Using Neutrosophic Soft Set to predict Higher Education Academic Performance

Sally Afchal[1,*], Muhammad Eid Balbaa[2]
[1]College of Business Administration, American University of the Middle East, Kuwait
[2]Tashkent State University of Economics, Uzbekistan
Emails: sally.afchal@aum.edu.kw; M.balbaa@tsue.uz

Abstract

Neutrosophic Logic is a neonate research field in which every proposition is assessed to have the proportion (percentage) of truth in a sub-set T, the proportion of indeterminacy in a sub-set I, and the proportion of falsity in a sub-set F. Neutrosophic set (NS) is effectively implemented for undetermined data processing and establishes benefits for handling the indeterminacy data. In the academic industries, early performance prediction of students is significant to the academic community so strategic interference might be planned before students attain the final semester. Forecasting the performance of students has turned into a challenging task owing to the rising number of data in educational procedures. The educational data mining (EDM) models are involved in extracting a pattern to explore hidden data from educational information. Currently, Machine learning (ML) and Artificial intelligence (AI) are implemented in numerous domains generally in the field of education to evaluate and analyze several features of educational datasets gathered from many educational institutions. This study develops a Leveraging Generalized Possibility Neutrosophic Soft Set with Feature Selection for Accurate Students' Academic Performance Prediction Model (GPNSSFS-SAPPM). The intention of the proposed GPNSSFS-SAPPM system relies on improving the prediction model of students' higher education performance using metaheuristic optimization algorithms. The data pre-processing model is employed at first by applying mean normalization for converting input data into a suitable format. In addition, the golf optimization algorithm (GOA) is exploited for the feature selection process. Followed by, the classification process is done by generalized possibility neutrosophic soft set (GPNSS). At last, the parameter tuning process is performed through henry gas solubility optimization (HGSO) algorithm to improve the classification performance of the GPNSS classifier. A wide-ranging experimentation was performed to prove the performance of the GPNSSFS-SAPPM method. The experimental results specified that the GPNSSFS-SAPPM model underlined advancement over other recent techniques.

Keywords: Generalized Possibility Neutrosophic Soft Set; Neutrosophic Logic; Students' Academic Performance; Feature Selection; Henry Gas Solubility Optimization

1. Introduction

In the mathematical model, dealing with inconsistency and uncertainty is one of the most significant subjects for experts to investigate. Experts have presented several estimates for difficulties comprising inconsistency and uncertain information [1]. Several popular estimates are fuzzy-set model introduced and the intuitionistic fuzzy-set model proposed. A fuzzy set can be recognized by membership functionalities and the intuitionistic fuzzy set can be recognized by non-membership and membership functionalities [2]. However, fuzzy set and the intuitionistic fuzzy set do not control the inconsistent and indeterminant data. Hence, the neutrosophic set model was proposed as an overview of the intuitionistic fuzzy set and the fuzzy set, which depends on the Neutrosophy [3]. In the latest year, the interest has increased in identifying the significant aspects that influence the performance of the students in education, particularly with the usage of data mining techniques and methods. This research area is typically recognized as educational data mining (EDM) [4]. The motive for these interests is credited to the relevance of some studies to help in identifying low-performing students earlier for overcoming the challenges

DOI: https://doi.org/10.54216/IJNS.260302

in learning which in turn improves learning outcome of the students as a result it supports the goal of the institution by offering high-quality education system [5]. Moreover, EDM was developed rapidly as it became a crucial research area because it can mine fresh understanding from the data of the students. EDM is involved with removing a design for discovering unseen data from the education information [6].

The estimation of student's performance in higher education has been a critical task that is being investigated by employing EDM. These tasks predict the importance of an anonymous variable that defines the student's results, marks, grades, and so on. Forecasting attrition of the student, success, and failures are the key fields that were debated in the related work of these studies. Every stakeholder belonging to this area needs prior warning systems to forecast learning [7]. This initial warning system does not just reduce the learning expenses also; it reduces the space and time necessities. The major difficulty is improving the quality of the educational process and enhancing the performance of the student. Teachers may upgrade their teaching method by fulfilling the necessity of poor-performing students and further guiding worthy students [8]. These forecast outcomes might assist learners in developing a proper realization of in what way they are performing and taking measures correspondingly. Increasing the retention of the students has been a long-term goal for educational institutions in the world [9]. In current, famous procedures like machine learning (ML) and deep learning (DL) are utilized for estimation. The ML and DL-enabled methods can discover unidentified designs precisely [10].

This study develops a GPNSSFS-SAPPM system relies on improving the prediction model of students' higher education performance using metaheuristic optimization algorithms. The processes of our paper are given below:

- The data pre-processing model applies mean normalization model.
- The golf optimization algorithm (GOA) is exploited for feature selection process.
- The classification process is done by generalized possibility neutrosophic soft set (GPNSS).
- The parameter tuning process is performed through henry gas solubility optimization (HGSO) algorithm

2. Related Works

Wang and Yu [11] proposed an ML technique to predict the performance of the student on the basis of online learning. The crucial thinking is that eleven learning behavior pointers was made as per the process of online learning. By examining the connection among these eleven learning behavior pointers as well as the score attained by learners in online learning, this study filters learning behavior pointers which is dimly connected with scores of the student, in the meantime, retaining those learning behavior pointers are connected strongly with the learners score, which is utilized as the characteristic root pointers. Chami and Cockburn [12] explored the opportunities and challenges proposed by AI in Higher education, concentrating on their ability to customize learning, improve accessibility and enhance efficiency. Issues associated to the development of critical thinking skills, preservation of teacher-student relationships, bias and data privacy are well examined. Certainly, AI has been a powerful tool to enrich and support education however; it cannot substitute quality of the human contribution. It stays significant for emphasizing teacher training, ensuring quality education and ethical considerations. Ryabko et al. [13] presented a new procedure to evaluate educational programs quality and performance of the institution in higher education employing progressive AI systems, particularly the multi-layer neural networks and Adaptive Neuro-Fuzzy Inference System (ANFIS). The objective of this research is addressing the difficulties of bias in self-assessment process and pro-active identification of possible problems and lack of educational events before authorization analyses. The introduced methodology used ratings of the student in four level valuation scale as input information to the multi-layer neural system; however, the standards to assess quality of the educational program worked as an input variable to the ANFIS method.

