

## Feature Selection with Optimal Stacked Sparse Autoencoder for Data Mining

Manar Ahmed Hamza<sup>1,\*</sup>, Siwar Ben Haj Hassine<sup>2</sup>, Ibrahim Abunadi<sup>3</sup>, Fahd N. Al-Wesabi<sup>2,4</sup>,  
Hadeel Alsolai<sup>5</sup>, Anwer Mustafa Hilal<sup>1</sup>, Ishfaq Yaseen<sup>1</sup> and Abdelwahed Motwakel<sup>1</sup>

<sup>1</sup>Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, Al-Kharj, 16278, Saudi Arabia

<sup>2</sup>Department of Computer Science, College of Science and Arts at Mahayil, King Khalid University, Muhayel Aseer, 62529, Saudi Arabia

<sup>3</sup>Department of Information Systems, Prince Sultan University, Riyadh, 11586, Saudi Arabia

<sup>4</sup>Faculty of Computer and IT, Sana'a University, Sana'a, 61101, Yemen

<sup>5</sup>Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Saudi Arabia

\*Corresponding Author: Manar Ahmed Hamza. Email: ma.hamza@psau.edu.sa

Received: 30 October 2021; Accepted: 05 January 2022

**Abstract:** Data mining in the educational field can be used to optimize the teaching and learning performance among the students. The recently developed machine learning (ML) and deep learning (DL) approaches can be utilized to mine the data effectively. This study proposes an Improved Sailfish Optimizer-based Feature Selection with Optimal Stacked Sparse Autoencoder (ISOFS-OSSAE) for data mining and pattern recognition in the educational sector. The proposed ISOFS-OSSAE model aims to mine the educational data and derive decisions based on the feature selection and classification process. Moreover, the ISOFS-OSSAE model involves the design of the ISOFS technique to choose an optimal subset of features. Moreover, the swallow swarm optimization (SSO) with the SSAE model is derived to perform the classification process. To showcase the enhanced outcomes of the ISOFS-OSSAE model, a wide range of experiments were taken place on a benchmark dataset from the University of California Irvine (UCI) Machine Learning Repository. The simulation results pointed out the improved classification performance of the ISOFS-OSSAE model over the recent state of art approaches in terms of different performance measures.

**Keywords:** Data mining; pattern recognition; feature selection; data classification; SSAE model

### 1 Introduction

Data mining (DM) is an approach of extracting hidden prediction data from the larger database; also, it is one of the modern technologies having greater potential to assist institutions/Universities to focus on the fundamental data in their data warehouses [1]. The DM tool predicts behaviors and future trends, enabling institutions to make knowledge-driven, proactive decisions. The automatic,



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

prospective analysis provided by the DM tool develops the analysis of historical events offered by retrospective tool typical of decision support system. The DM tool is capable of answering institution questions that usually take a lot of time to solve [2]. Nowadays, several industries utilize the DM tool to prepare marketing strategies and to make decisions towards target segmented customers for achieving their goals. However, the various university ignored practicing DM methods. The DM Application in the field of education is a new tendency in the globally competitive business. Understanding the DM application, terms, tasks, and techniques are the basis for the development of DM tools in educational sectors. Hence, it is necessary to examine the function of DM in educational sectors [3].

Educational Data Mining (EDM) is the application of the DM system on education information. EDM aims to study this information and solve education research problems. EDM acts toward the development of modern technology to examine the education information, and utilize DM tools for better understanding student learning environments [4–6]. The EDM method converts transform raw data coming from the education system into helpful data that might have a greater impact on education practice and research. The ever-growing technologies in the education system have made a great number of data available. EDM offers a clearer picture of learners and their learning processes [7] and provides a significant amount of relevant information. It employs the DM technique to examine education information and resolve education problems. Like other DM methodologies extraction process, EDM extract novel, interesting, interpretable, and useful data from the education system. But EDM is specially designed for using different kinds of information in the education system [8]. Then, this technique is utilized for enhancing the knowledge about the settings, educational phenomena, and students where they can learn [9]. The development of computational methods that integrate data and theory would assist to improve the quality of teaching and learning (T& L) procedures.

EDM research studies several fields, involving specific learning strategies from computer-adoptive testing (as well as testing at a larger scale), education software, computer-assisted collaborative learning, and the factors i.e., related to the non-retention/student failure in courses [10]. Several critical areas include the development of the student system; applications of the EDM method have been in examining pedagogical supports (in learning software, and other fields, namely collaborative learning behavior) and improving/determining the method of a domain knowledge structure. There has been growing interest in the field of an EDM system. This newly emerging field, named EDM, is a concern with the development of techniques that determine knowledge from data originating from the education environment. EDM system employs several methods, for example, K-nearest neighbor (NN), Decision Trees (DT), Naïve Bayes (NB), Neural Networks (NN), etc. analysis, and the Prediction of student performances is an important building block in education environments. academic performance of students is a critical factor in creating their future. Student Academic performance isn't a result of only one determining factor as well it depends heavily on socio-economic, personal, and psychological factors.

