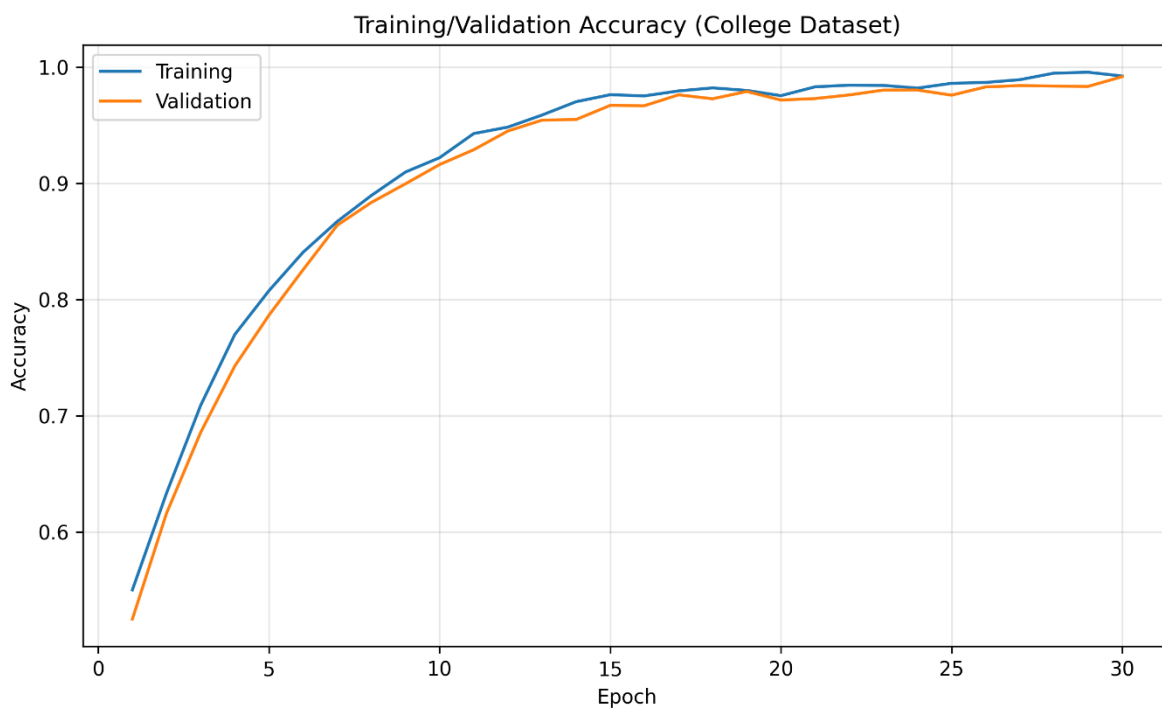


Fig 4. Training and Validation Accuracy of MARLIN-X on the College Dataset

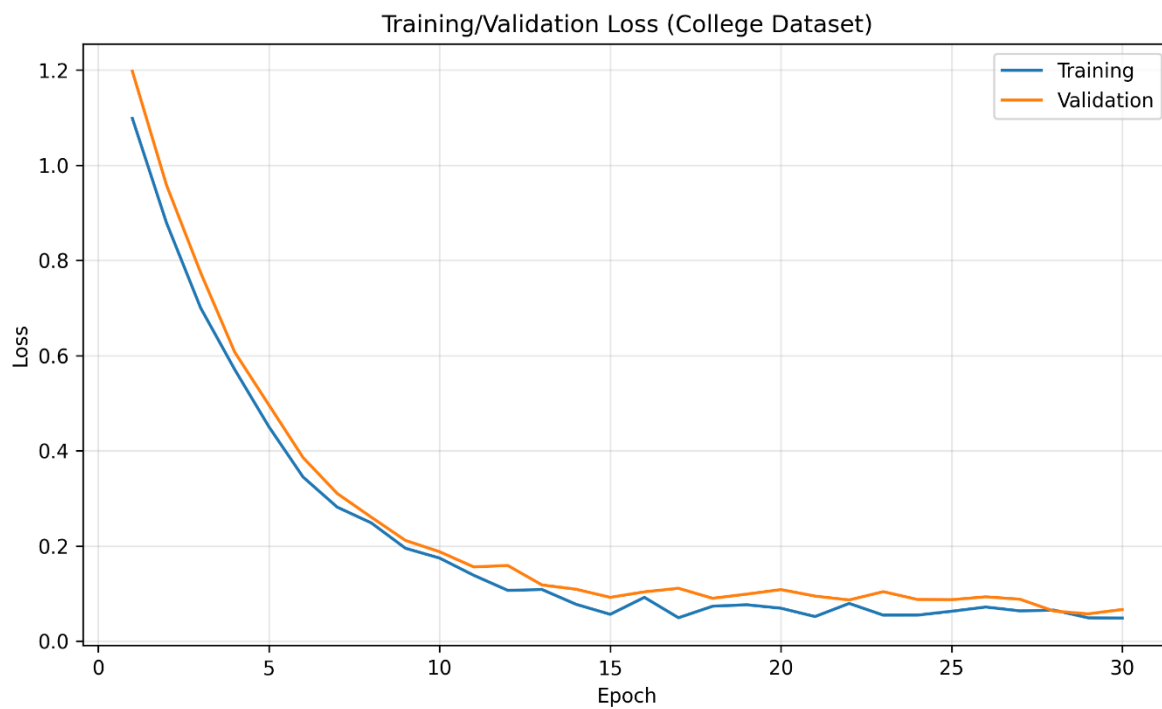


Fig 5. Training and Validation Loss of MARLIN-X on the College Dataset

models that are transformer based increase the overall accuracy but are sensitive to imbalances. The tri-label margin-based feature selection with structured hyperparameter optimization helps with the stable and balance classification results.

The hyperparameter optimization phase also leads to the increased efficiency. The objective function is a combination of predictive performance, parameter penalty and latency control. Consequently, MARLIN-X has greater accuracy using fewer parameters than regular deep models. Patch fragmentation strategy also avoids premature stagnation whereas patch migration increases the speed at which convergence is achieved on productive parts of the search space. The traditional PSO and GWO variants do not have this structured exploration mechanism.

The proposed work is novel in three aspects. To begin with, the MPMO proposes patch level migration and fragmentation as opposed to single-level migration. Second, the tri-label margin criterion directly aims at maximizing the multi-class separation in the process of feature selection. Third, rule-set distillation is incorporated in the optimization system and not used as a post-processing.

4 Conclusion

This study introduced MARLIN-X, an academic prediction of performance framework which is an optimization-based study based on smartphone behavioral predictors. The main aim was the enhancement of the predictive accuracy, better separation of the classes, lowering the dimensionality of features and offer a readable decision logic in a single optimization framework. The given framework manages to combine Tri-Label Margin Feature Selection, DCAR-Net hyperparameter optimization and Rule-Set Distillation with the help of the Mosaic Patch Migration Optimizer. The findings indicate a definite quantitative gain on the previous models. The dimensionality of the feature was decreased by 53.85% on the School dataset and by 51.85% on the College dataset. The tri-label margin went up to 1.93 and 1.90 which means that there is more separation of classes. Accuracy of classification was 97.8% and 98.3% at School and College datasets respectively. There was a strong level of Macro-F1 scores of over 97% and a balanced