Parkavi et al. [14] uses Fuzzy Logic in the initial phase, ML technique and Factor Analysis (FA) in the next phase for discovering crucial aspects related to the efficient usage of Active Learning Strategies (ALS) in Learning Management System (LMS) of information technology (IT) class students. Neural network topology and Fuzzy logic has been joined utilizing ANFIS, an adaptive system. Kukkar et al. [15] introduced fresh methods to predict student's results in some course. This system utilizes demographic, emotional, VLE, sequence, and academic student data. Classical estimation models frequently struggles for capturing the sequential dynamic characteristics of data namely study habits, evolving performance patterns and learning trajectories. In return, the study influences LSTM networks and Recurrent Neural Networks (RNNs) specially intended for modelling arrangements and long term dependence from OULAD and self-produced emotional datasets. By integrating those designs, the introduced methods excel to capture the complex relationship among several aspects in due time.

Abuzinadah et al. [16] proposed an ML based structure with deep convoluted characteristics to predict student academic performance. The presented structure is used for predicting the academic performance of the student employing both balanced and imbalanced dataset utilizing the synthetic minority oversampling technique (SMOTE). In addition, the performance was estimated through the unique and deep convoluted characteristics.

DOI: https://doi.org/10.54216/IJNS.260302

Maulana et al, [17] proposed ML and AI applications to predict performance of the students in high school to universities. Identifying the essential part of theoretical willingness, this paper highlights the necessity for personalized intrusions for enhancing the success of the student.

3. Methodology

In this manuscript, we have proposed a GPNSSFS-SAPPM model. The primary aim of proposed GPNSSFS-SAPPM system relies on improving prediction model of student's higher education performance using metaheuristic optimization algorithms. It comprises 4 major phases namely data pre-processing, feature selection using GOA, GPNSS-based classification process, and parameter tuning using HGSO

A. Data Pre-processing

Primarily, the data pre-processing model is employed by applying mean normalization for converting an input data into a suitable format [18]. Mean normalization is a data pre-processing model employed to forecast academic performance of students by scaling aspects to have a zero mean. It aids in managing changing data ranges by converting values depending on the range and mean of the dataset. This guarantees that no single aspect surpasses the prediction model, enhancing the performance of ML methodologies. By creating data, more similar, mean normalization improves the precision of academic performance prediction techniques. It is mainly beneficial in regression-based and neural network approaches, assisting stability and better convergence. This methodology helps in recognizing key academic indicators and increasing decision-making in educational analytics. Fig. 1 denotes the workings of GPNSSFS-SAPPM technique.

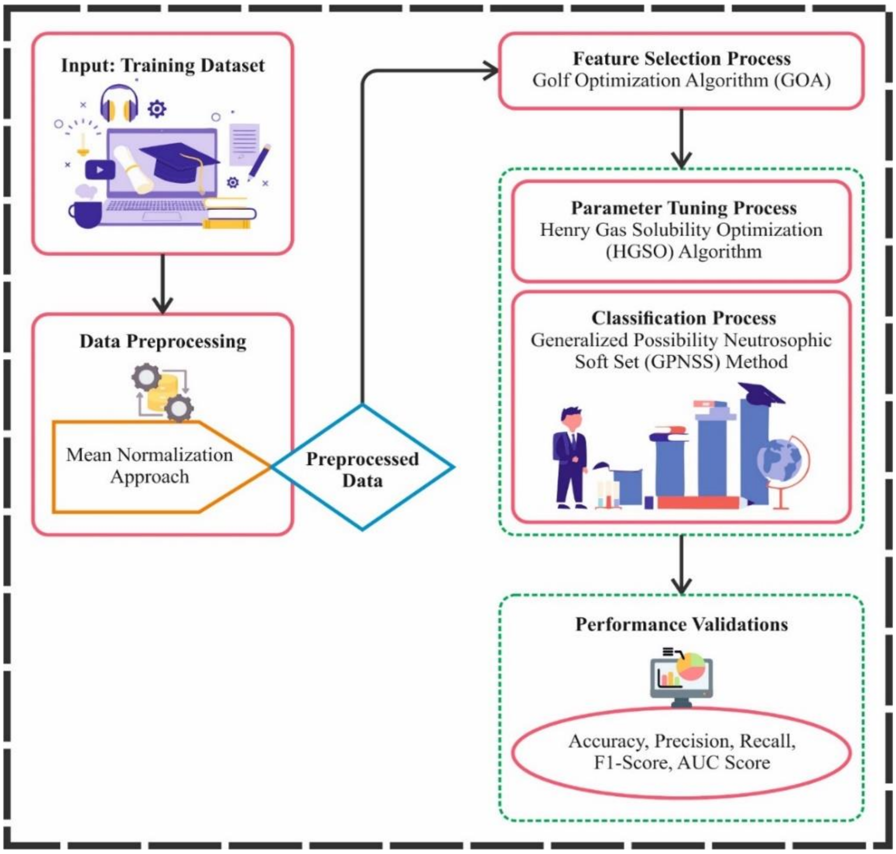

Figure 1. Workings of GPNSSFS-SAPPM technique

B. GOA-based Feature selection

Furthermore, GOA is exploited for the feature selection process. GOA is a game-based metaheuristic model, which operates in dual phases such as exploitation and exploration, with the novelty of the golf sport with the observation of player performance and considered dynamics [19]. It works by imitating the rules and behavior of players in the golf game. In golf, the swing motion consists of a mixture of controlled actions targeted for striking the ball precisely near the goal. In the same way, in GOA, the optimization procedure includes iteratively fine-tuning the solution candidates for converging near the best solution. GOA sets a population of possible solutions

DOI: https://doi.org/10.54216/IJNS.260302

inside the solution area and then, the fitness function (FF) is expressed by assessing all solution candidates using the estimation of results concerning the optimization aim. Golfers consistently use their swing approach based on consideration of the response from previous tries to increase their performance. Similarly, in the GOA, solutions are developed iteratively over the incorporation of exploitation and exploration to converge toward the best solution.