Serik et al. [11] suggest the incorporation of techniques like data analysis and artificial intelligence (AI) methods, with learning management systems to enhance learning. These goals are defined in novel normality which tries to find a strong education system, where particular activity can be performed in an online system, amongst technique allows the student to have a virtual assistant to assist them in their learning. In [12], a novel predictive method to evaluate student performances in academics has been designed based on clustering and classification systems and tested in a real-time manner with the student's database of several academic fields of high education institutions in Kerala, India. The results show that the hybrid method integrating classification and clustering methods yields outcomes i.e., much greater in terms of attaining accuracy in prediction of student academic performances.

Francis et al. [13] aimed at finding fast, slow and average learners amongst students and displays it by prediction DM tool with classification-based models. Mhetre et al. [14] determine the student performances with various classification methods and discover the better one that produces optimum outcomes. Education Database is gathered from a Saudi University database. The database is preprocessed for filtering duplicate records; missing areas are recognized and filled with certain information. DL methods such as DNN and DM methods like random forest (RF), support vector machine (SVM), DT, and NB are applied to the dataset with Rapid Miner and Weka methods.

Aslam et al. [15] separate the significance of 2 blooming techniques, ML and Blockchain, in the education domain. Blockchain technique, with data immutability as its major benefits, was employed in the miscellaneous field for security factors. It is utilized for storing securely the achievement/degree certificates. This data will be included by the university/college to the blockchain technology, that is shared/accessed by the students via online resume with employers. This method is highly secured since there is no need to worry about the loss of data/modification to the institution. Shah et al. [16] gathered several samples of distinct kinds of student attributes with the survey forms which are related to academic performances. Then, select a few significant features with distinct feature extraction methods. Next, employed a few ML methods to that pre-processed dataset. Malhotra et al. [17] examine the educational quality that is tightened closely with the sustainable developmental objectives. The performance of the method has generated excessive information that should be suitably processed to attain useful data which is highly beneficial for future planning and development. Student grade and mark predictions from their historic academic data are useful and popular applications in the EDM system, hence it become a useful data provider that is employed in various forms to enhance the educational quality in the country. The classification method would forecast the grades whereas the regression method would forecast the mark, lastly, the outcomes are attained from both the models were investigated.

This study proposes an Improved Sailfish Optimizer-based Feature Selection with Optimal Stacked Sparse Autoencoder (ISOFS-OSSAE) for data mining and pattern recognition in the educational sector. The proposed ISOFS-OSSAE model aims to mine the educational data and derive decisions based on the feature selection and classification process. Moreover, the ISOFS-OSSAE model involves the design of the ISOFS technique to choose an optimal subset of features. Moreover, the swallow swarm optimization (SSO) with the SSAE model is derived to perform the classification process. To showcase the enhanced outcomes of the ISOFS-OSSAE model, a wide range of experiments were taken place on a benchmark dataset from the UCI repository.

## 2 The Proposed Model

In this work, a novel ISOFS-OSSAE technique is derived aims to mine the educational data and derive decisions based on the feature selection and classification process. The proposed ISOFS-OSSAE model involves the design of the ISOFS technique to choose an optimal subset of features. Followed by, the OSSAE model-based classifier is derived in which the parameter tuning of the SSAE model is done by the use of the SSO algorithm.

## 2.1 Design of ISOFS Technique

The SFO approach is depending upon the behavior of sailfishes. The sailfish population can be determined by a candidate solution in the SFO method [18]. The population in solution space is arbitrarily created. The location matrix of initiated sailfish has been demonstrated by:

$$X_{SF}^{(t)} = [X_{i,j}^{(t)}]_{m \times d} = \begin{bmatrix} X_{1,1}^{(t)} & X_{1,2}^{(t)} & X_{1,d}^{(t)} \\ X_{2,1}^{(t)} & X_{2,2}^{(t)} & X_{2,d}^{(t)} \\ \vdots & \vdots & \vdots \\ X_{m,1}^{(t)} & X_{m,2}^{(t)} & X_{m,d}^{(t)} \end{bmatrix} \quad (1)$$

In which  $m$  signifies the quantity of sailfish in sailfish population,  $d$  denotes the dimensional parameter,  $X_{i,j}^{(t)}$  characterizes the value of  $j_{th}$  variable of  $i_{th}$  sailfish,  $t$  indicates the number of existing iterations. The matrix afterward the initiation of the sardine location has been demonstrated by:

$$X_s^{(t)} = [X_{i,j}^{(t)}]_{n \times d} = \begin{bmatrix} X'_{1,1}^{(t)} & X'_{1,2}^{(t)} & X'_{1,d}^{(t)} \\ X'_{2,1}^{(t)} & X'_{2,2}^{(t)} & X'_{2,d}^{(t)} \\ \vdots & \vdots & \vdots \\ X'_{n,1}^{(t)} & X'_{n,2}^{(t)} & X'_{n,d}^{(t)} \end{bmatrix} \quad (2)$$

whereas  $n$  indicates the number of sardines in the sardine population,  $X_{i,j}^{(t)}$  means the value of  $j_{th}$  parameter of  $i_{th}$  sardine.