accuracy of 96.9% and 97.4% hence confirming that multi-class performance was stable. The small rule set with 15 rules came up with a fidelity of 96.5% to the deep model predictions without compromising predictive ability. The most important innovation of this work is the MPMO. The proposed optimizer is based at the habitat patch level as compared to the traditional swarm-based algorithms which are based on individual movement of leaders. The combination of margin-based feature selection, optimization and rule discovery in a single optimization model gives a predictive architecture.

Even though the proposed work has some limitations. The datasets are restricted to school and college groups under a particular academic setting. Live smartphone sensor records were not integrated as such. Larger population sizes and deep architectures raise the cost of computation. The rule set despite its small size is still threshold-dependent and can need re-calibration to other demographic distributions. The framework will be scaled to multi-institutional datasets and streams of real-time behavior in the future. It can be multi-objective optimization to minimize the energy consumption and the cost of deployment. Generalization of rules in various educational systems can also be developed. Combination with mental health risk prediction models can offer a larger applicability.

5 Data Availability Statement

The datasets used in this study are available from the corresponding author upon reasonable request.

6 Details of Ethical Clearance

The study was conducted following ethical guidelines approved by the affiliated College and participating School administrations. Informed consent was obtained from participants. Data were anonymized prior to analysis. No personally identifiable information was retained.

7 Acknowledgements

The authors thank, DST-FIST, Government of India for funding towards infrastructure facilities at St. Joseph's College (Autonomous), Tiruchirappalli – 620002, Affiliated to Bharathidasan University, Tamil Nadu, India.

About authors: S,Vimala is a Full-time Ph.D. Scholar and Dr. G. Arockia Sahaya Sheela serves as Assistant Professor

References

- 1) Vimala S, Sheela GA. Impact of smartphone usage on students' academic performance using contemporary deep learning models. *International Journal of Information Technology, Research and Applications*. 2025;4(4):1-8. <https://doi.org/10.59461/ijitra.v4i4.222>.