Exploration: The players are required to make their initial swing to the target named a hole that can be exemplified by the exploration phase, whereas GOA members search dissimilar regions of the solution area. GOA members are upgraded by calculating novel locations based on the expression of the powerful shot by the player toward the ball. The preceding position of the GOA member is upgraded whereas a new location is enhanced by the value of the objective function. It was mathematically obtained as demonstrated.

$$G_a^{B1}: g_{a,m}^{B1} = g_{a,m} + h \times \left(BM_m - Q \times g_{a,m}\right) \quad (1)$$

$$G_a = \begin{cases} G_a^{B1}, & F_a^{B1} < F_a \\ G_a, & else \end{cases} \quad (2)$$

Here, a randomly generated number within the range [0-1] was represented as h, the top member of GOA, and its m^{th} dimensions are consistently stated as BM, and BM_m. Now, a randomly selected random number from the set $\{1, 2\}$ was represented as Q. The parameter Q in the updated equation defines whether the shot approaches the hole $(Q = 1)$ or explores many places of the solution area $(Q = 2)$, improving the model's global search ability. The updated location of GOA members follows the exploration stage and is represented as G_a^{B1}, its m^{th} dimension is formulated as $g_{a,m}^{B1}$, and their value of the objective function is exposed as F_a^{B1}.

Exploitation: In golf, the region near the hole is named the green, where players perform accurate shots. Therefore, these controlled shots are carried out with the minimal force to ensure the existence of the ball inside the green area near the hole. It is implemented with the exploration of each GOA location of the member, establishing the capacity of skill for exploiting local searching regions. The succeeding equations are applied to update the exploitation phase.

$$G_a^{B2}: g_{a,m}^{B2} = g_{a,m} + (1 - 2h) \times \frac{l_a + h \times (u_a - l_a)}{t} \quad (3)$$

$$G_a = \begin{cases} G_a^{B2}, & F_a^{B2} < F_a \\ G_a, & else \end{cases} \quad (4)$$

Now, novel locations are calculated for all members on reviewing delicate shots near the ball. The iteration counter is gained as t, and the upgraded location of GOA member follows the exploitation stage and is represented as G_a^{B2} whereas its m^{th} dimension is designed as $g_{a,m}^{B2}$ and their value of the objective function is presented as F_a^{B2}, and the lower and upper limits of the m^{th} variable is correspondingly expressed as u_a and l_a. At last, the best solution is developed and designated for getting best solutions. The fitness function (FF) employed in the GOA method is intended to hold a balance among the various selected aspects in every solution (minimum) and the classification precision (maximum) attained by employing these chosen features, Eq. (5) depicts the FF to assess solutions.

$$Fitness = \alpha\gamma_R(D) + \beta \frac{|R|}{|C|} \quad (5)$$

Here $\gamma_R(D)$ is the classification rate of error in specified classifier. $|R|$ signifying the cardinality of the chosen subset and $|C|$ denotes the overall number of features in the database, α and β are dual parameters equivalent to the significance of classification quality and length of subset. $\in [1,0]$ and $\beta = 1 - \alpha$.

C. GPNSS-based Classification Process

In addition, the classification process is performed by GPNSS. This section presents a novel kind of set named Generalized Possibility Neutrosophic Soft Set (GPNSS) and examines a Generalized Possibility Neutrosophic Soft Decision Making Method (GPNSDMM) [20].

DEFINITION2.1

Assume that U is an initial universe, E represents a parameter set, $N(U)$ indicates the collection of all neutrosophic sets of $F(U)$ and U is collection of each fuzzy subset U. GPNS-set $P_{[\alpha,\beta]}$ through U specifies a set of triples described as

$$P_{[\alpha,\beta]} = \left\{\left(e_k, \left\{\left(\frac{u_j}{P(e_k)(u_j)}, \alpha(e_k)(u_j)\right)\right\}, \beta\left(P(e_k)\right)\right) : e_k \in E\right\}$$

DOI: https://doi.org/10.54216/IJNS.260302

or a mapping expressed by $P_{[\alpha,\beta]}: E \to N(U) \times F(U) \times N(U)$, here, $i \in \wedge_1$ and $k \in \wedge_2$, P denotes a mapping given by $P: E \to N(U)$ and $\alpha(e_k)$ represents a fuzzy set so that $\alpha: E \to F(U)$.

For every parameter

$$e_k \in E, P(e_k) = \{\langle u_j, T_{P(e_k)}(u_j), I_{P(e_k)}(u_j), F_{P(e_k)}(u_j)\rangle u_j \in U\}$$

signifies value of neutrosophic set of parameter e_k and now I, F, T, are the falsity, truth, and indeterminacy values correspondingly the element $u_i \in U$. For every $u_i \in U$ and $e_k \in E$, $0 \le T_{P(e_k)}(u_j) + I_{P(e_k)}(u_j) + F_{P(e_k)}(u_j) \le 3$. Likewise, α degrees of priority specified for the belongingness of elements of U in $P(e_k)$ and $\beta(e_k)$ the likelihood degree and weightage specified to the parameters by the specialists.

$$P_{[\alpha,\beta]}(e_k) = \left\{\left(\frac{u_1}{P(e_k)(u_1)}, \alpha(e_k)(u_1)\right)\left(\frac{u_2}{P(e_k)(u_2)}, \alpha(e_k)(u_2)\right), \dots, \left(\frac{u_n}{P(e_k)(u_n)}, \alpha(e_k)(u_n)\right), \beta(e_k)\right\}.$$

EXAMPLE2.2

Assume that $U = \{u_1, u_2, u_3\}$ be a set of 3 restaurants and $E = \{e_1, e_2, e_3\}$ is a set of qualities here $e_1 =$Taste, $e_2 =$Variety, $e_3 =$Service and let $\alpha: E \to F(U)$ and $\beta: E \to N(U)$. It an be expressed as a function $P_{[\alpha,\beta]}: E \to N(U) \times F(U) \times N(U)$

$$P_{[\alpha,\beta]} = \left\{\begin{array}{l} P_{[\alpha,\beta]}(e_1) = \left\{\left[\left(\frac{u_1}{(0.4, 0.1, 0.S)}, 0.7\right), \left(\frac{u_2}{(0.6, 0.2, 0.4)}, 0.3\right)\left(\frac{u_3}{(0.3, 0.4, 0.7)}, 0.6\right)\right], (0.3,0.5,0.7)\right\} \\ P_{[\alpha,\beta]}(e_2) = \{\left[\left(\frac{u_1}{(0.7, 0.3, 0.4)}, 0.5\right), \left(\frac{u_2}{(0.4, 0.6, 0.1)}, 0.7\right), \left(\frac{u_3}{(0.6, 0.2, 0.8)}, 0.3\right)\right], (0.6,0.8,0.2)\} \\ P_{[\alpha,\beta]}(e_3) = \{[\left(\frac{u_1}{(0.S, 0.6, 0.7)}, 0.2\right), \left(\frac{u_2}{(0.4, 0.2, 0.6)}, 0.5), \left(\frac{u_3}{(0.S, 0.4, 0.3)}, 0.4\right)\right], (0.5,0.3,0.8)\} \end{array}\right\}$$