To estimate the quality of sardines and sailfish, the fitness of sardine and sailfish solution can be resolved and  $F$  represent fitness function as well as stored in matrix format. The fitness matrix of sailfish can be expressed as:

$$F_{SF} = (t) \begin{Bmatrix} F_1^{(t)} \\ F_2^{(t)} \\ \vdots \\ F_m^{(t)} \end{Bmatrix} = \begin{Bmatrix} F(X_{1,1}^{(t)}) & X_{1,2}^{(t)} & x_{1,d}^{(t)} \\ F(X_{2,1}^{(t)}) & X_{2,2}^{(t)} & x_{2,d}^{(t)} \\ \vdots & & \\ F(x_{m,1}^{(t)}) & X_{m,2}^{(t)} & x_{m,d}^{(t)} \end{Bmatrix} \quad (3)$$

Now,  $F_{SF}^{(t)}$  signifies the fitness matrix of sailfish,  $m$  implies the amount of sailfish,  $F_i^{(t)}$  represent the fitness value of the  $i_{th}$  sailfish. It can be formulated by:

$$F_s^{(t)} = \begin{Bmatrix} F_1^{(t)} \\ F_2^{(t)} \\ \vdots \\ F_n^{(t)} \end{Bmatrix} = \begin{Bmatrix} F & X'_{1,1}^{(t)} & X'_{1,2}^{(t)} & X'_{1,d}^{(t)} \\ F & x_{2,1}^{(t)} & x_{2,2}^{(t)} & x_{2,d}^{(t)} \\ F & X'_{n,1}^{(t)} & X'_{n,2}^{(t)} & X'_{n,d}^{(t)} \end{Bmatrix} \quad (4)$$

Here,  $F_s^{(t)}$  denotes the sardine fitness matrix,  $n$  signifies the quantity of sardines, and  $F_i^{(t)}$  indicates the fitness value of  $i_{th}$  sardine. The hunting nature of sailfish alters its location based on the location of other sailfish near the sardines. The equation to update the location of sailfish is given in the following:

$$X_i^{(t)} = X_{elite}^{(t-1)} - \lambda_i^{(t-1)} \times \left( rand \times \left( \frac{X_{elite}^{(t-1)} - X_{injured}^{(t-1)}}{2} \right) - X_i^{(t-1)} \right) \quad (5)$$

where  $X_{elite}^{(t-1)}$  indicates the elite sailfish in the t-l iteration,  $X_{injured}^{(t-1)}$  represent the sardine i.e., severely injured in t-l iteration,  $X_i^{(t-1)}$  shows the  $i_{th}$  sailfish in the t-l iteration,  $X_i^{(t)}$  refers to the  $i_{th}$  sailfish in  $t$  iteration,

rand denotes an arbitrary number within [0,1],  $\lambda_i$  indicates the upgrade coefficient:

$$\begin{cases} \lambda_i^{(t)} = 2 \times \text{rand} \times D^{(t)} - D^{(t)} \\ D^{(t)} = 1 - \left( \frac{M_{SF}^{(t)}}{M_{SF}^{(t)} + M_S^{(t)}} \right) \end{cases} \quad (6)$$

Let  $M_{SF}^{(t)}$  be the amount of sailfish in  $t$  iteration and  $M_S^{(t)}$  signifies the amount of sardines in  $t$  iteration.

To prevent from sailfish attack, the sardine would consider the location of the elite sailfish and the attack power of sailfish in all the iterations and upgrade the location. The location upgrade equation of sardines is expressed as:

$$\begin{cases} X_i'^{(t)} = \text{rand} \times (X_{elite}^{(t-1)} - X_i'^{(t-1)} + Q^{(t-1)}) \\ Q^{(t-1)} = A \times (1 - (2 \times (t - 1) \times \xi)) \end{cases} \quad (7)$$

whereas  $X_i'^{(t-1)}$  indicates the  $i_{th}$  sardine in  $t-1$  iteration,  $X_i'^{(t)}$  represent the  $i_{th}$  sardine in  $t$  iteration,  $Q^{(t-1)}$  signifies the attack power of sailfish in  $t-1$  iteration,  $A$  &  $\xi$  denotes attack power coefficient, within 4 and 0.001,  $t$  indicates the number of present iterations, the amount of upgrade sardines  $\gamma$  and the amount of parameter  $\eta$  based on the attack power of the sailfish can be represented by:

$$\begin{cases} \gamma^{(t)} = M_S^{(t)} \times Q(t) \\ \eta^{(t)=d_i^{(t)}} \times Q^{(t)} \end{cases} \quad (8)$$

Now  $M_S^{(t)}$  signifies the quantity of sardines in  $= t$  iteration and  $d_i^{(t)}$  characterizes the amount of  $i_{th}$  sardine parameter in the  $t$  iteration.