- 2) Kalita E, Alfarwan AM, Aouifi HE, Kukkar A, Hussain S, Ali T, et al. Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation. *Frontiers in Education*. 2025;10:1581247. <https://doi.org/10.3389/feduc.2025.1581247>.
- 3) Boujmiraz S, Darhmaoui H, Maliani ADE. Predicting student performance: a comprehensive review of machine learning, deep learning, and explainable AI approaches. *Computers and Education: Artificial Intelligence*. 2026;10:100548. <https://doi.org/10.1016/j.caeai.2026.100548>.
- 4) Vimala S, Sheela DG. Predictive analytics for mobile phone impact on student academic achievement: a deep learning framework for digital wellness monitoring. *International Journal of Research Publication and Reviews*. 2025;6(11):629–636. Available from: <https://ijrpr.com/uploads/V6ISSUE11/IJRPR54986.pdf>.
- 5) Ali EA, Rahman A, Kshetri N. Predicting student academic performance using deep learning: a PyTorch-based approach. *Preprints*. 2025. <https://doi.org/10.20944/preprints202507.2493.v1>.
- 6) Devaraje U, Sudharshan PKA. A multi-factor machine learning framework for predicting and profiling student academic performance using behavioral, financial, and wearable data. *MethodsX*. 2025;p. 103673. <https://doi.org/10.1016/j.mex.2025.103673>.
- 7) Ludwig SA. Guided particle swarm optimization for feature selection: application to cancer genome data. *Algorithms*. 2025;18:220. <https://doi.org/10.3390/a18040220>.
- 8) Wang Y, Yin Y, Zhao H, Liu J, Xu C, Dong W. Grey wolf optimizer with self-repulsion strategy for feature selection. *Scientific Reports*. 2025;15:12807. <https://doi.org/10.1038/s41598-025-97224-8>.
- 9) Sutikno HA, Wibawa A, Waspada I. Introducing the mutated binary whale optimization algorithm for rice leaf disease classification. *Engineering, Technology and Applied Science Research*. 2025;15(6):29362–29367. <https://doi.org/10.48084/etasr.13915>.
- 10) Isabella S, Lokeshraja B, Narayan NM, Dharshan S, Manikandan G. Harris hawks optimization based deep learning models for heart disease diagnosis. *Scientific Reports*. 2025;15:38395. <https://doi.org/10.1038/s41598-025-22326-2>.
- 11) Alnasyan B, Basher M, Alassafi M. Kanformer: an attention-enhanced deep learning model for predicting student performance in virtual learning environments. *Social Network Analysis and Mining*. 2025;15:25. <https://doi.org/10.1007/s13278-025-01446-7>.
- 12) Gong Y, Wang F, Zhang Y, Geng J. TGEL-transformer: fusing educational theories with deep learning for interpretable student performance prediction. *PLoS ONE*. 2025;20(6):e0327481. <https://doi.org/10.1371/journal.pone.0327481>.
- 13) Junejo NUR, Nawaz MW, Huang Q, Dong X, Wang C, Zheng G. Accurate multi-category student performance forecasting at early stages of online education using neural networks. *Scientific Reports*. 2025;15:16251. <https://doi.org/10.1038/s41598-025-00256-3>.
- 14) Saini BK. Optimizing student academic performance prediction using heterogeneous ensemble learning. *European Journal of Artificial Intelligence and Machine Learning*. 2025;4(4):1–6. <https://doi.org/10.24018/ejai.2025.4.4.77>.
- 15) Ahmed W, Wani MA, Plawiak P, Meshoul S, Mahmoud A, Hammad M. Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Scientific Reports*. 2025 Jul;15(1):26879. <https://doi.org/10.1038/s41598-025-12353-4>.
- 16) Rochy EA, Ferdaus J, Biswas U, Tiang JJ, Nahid AA. Metaheuristic-based feature selection and machine learning workflow for predictive modeling. *Journal of Computer Science*. 2026;22(2):540–551. <https://doi.org/10.3844/jcssp.2026.540.551>.
- 17) Alsabhan AH, Malik S, Patro SG, Mahanty C, Hadi AA, Ghouse M, et al. SailMutLoc: a sailfish-mutation enhanced optimization algorithm for student performance prediction. *Ain Shams Engineering Journal*. 2026 Jan;17(1):103898. <https://doi.org/10.1016/j.asej.2025.103898>.
- 18) Pan L, Cheng WL, Lim WH, Sharma A, Jatily V, Tiang SS, et al. A robust wrapper-based feature selection technique based on modified teaching learning based optimization with hierarchical learning scheme. *Engineering Science and Technology, an International Journal*. 2025 Jan;61:101935. <https://doi.org/10.1016/j.jestch.2024.101935>.