For the objective of storing a GPNS in a computer, employ matrix notation of generalized likelihood soft set of neutrosophic $P_{[\alpha,\beta]}$. For instance, matx notation of GPNS $P_{[\alpha,\beta]}$ might be specified as: for $m, n \in \wedge$,

$$P_{[\alpha,\beta]} = \begin{pmatrix} (\langle 0.4,0.1,0.5\rangle, 0.7)(\langle 0.6,0.2,0.4\rangle, 0.3)(\langle 0.3,0.4,0.7\rangle, 0.6)(0.3,0.5,0.7) \\ (\langle 0.7,0.3,0.4\rangle, 0.5)(\langle 0.4,0.6,0.1\rangle, 0.7)(\langle 0.6,0.2,0.8\rangle, 0.3)(0.6,0.8,0.2) \\ (\langle 0.5,0.5,0.4\rangle, 0.1)(\langle 0.4,0.2,0.6\rangle, 0.5)(\langle 0.5,0.4,0.3\rangle, 0.4)(0.5,0.3,0.8) \end{pmatrix}$$

Now the m-th row vector exhibits $P(e_m)$ and n-th column vector displays u_n

DEFINITION2.3

Assume that $P_{[\alpha,\beta]}$, $Q_{[\gamma,\delta]} \in GPN(U,E)$. Subsequently, $P_{[\alpha,\beta]}$ denotes to a generalized probability GPNS- subset of $Q_{[\gamma,\delta]}$, and signified by $P_{[\alpha,\beta]} \subseteq Q_{[\gamma,\delta]}$ if

1. $\alpha(e)$ and $\beta(e)$ are a fuzzy sub-set of $\gamma(e)$ and $\delta(e)$, for each $e \in E$
2. P represents a neutrosophic sub-set of Q.

DEFINITION2.5

Assume that $P_{[\alpha,\beta]}, Q_{[\gamma,\delta]} \in GPN(U,E)$. Subsequently, $P_{[\alpha,\beta]}$ and $Q_{[\gamma,\delta]}$ are named generalized possibility neutrosophic soft equivalent-set and indicated $by P_{[\alpha,\beta]} = Q_{[\gamma,\delta]}$ if $P_{[\alpha,\beta]} \subseteq Q_{[\gamma,\delta]}$ and $P_{[\alpha,\beta]} \supseteq Q_{[\gamma,\delta]}$.

DEFINITION2.6

Assume that $P_{[\alpha,\beta]} \in GPN(U,E)$. Afterwards, $P_{[\alpha,\beta]}$ represents generalized possibility neutrosophic soft null-set, specified as $\phi_{[\alpha,\beta]}$, if $\forall e \in$, $\phi_{[\alpha,\beta]}: E \to N(U) \times F(U) \times N(U)$ so that $\phi_{[\alpha,\beta]}(e) = \left\{\left[\frac{u}{\phi(e)(u)}, \alpha(e)(u): u \in U\right] \beta\big(P(e)\big)\right\}$ now $\phi(e) = \{\langle u, 0,0,1\rangle: u \in U\}$ and $\alpha(e) = \{(u, 0): u \in U\}$ and $\beta(e) = \{(u, 0,0,1): u \in U\}$.

DEFINITION2.7

Assume that $P_{[\alpha,\beta]} \in GPN(U,E)$. Moreover, $P_{[\alpha,\beta]}$ denotes generalized possibility neutrosophic soft universal-set, designated as $U_{[\alpha,\beta]}$, if $\forall e \in E$, $U_{[\alpha,\beta]}: E \to N(U) \times F(U) \times N(U)$ so that $U_{[\alpha,\beta]}(e) = \left\{\left[\frac{u}{U(e)(u)}, \alpha(e)(u): u \in U\right], \beta(e)\right\}$ now $U(e) = \{\langle u, 0,0,1\rangle: u \in U\}$ and $\alpha(e) = \{(u, 1): u \in U\}$ and $\beta(e) = \{(u, 1,1,0): u \in U\}$.

PROPOSITION2.8

DOI: https://doi.org/10.54216/IJNS.260302

Assume $P_{[\alpha,\beta]}$, $Q_{[\gamma,\delta]}$ and $H_{[\theta,\varphi]} \in GPN(U,E)$. Formerly,

1. $\phi_{[\alpha,\beta]} \subseteq P_{[\alpha,\beta]}$
2. $P_{[\alpha,\beta]} \subseteq U_{[\alpha,\beta]}$
3. $P_{[\alpha,\beta]} \subseteq \phi_{[\alpha,\beta]}$ and $\phi_{[\alpha,\beta]} \subseteq H_{[\theta,\varphi]}$ implies $P_{[\alpha,\beta]} \subseteq H_{[\theta,\varphi]}$

Proof. The proof arises from the Definitions $(2.5) - (2.7)$

DEFINITION2.9

Consider $P_{[\alpha,\beta]} \in GPN(U,E)$, Now

$$P_{[\alpha,\beta]}(e_k) = \{([P(e_k)(u_i), \alpha(e_k)(u_i)], \beta(P(e_k))): e_k \in E_t u_i \in U\}$$

$$P(e_k) = \{\langle u_t T_{P(e_k)}(u_i), I_{P(e_k)}(u_i), F_{P(e_k)}(u_i)\rangle\} \forall e_k \in E_t u \in U.$$

Thus $e_k \in E$ and $u_i \in U$, and

1. $P^T_{[\alpha,\beta]}$ specifies truth-membership part of $P_{[\alpha,\beta]}$

$P^T_{[\alpha,\beta]} = \left\{\left(P^T_{kj}(e_k), \alpha_{kj}(e_k)\right)\right\}$ and $P^T_{kj}(e_k) = \left\{\left(u_j, T_{P(e_k)}(u_j)\right)\right\}, \alpha_{kj}(e_k) = \left\{\left(u_j, \alpha(e_k)(u_j)\right)\right\}$