Once the fitness of sardine is minimum when compared to the sailfish, the location of the captured sardine is substituted with the location of the sailfish:

$$X_i^{(t)} = X_i'^{(t)} \text{ if } F(X_i^{(t)}) > F(X_i'^{(t)}) \quad (9)$$

Let  $X_i^{(t)}$  be the location of sailfish in  $t$  iteration and  $X_i'^{(t)}$  represent the location of sardine in  $t$  iteration.

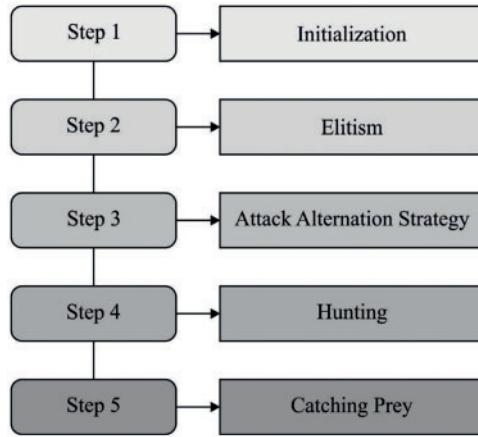
The  $u$  weight inertia is presented into the alternate attack and pursuit procedure of sailfish [19], and the local searching capacity of sardines and sailfish is improved using weight inertia  $u$ . The location upgrade equation can be given below:

$$\begin{cases} X_i'^{(t)} = \text{rand} \times (u(t - 1) \times X_{elite}^{(t-1)} - X_i'^{(t-1)} + Q^{(t-1)}) \\ X_i^{(t)} = u(t - 1) \times X_{elite}^{(t-1)} - \lambda_i^{(t-1)} \times \left( \text{rand} \times \left( \frac{X_{elite}^{(t-1)} - X_{injured}^{(t-1)}}{2} \right) - X_i^{(t-1)} \right) \end{cases} \quad (10)$$

In the equation, rand denotes an arbitrary number within [0,1] The weight inertia  $u$  can be expressed by:

$$u(t) = u_{\max} + (u_{\max} - u_{\min}) \times \exp \left( -25 \times \left( \frac{t}{T} \right)^3 \right) \quad (11)$$

Now  $u_{\min}$  &  $u_{\max}$  represents the lower and upper bounds of weight inertia  $u$ .  $u_{\min} = 0.4$ ,  $u_{\max} = 0.9$ .  $t$  as well as  $T$  represents the present number of iterations and the maximal amount of iterations. The steps involved in the SFO algorithm are given in Fig. 1.



**Figure 1:** Steps involved in SFO algorithm

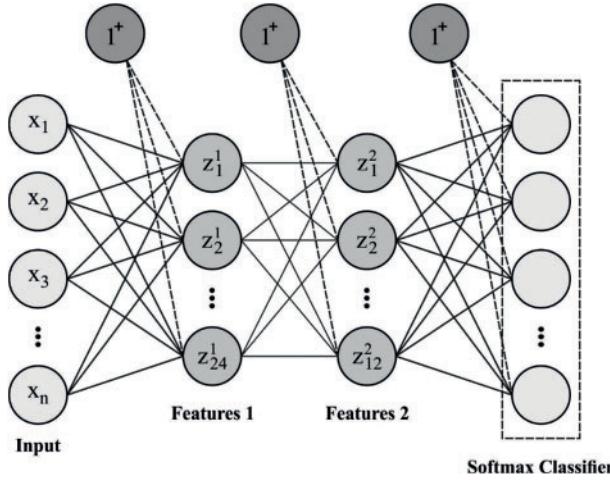
The ISOFS technique is applied to discover the feature area effectively and generate a proper set of features. The feature selection can be considered as a multiobjective optimization problem as it requires fulfilling distinct aims to get optimum solutions that reduce the features and improve the classification performance. Therefore, a fitness function (FF) is derived to obtain solutions for attaining a balance amongst two objectives as defined below:

$$\text{fitness} = \alpha \Delta_R(D) + \beta \frac{|Y|}{|T|} \quad (12)$$

$\Delta_R(D)$  denotes the error level of the classifier,  $|Y|$  represents the number of features, and  $|T|$  denotes the total feature count that exists in the dataset. Here,  $\alpha$  indicates a parameter  $\in [0, 1]$  related to the weight of error rate of classifier, and  $\beta = 1 - \alpha$  characterizes the standing of reduced features.

## 2.2 Design of OSSAE Based Classification Model

The reduced feature subsets are fed into the OSSAE model to carry out the classification process. The auto encoder (AE) approach is an asymmetrical NN that extracts the features with minimal recreation error. Li et al. [20] projected a perspective, termed ‘pre-train’, that separates a complex network to stack sub-networks. The training failure could be prevented since the network parameter of all the layers are allocated certain values, instead of arbitrary initiation. Nevertheless, the stacked sub-network provides generalization capacity and low training efficacy because of the easiness of the single-hidden neuron and the complexity in the selection of variables. To resolve these above-mentioned problems, the SSAE method is presented based on 2-phase networks with 5 hidden layers in all the networks. The overall framework of the SSAE method is demonstrated in Fig. 2. Initially, the  $X$  input layer is mapped to a  $h_3$  hidden neuron (termed low-level hidden layer). Next,  $h_3$  is mapped back to the recreated layer,  $X_{rec}$ . Then, the  $h_3$  hidden feature is converted into the input parameter for acquiring  $h_6$  (high-level hidden layer) [21].



**Figure 2:** Structure of SSAE

In the SSAE, the feature learning method follows a sequence of processes, like convolution, denoising, pooling, activation, and batch normalization (BN). A summary of this operation can be thoroughly explained in the following: Denoising to attain the strong and illustrative learning feature of the flame image, a denoising AE learning method is utilized to add distinct noises with the input signal. The white Gaussian noises are taken into account, such as the corrupted version  $X_n$  is attained by a fixed corruption ratio to the input  $X$ :

$$I_{x_n} = I_X + \varphi \xi \quad (13)$$

A rectified linear unit (ReLU) is utilized as an activation function of the hidden *neuron*  $\gamma$ , determined by:

$$y(\gamma) = \max(0, \gamma) \quad (14)$$

The ReLU is an unsaturated piecewise linear function, i.e., quicker when compared to the saturated non-linear functions, like TanH and Sigmoid. Particularly, the Sigmoid function is employed in the 3rd decoders for ensuring the intensity range of the recreated layer  $X_{rec}$  is reliable with the input layer  $X$ , as:

$$y(\gamma) = \frac{1}{1 + e^{-\gamma}} \quad (15)$$

**Pooling and upsampling:** The pooling function is performed for reducing the parameter of the network. Here,  $P(r \times r + t)$  represents the pooling function which condenses the feature map through choosing a maximal value with a step of  $t$  and a  $r \times r$  conversion kernel. The pooling function helps improve the translation invariance as:

$$\delta_{ij}^k = \max_{0 \leq n \leq p} \{\gamma_{i:r+t, j:r+t}^k\} \quad (16)$$

In which  $\gamma_{ij}^k$  &  $\delta_{ij}^k$  represents the value of position  $(i, j)$  in the  $k^{th}$  feature map of outputs and inputs [22]. Consider  $s_{ij}$  ( $i \in (1, E)$ ,  $j \in (1, F)$ ) signifies the activation of hidden layer  $j$ , in which  $E$  denotes the number of images in the trained data set and  $F$  denotes the number of layers in the hidden neuron. Next, the average activation of all the hidden neurons  $j$  are evaluated as:

$$p_j = \frac{1}{E} \sum_{i=1}^E s_{ij} \quad (17)$$

whereas  $p_j$  projected to be closer to 0, which implies the neuron of the hidden neuron is generally inactive". A  $P_{\text{penalty}}$  the penalty term is added to the loss function, that penalize  $p_j$  once it significantly deviates from the  $p_{\text{target}}$  sparse target. The  $P_{\text{penalty}}$  penalty term can be determined by:

$$P_{\text{penalty}} = \sum_{j=1}^F KL(p_{\text{target}} || p_j) \quad (18)$$

Let  $KL(p_{\text{target}} || p_j)$  be the KullbackLeibler divergence (KL divergence) that performs as a sparsity limitation:

$$KL(p_{\text{target}} || p_j) = p_{\text{target}} \log \frac{p_{\text{target}}}{p_j} + (1 - p_{\text{target}}) \log \frac{1 - p_{\text{target}}}{1 - p_j} \quad (19)$$

When  $p_{\text{target}} = p_j$ ,  $KL(p_{\text{target}} || p_j) = 0$ . Or else, the  $KL(p_{\text{target}} || p_j)$  monotonically increased as  $p_j$  deviate from  $p_{\text{target}}$ .