2. $P^I_{[\alpha,\beta]}$ indicates indeterminacy-membership part of $P_{[\alpha,\beta]}$

$P^I_{[\alpha,\beta]} = \left\{\left(P^I_{kj}(e_k), \alpha_{kj}(e_k)\right)\right\}$ and $P^I_{kj}(e_k) = \left\{\left(u_j, I_{P(e_k)}(u_j)\right)\right\}, \alpha_{kj}(e_k) = \left\{\left(u_j, \alpha(e_k)(u_j)\right)\right\}$

3. $P^F_{[\alpha,\beta]}$ depicts falsity-membership part of $P_{[\alpha,\beta]}$

$P^F_{[\alpha,\beta]} = \left\{\left(P^F_{kj}(e_k), \alpha_{kj}(e_k)\right)\right\}$ and $P^F_{kj}(e_k) = \left\{\left(u_j, F_{P(e_k)}(u_j)\right)\right\}, \alpha_{kj}(e_k) = \left\{\left(u_j, \alpha(e_k)(u_j)\right)\right\}$

It specifies GPNS in the form of $P_{[\alpha,\beta]} = \left(P^T_{[\alpha,\beta]t} P^I_{[\alpha,\beta]t} P^F_{[\alpha,\beta]}\right)$.

A GPNS might be stated in form of matrix.

Assume that GPNSS $P_{[\alpha,\beta]}$ given in instance 2.4. Subsequently GPNSS $P_{[\alpha,\beta]}$ might be shown.

matrix form is given:

$$P^T_{[\alpha,\beta]} = \begin{pmatrix} (0.4,0.8)(0.8,0.3)(0.4,0.7)(0.3) \\ (0.7,0.6)(0.5,0.7)(0.7,0.5)(0.6) \\ (0.5,0.2)(0.6,0.5)(0.6,0.4)(0.5) \end{pmatrix}$$

$$P^I_{[\alpha,\beta]} = \begin{pmatrix} (0.7,0.8)(0.6,0.3)(0.5,0.7)(0.5) \\ (0.4,0.6)(0.7,0.7)(0.4,0.5)(0.8) \\ (0.6,0.2)(0.5,0.5)(0.5,0.4)(0.3) \end{pmatrix}$$

$$P^F_{[\alpha,\beta]} = \begin{pmatrix} (0.6,0.8)(0.8,0.3)(0.8,0.7)(0.7) \\ (0.6,0.6)(0.5,0.7)(0.8,0.5)(0.2) \\ (0.7,0.2)(0.8,0.5)(0.7,0.4)(0.8) \end{pmatrix}$$

DEFINITION2.10

Assume $P_{[\alpha,\beta]}$, $Q_{[\gamma,\delta]} \in GPN(U,E)$. The union of dual GPNSs $P_{[\alpha,\beta]}$ and $Q_{[\gamma,\delta]}$ over U, represented as $P_{[\alpha,\beta]} \cup Q_{[\gamma,\delta]}$.

$$P_{[\alpha,\beta]} \cup Q_{[\gamma,\delta]} = \left\{\left(e_k, \left\{\left(\rho_t \alpha_{kj}(e_k) \oplus \gamma_{kj}(e_k)\right)\right\}, \beta_{kj}\left(P(e_k)\right) \oplus \delta_{kj}(e_k)\right): e_k \in E\right\}$$

Now

$$\rho = \frac{u_j}{\left(P^T_{kj}(e_k) \oplus Q^T_{kj}(e_k), P^I_{kj}(e_k) \oplus Q^I_{kj}(e_k), P^F_{kj}(e_k) \otimes Q^F_{kj}(e_k)\right)}$$

and $\oplus$ denotes n-conorm and $\otimes$ is n-norm functions correspondingly.

DOI: https://doi.org/10.54216/IJNS.260302

D. HGSO-based Parameter Tuning

Finally, the parameter tuning process is performed through HGSO algorithm for improving the classification performance of the GPNSS classifier [21]. The HGSO technique pretends Henry's gas law behaviour, which relates to a quantity of an assumed gas that is melted to a given category and amount of liquid at a fixed temperature, to resolve challenging optimizer issues. The HGSO technique imitates the gas huddling behaviour to balance exploration and exploitation in the searching space and to evade local goals. The locations of gases are set as per Eq. (3).

$$X_i(t+1) = X_{min} + r(X_{max} - X_{min}) \tag{6}$$

Here, X_i denotes a location of ith gas of size N; r denotes a randomly generated amount among 0 and 1; X_{min} and X_{max} refers to lower and upper limits, correspondingly; and t denotes an iteration time. L_1, L_2, and L_3 denotes constants; j means a cluster number; T refers to a temperature; $P_{i,j}$ indicates the partial pressure of ith gas in jth cluster; T^{θ} and K are constants; $iter$ represents a maximum amount of iterations; γ refers to a measure of capability of gases in a group to interrelate with each other; α indicates a measure of effect of other gases; β refers to a constant; $F_{i,j}$ specifies the fitness of ith gas in jth cluster; and F is the fitness of best gas; $G(i,j)$ means a location of gas i in cluster j; r signifies a randomly generated amount; G_{min} and G_{max} denotes the limits of problem. The fitness selection is the substantial feature inducing the execution of the HGSO method. The process of hyper-parameter selection comprises the solution encoding method to assess the effectiveness of candidate solutions. In this novel, the HGSO technique deliberates precision as the main criterion to intend the FF can be expressed

$$Fitness = \max(P)$$

$$P = \frac{TP}{TP+FP} \tag{7}$$

From the expression, TP specifies true positive and FP is false positive value.

4. Result and Experimental Analysis

The experimental analysis of GPNSSFS-SAPPM model is verified under Students Performance dataset [22]. This dataset holds 2392 samples under 4 grades such as 0 of 'A' (GPA >= 3.5), 1 of 'B' (3.0 <= GPA < 3.5), 2 of 'C' (2.5 <= GPA < 3.0), 3 of 'D' (2.0 <= GPA < 2.5), and 4 of 'F' (GPA < 2.0). It has 14 attributes in total but only 10 attributes were chosen.

Table 1 exemplifies the students' performance prediction result of GPNSSFS-SAPPM model based on 80%TRAPHA and 20%TESPHA. The performances indicate that the GPNSSFS-SAPPM model accurately recognized the samples. On 80%TRAPHA, the GPNSSFS-SAPPM approach obtains average $accu_y$ of 98.81%, $prec_n$ of 96.12%, $reca_l$ of 95.39%, $F1_{score}$ of 95.75%, and AUC_{score} of 97.27%. Besides, on 20%TESPHA, the GPNSSFS-SAPPM method achieves average $accu_y$ of 99.25%, $prec_n$ of 97.14%, $reca_l$ of 97.58%, $F1_{score}$ of 97.34%, and AUC_{score} of 98.52%.

Table 1: Student's performance prediction outcome of GPNSSFS-SAPPM model under 80%TRAPHA and 20%TESPHA

Class Labels	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$	AUC_{score}
TRAPHA (80%)					
0	99.27	94.12	89.89	91.95	94.81
1	99.22	96.67	96.21	96.44	97.90
2	98.27	94.62	94.92	94.77	96.93
3	99.06	97.04	97.62	97.33	98.49
4	98.22	98.13	98.34	98.23	98.22
Average	**98.81**	**96.12**	**95.39**	**95.75**	**97.27**

DOI: https://doi.org/10.54216/IJNS.260302

TESPHA (20%)					
0	99.79	94.74	100.00	97.30	99.89
1	98.75	94.83	94.83	94.83	97.06
2	98.96	97.33	96.05	96.69	97.78
3	99.58	100.00	97.44	98.70	98.72
4	99.16	98.80	99.60	99.20	99.15
Average	**99.25**	**97.14**	**97.58**	**97.34**	**98.52**

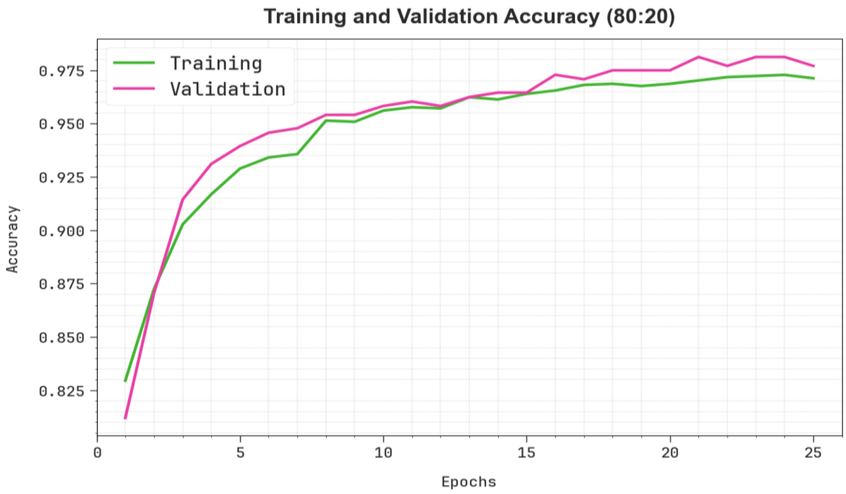

Figure 2. $Accu_y$ Curve of GPNSSFS-SAPPM model under 80:20

In Fig. 2, the training (TRAN) $accu_y$ and validation (VALN) $accu_y$ performances of the GPNSSFS-SAPPM approach based on 80:20 are showcased. The values of $accu_y$ are computed across an interlude of 0-25 epochs. The figure underscored that the both $accu_y$ values express a cumulative propensity, indicating the proficiency of the GPNSSFS-SAPPM algorithm with enhanced performance through multiple repetitions. In addition, the both $accu_y$ ruins nearer across the epochs, which notified diminish overfitting and presents superior outcomes of the GPNSSFS-SAPPM system, assuring consistent prediction on unseen samples.

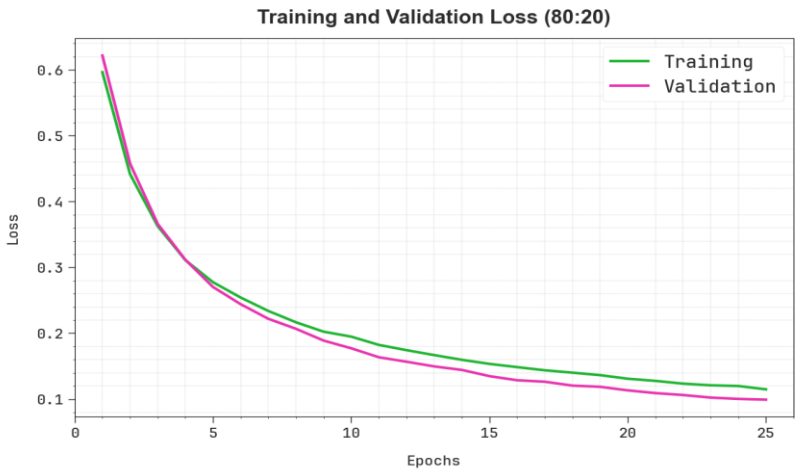

Figure 3. Loss curve of GPNSSFS-SAPPM technique under 80:20

DOI: https://doi.org/10.54216/IJNS.260302

In Fig. 3, the TRAN loss (TRANLOS) and VALN loss (VALNLOS) graph of the GPNSSFS-SAPPM technique based on 80:20 is shown. The loss values are computed through an interlude of 0-25 epochs. It is exemplified that both values represent a lessening propensity, which indicates the proficiency of the GPNSSFS-SAPPM approach in corresponding to a tradeoff between data fitting and generalization. The successive reduction in values of loss as well as securities the maximum outcome of the GPNSSFS-SAPPM model and tunes the prediction results gradually.

Table 2 displays the students' performance prediction analysis of GPNSSFS-SAPPM model based on 70%TRAPHA and 30%TESPHA. With 70%TRAPHA, the GPNSSFS-SAPPM model reaches average $accu_y$ of 98.83%, $prec_n$ of 95.78%, $reca_l$ of 96.16%, $F1_{score}$ of 95.96%, and AUC_{score} of 97.69%. Furthermore, with 30%TESPHA, the GPNSSFS-SAPPM algorithm obtains average $accu_y$ of 98.72%, $prec_n$ of 94.87%, $reca_l$ of 93.65%, $F1_{score}$ of 94.21%, and AUC_{score} of 96.42%.