For optimally adjusting the parameters involved in the SSAE model, the SSO algorithm is utilized.

SSO approach was inspired by the collaboration of swallow and the interface between flock members has accomplished optimal results [22,23]. This method was proposed a meta-heuristic model based on unique features of swallow comprised of intelligent social relation, fast flight, and hunting skill. Now, this approach is similar to the particle swarm optimization (PSO) method however it could be particular characteristics that couldn't be initiated in common methods which contain: Leader Particles ( $l_i$ ), Explorer Particles ( $e_j$ ), and Aimless Particles ( $o_i$ ), which has certain responsibilities in the group. The particle  $e_j$  is accountable for searching the problem space. It performed these exploring performances in the control of the number of variables. It can be expressed by

$$V_{HL_{i+1}} = V_{HL_j} + \alpha_{HL} \text{rand}() (e_{best} - e_j) + \beta_{HL} \text{rand}() (HL_j - e_j) \quad (20)$$

Eq. (20) exhibits the velocity vector parameter in the path of the global leader.

$$\alpha_{HL} = \{\text{if } (e_j = 0 || e_{best} = 0) \rightarrow 1.5\} \quad (21)$$

Eqs. (21)–(12) evaluates the acceleration coefficient parameter ( $\alpha_{HL}$ ) which straightforwardly affects individual knowledge of each particle.

$$\alpha_{HL} = \begin{cases} \text{if } (e_i < e_{best}) \&\& (< HL_i) \rightarrow \frac{\text{rand}() \cdot e_i}{e_i \cdot e_{best}} e_i, e_{best} \neq 0 \\ \text{if } (e_i < e_{best}) \&\& (e_i > HL_i) \rightarrow \frac{2\text{rand}() \cdot e_i}{1} e_i \neq 0 \\ \text{if } (e_i > e_{best}) \rightarrow \frac{e_{best}}{1} \end{cases} \quad (22)$$

$$\beta_{HL} = \{\text{if } (e_j = 0 || e_{best} = 0) \rightarrow 1.5\} \quad (23)$$

$$\beta_{HL} = \begin{cases} \text{if } (e_i < e_{best}) \& \& (e_i < HL_i) \rightarrow \frac{\text{rand}() \cdot e_i}{e_i \cdot HL_i} e_i, \quad HL_i \neq 0 \\ \text{if } (e_i < e_{best}) \& \& (e_i > HL_i) \rightarrow \frac{2\text{rand}() \cdot e_i}{1} e_i \neq 0 \\ \text{if } (e_i > e_{best}) \rightarrow \frac{HL_i}{\frac{1}{(2 \cdot e_i)}} \end{cases} \quad (24)$$

Eqs. (23) and (24) calculates the acceleration coefficient parameter ( $\beta_{HL}$ ) which straightforwardly affects the integrated knowledge of each particle. Both acceleration coefficients are quantified allows placing each particle with global leaders and optimal individual knowledge. The  $o_i$  particle utilizes the succeeding Eq. (25) for arbitrary motions:

$$o_{i+1} = o_i + \left[ \text{rand}(\{-1, 1\}) * \frac{\text{rand}(\min, \max)}{1 + \text{rand}()} \right] \quad (25)$$

In the SSO method, there are 2 types of leaders: global and local leaders. The particle is separated into groups. The particle in each group is often similar. Next, an optimal particles in each group is chosen and termed as a local leader. After, an optimal particle amongst the local leaders is selected and termed as a global leader. The particle modification converges and way depends on the place of this article.

### 3 Results and Discussion

The experimental results analysis of the ISOFS-OSSAE technique takes place using the benchmark dataset from the UCI repository [24]. It includes a total of 649 samples with 33 attributes and 2 class labels. The results are investigated under varying dimensions.

Tab. 1 and Fig. 3 demonstrate the best cost and number of chosen features offered by the ISOFS with other FS methods. The results portrayed that the Information gain and CFSSubsetEval techniques have attained poor performance with the maximum best cost of 0.386920 and 0.366000 respectively. Followed by, the genetic algorithm (GA) and PSO models have obtained moderate best cost of 0.165283 and 0.183638 respectively. In line with this, the ant colony optimization (ACO) algorithm has accomplished near-optimal best cost of 0.030509. However, the proposed ISOFS technique has accomplished superior outcomes with the least good cost of 0.02156.