Table 2: Students' performance prediction outcome of GPNSSFS-SAPPM model under 70%TRAPHA and 30%TESPHA

Class Labels	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$	AUC_{score}
TRAPHA (70%)					
0	99.34	92.59	93.75	93.17	96.69
1	99.16	97.31	95.26	96.28	97.46
2	98.69	95.05	97.11	96.07	98.05
3	98.63	95.14	96.82	95.97	97.91
4	98.33	98.80	97.87	98.33	98.33
Average	**98.83**	**95.78**	**96.16**	**95.96**	**97.69**
TESPHA (30%)					
0	99.16	92.00	85.19	88.46	92.45
1	98.05	93.33	88.61	90.91	93.91
2	98.89	95.69	97.37	96.52	98.27
3	98.61	94.16	98.47	96.27	98.56
4	98.89	99.18	98.64	98.91	98.89
Average	**98.72**	**94.87**	**93.65**	**94.21**	**96.42**

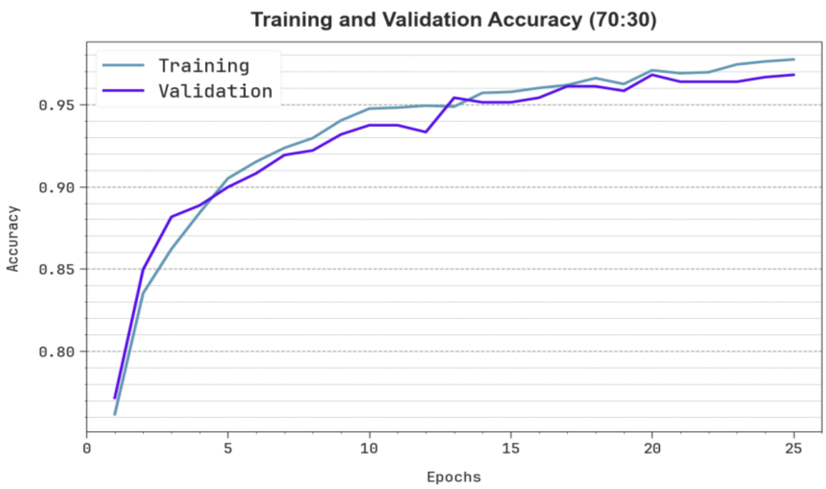

Figure 4. $Accu_y$ Curve of GPNSSFS-SAPPM model under 70:30

DOI: https://doi.org/10.54216/IJNS.260302

In Fig. 4, the training (TRAN) $accu_y$ and validation (VALN) $accu_y$ performances of the GPNSSFS-SAPPM method based on 70:30 is depicted. The $accu_y$ values are computed through a period of 0-25 epochs. The figure underscored that the both $accu_y$ values express a growing propensity, indicating the proficiency of the GPNSSFS-SAPPM system by maximum outcome through numerous repetitions. Moreover, the both $accu_y$ ruins nearer through the epochs, indicating decreased overfitting and expressing superior outcomes of the GPNSSFS-SAPPM approach, ensuring steady prediction on unseen samples.

In Fig. 5, the TRAN loss (TRANLOS) and VALN loss (VALNLOS) graph of the GPNSSFS-SAPPM method based on 70:30 is showcased. The loss values are computed through an interlude of 0-25 epochs. It is exemplified that the both values represent a lessening propensity, indicating the proficiency of the GPNSSFS-SAPPM approach in equalizing an equilibrium between data fitting and generalization. The progressive decrease in values of loss also assurances the higher performance of the GPNSSFS-SAPPM system and tunes the prediction results after a while.

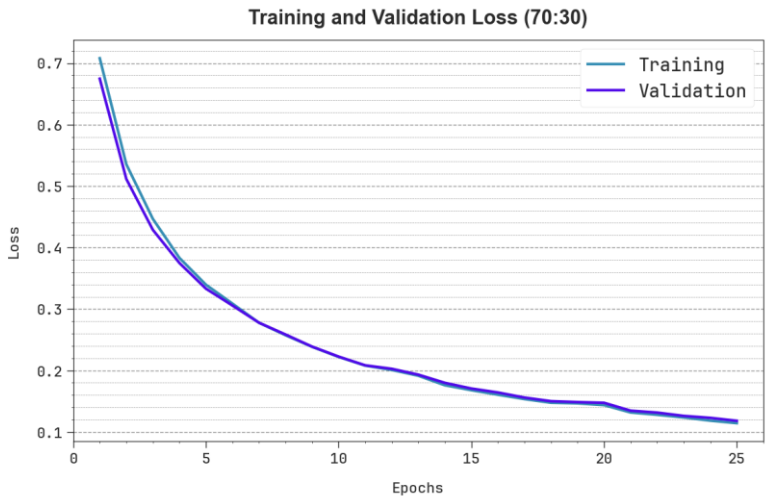

Figure 5. Loss curve of GPNSSFS-SAPPM technique under 70:30

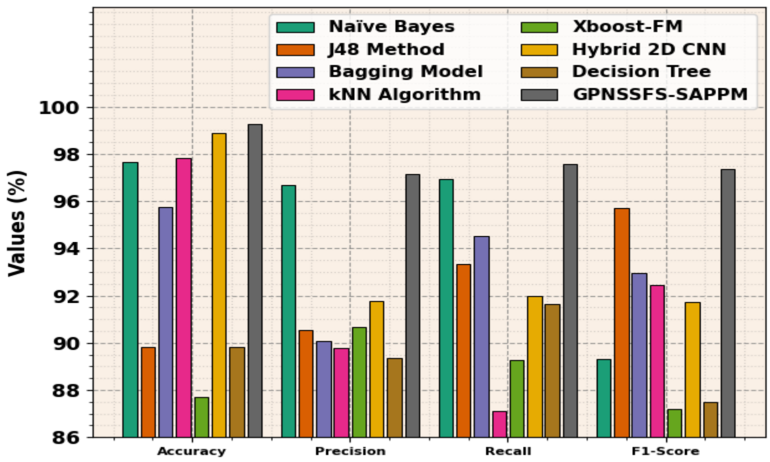

Figure 6. Comparative study of GPNSSFS-SAPPM model with existing approaches

Fig. 6 signifies the comparative study of GPNSSFS-SAPPM technique with existing methodologies [23-25]. The performance implies that the proposed GPNSSFS-SAPPM system has gained efficient performance with $accu_y$ of 99.25%, $prec_n$ of 97.14%, $reca_l$ of 97.58%, and $F1_{score}$ of 97.34%. Whereas, the Naïve Bayes method, J48 technique, Bagging approach, kNN system, Xboost-FM algorithm, Hybrid 2D CNN, model, and Decision Tree methodology have minimal outcomes based on several measures.

DOI: https://doi.org/10.54216/IJNS.260302

5. Conclusion

This study develops a Leveraging GPNSSFS-SAPPM. The intention of proposed GPNSSFS-SAPPM system relies on improving prediction model of student's higher education performance using metaheuristic optimization algorithms. The data pre-processing model is employed at first by applying mean normalization for converting an input data into a suitable format. In addition, GOA is exploited for the feature selection process. Followed by, the classification process is done by GPNSS. At last, the parameter tuning process is performed through HGSO algorithm for improving the classification performance of the GPNSS classifier. An extensive experiment was conducted to validate the performance of the GPNSSFS-SAPPM technique. The simulation outcomes indicated that the GPNSSFS-SAPPM method highlighted betterment over other recent approaches.

References

[1] I. Deli, V. Uluçay, and Y. Polat, "N-valued neutrosophic trapezoidal numbers with similarity measures and application to multi-criteria decision-making problems," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, pp. 4493–4518, 2022. [Online]. Available: https://doi.org/10.1007/s12652-021-03294-7

[2] J. Wang, J. Wang, and Y. Ma, "Possibility degree and power aggregation operators of single-valued trapezoidal neutrosophic numbers and applications to multi-criteria group decision-making," *Cogn. Comput.*, vol. 13, pp. 657–672, 2021. [Online]. Available: https://doi.org/10.1007/s12559-020-09736-2

[3] S. Liu, "Evaluation of physical education teaching based on analytic hierarchy process," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 16, no. 19, pp. 32–43, Oct. 2021. [Online]. Available: https://doi.org/10.3991/ijet.v16i19.26157

[4] R. K. Saini, A. Sangal, and Manisha, "Application of single valued trapezoidal neutrosophic numbers in transportation problem," *Neutrosophic Sets Syst.*, vol. 35, pp. 563–583, Feb. 2020. [Online]. Available: https://fs.unm.edu/nss8/index.php/111/article/view/4024

[5] Z. Yi, L. Yao, and H. Garg, "Power geometric operations of trapezoidal Atanassov's intuitionistic fuzzy numbers based on strict t-norms and t-conorms and its application to multiple attribute group decision making," *Int. J. Fuzzy Syst.*, vol. 26, pp. 239–259, 2024. [Online]. Available: https://doi.org/10.1007/s40815-023-01591-1

[6] E. Alyahyan and D. Düştegör, "Predicting academic success in higher education: literature review and best practices," *Int. J. Educ. Technol. Higher Educ.*, vol. 17, no. 1, p. 3, 2020.

[7] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, p. 552, 2021.

[8] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, p. 237, 2020.

[9] H. Waheed et al., "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Hum. Behav*, vol. 104, p. 106189, 2020.

[10] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ*, vol. 9, no. 1, p. 11, 2022.

[11] J. Wang and Y. Yu, "Machine learning approach to student performance prediction of online learning," *PLoS One*, vol. 20, no. 1, p. e0299018, 2025.

[12] G. Chami and B. Cockburn, "Higher education in the age of artificial intelligence: striking a balance between embracing innovation and preserving authenticity," *Caribbean Teach. Scholar*, vol. 8, no. 1, 2025.

[13] A. V. Ryabko et al., "A novel neuro-fuzzy approach for evaluating educational programme quality and institutional performance in higher education," in *CEUR Workshop Proc.*, 2025, pp. 102–124.

[14] R. Parkavi, P. Karthikeyan, and A. S. Abdullah, "Predicting academic performance of learners with the three domains of learning data using neuro-fuzzy model and machine learning algorithms," *J. Eng. Res.*, vol. 12, no. 3, pp. 397–411, 2024.

DOI: https://doi.org/10.54216/IJNS.260302

[15] A. Kukkar et al., "A novel methodology using RNN+ LSTM+ ML for predicting student's academic performance," *Educ. Inf. Technol.*, vol. 29, no. 11, pp. 14365–14401, 2024.

[16] N. Abuzinadah et al., "Role of convolutional features and machine learning for predicting student academic performance from MOODLE data," *PLoS One*, vol. 18, no. 11, p. e0293061, 2023.

[17] A. Maulana et al., "Leveraging artificial intelligence to predict student performance: a comparative machine learning approach," *J. Educ. Manage. Learn*, vol. 1, no. 2, pp. 64–70, 2023.

[18] Y. S. Kim et al., "Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models," *Sustain. Cities Soc.*, vol. 118, p. 105570, 2025.

[19] K. U. Kiran, T. Swati, and W. Yasmeen, "Maximizing mutual information: optimal keypoint selection using golf optimizer for improved multimodal image registration," in *Int. Conf. Adv. Mater. Manuf. Sustain. Dev. (ICAMMSD 2024)*, Atlantis Press, 2025, pp. 193–207.

[20] S. Bhuvaneshwari and C. Sweety, "Generalized possibility neutrosophic soft set and its application," *Int. J. Neutrosophic Sci.*, vol. 15, no. 2, 2021.

[21] N. B. R. Nayana et al., "Scalable bearing fault diagnosis using metaheuristic feature selection and machine learning for diverse operating conditions," *Syst. Sci. Control Eng.*, vol. 13, no. 1, p. 2469606, 2025.

[22] R. Elkharoua, "Students performance dataset," *Kaggle*, Accessed Nov. 2024.

[23] N. A. Butt et al., "Performance prediction of students in higher education using multi-model ensemble approach," *IEEE Access*, vol. 11, pp. 136091–136108, 2023.

[24] S. D. A. Bujang et al., "Multiclass prediction model for student grade prediction using machine learning," *IEEE Access*, vol. 9, pp. 95608–95621, 2021.

[25] K. Qin, X. Xie, Q. He, and G. Deng, "Early warning of student performance with integration of subjective and objective elements," *IEEE Access*, vol. 11, pp. 72601–72617, 2023.

DOI: https://doi.org/10.54216/IJNS.260302