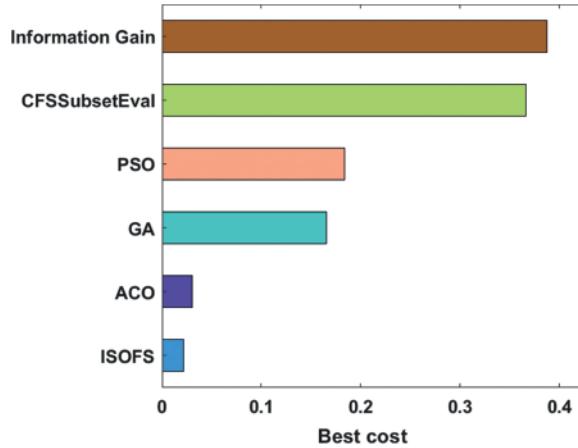
**Table 1:** Best cost analysis of different FS models

Methods	Best cost	Selected features
ISOFS	0.02156	21, 16, 6, 15, 30, 29, 10, 5, 17, 28
ACO	0.030509	28, 11, 16, 14, 32, 18, 24, 4, 15, 30, 25, 29, 3, 10, 5, 17, 9, 19, 23, 8, 21, 31, 12, 6
GA	0.165283	23, 25, 27, 1, 2, 5, 4, 7, 8, 10, 12, 13, 16, 17, 19, 21, 29, 31
PSO	0.183638	1, 2, 3, 5, 6, 8, 9, 11, 12, 15, 16, 18, 17, 19, 22, 23, 28, 29, 30, 32
CFSSubsetEval	0.366000	1, 15, 21, 31, 32

(Continued)

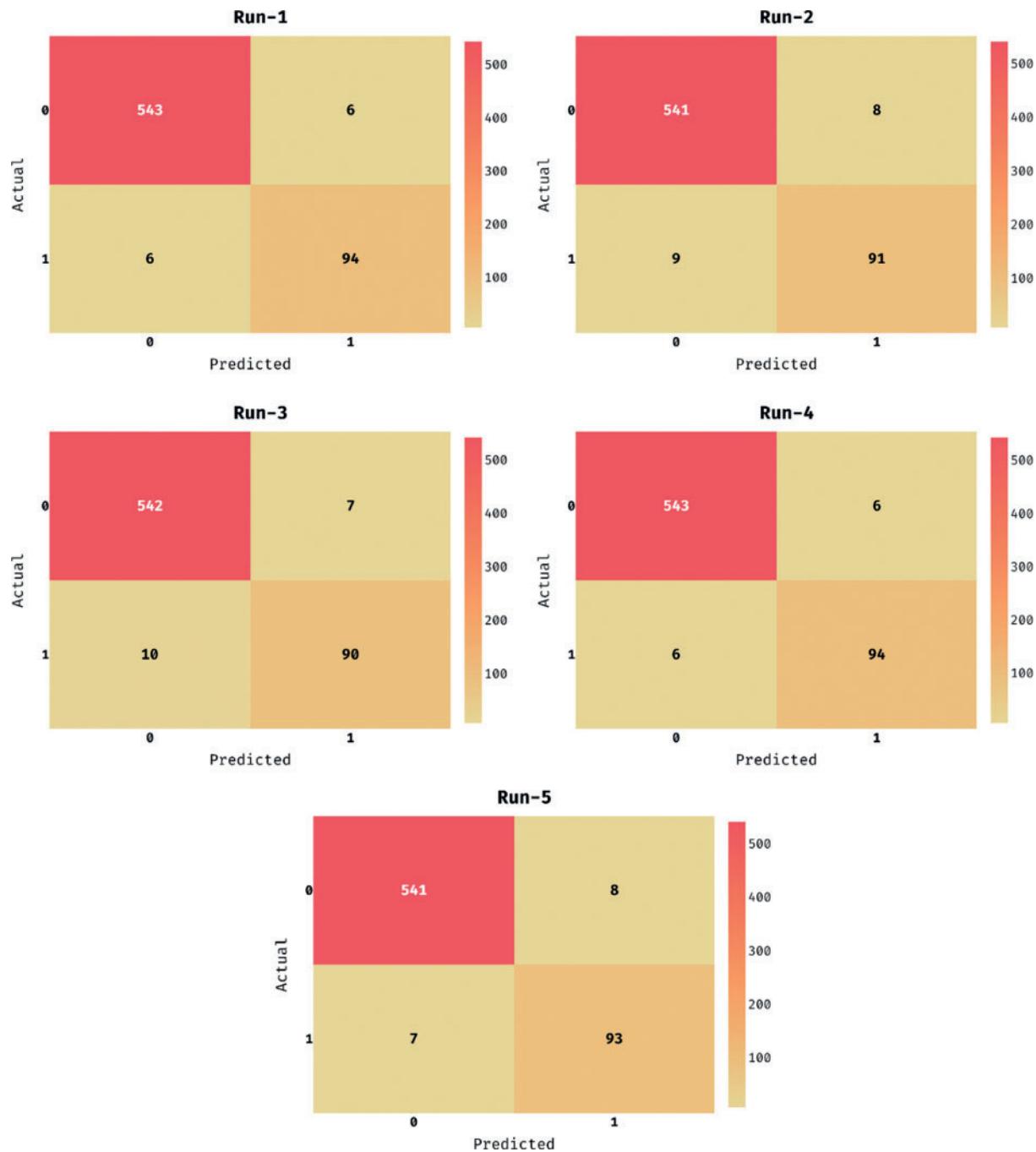
**Table 1:** Continued

Methods	Best cost	Selected features
Information gain	0.386920	32, 31, 15, 1, 21, 8, 11, 14, 7, 9, 4, 26, 12, 22, 23, 2, 10, 5, 18, 19, 17, 16, 20, 6, 24, 3, 25, 13, 27, 29, 30, 28

**Figure 3:** Best cost analysis of ISOFS with other techniques

[Fig. 4](#) portrays the confusion matrices generated by the ISOFS-OSSAE technique under five test runs. On test run-1, the ISOFS-OSSAE technique has identified 543 instances into class 0 and 94 instances into class 1. Besides, on test run-2, the ISOFS-OSSAE technique has identified 541 instances into class 0 and 91 instances into class 1. Moreover, on test run-3, the ISOFS-OSSAE technique has identified 542 instances into class 0 and 90 instances into class 1. Likewise, on test run-4, the ISOFS-OSSAE technique has identified 543 instances into class 0 and 94 instances into class 1. At last, on test run-5, the ISOFS-OSSAE technique has identified 541 instances into class 0 and 93 instances into class 1. The values present in the confusion matrix are transformed in terms of TP, TN, FP, and FN in [Tab. 2](#).

[Tab. 3](#) provides a detailed overall classification results analysis of the ISOFS-OSSAE technique under five test runs. The table values denoted that the ISOFS-OSSAE technique has gained effective classifier results. For instance, with run-1, the ISOFS-OSSAE technique has attained  $prec_n$ ,  $reca_l$ ,  $F_{measure}$ ,  $acc_y$ , and  $kappa$  of 0.9891, 0.9891, 0.9891, 0.9815, and 0.8258 respectively. In line with, with run-2, the ISOFS-OSSAE technique has offered  $prec_n$ ,  $reca_l$ ,  $F_{measure}$ ,  $acc_y$ , and  $kappa$  of 0.9836, 0.9854, 0.9845, 0.9738, and 0.8122 respectively. Followed by, with run-3, the ISOFS-OSSAE technique has achieved  $prec_n$ ,  $reca_l$ ,  $F_{measure}$ ,  $acc_y$ , and  $kappa$  of 0.9819, 0.9872, 0.9846, 0.9738, and 0.8223. Along with that, with run-4, the ISOFS-OSSAE technique has gained  $prec_n$ ,  $reca_l$ ,  $F_{measure}$ ,  $acc_y$ , and  $kappa$  of 0.9891, 0.9891, 0.9891, 0.9815, and 0.8258 respectively. Lastly, with run-5, the ISOFS-OSSAE technique has attained  $prec_n$ ,  $reca_l$ ,  $F_{measure}$ ,  $acc_y$ , and  $kappa$  of 0.9891, 0.9891, 0.9891, 0.9815, and 0.8258 respectively.



**Figure 4:** Confusion matrices of ISOFS-OSSAE technique under five test runs

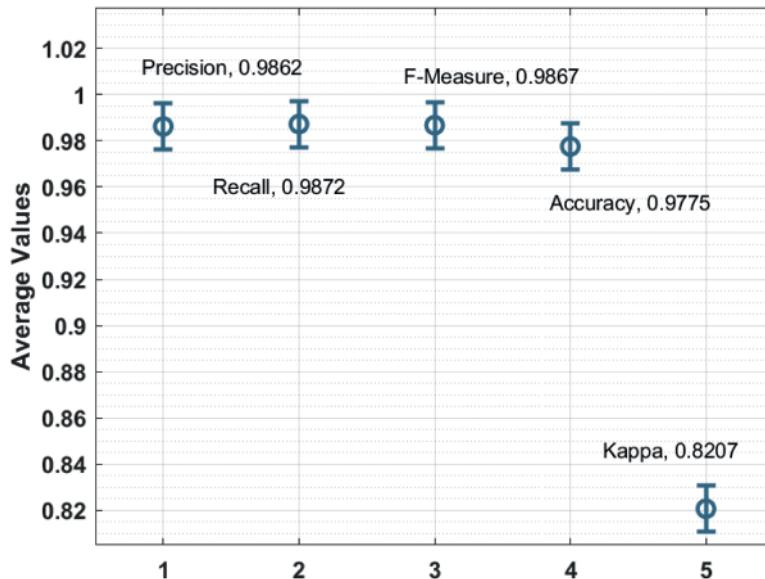
**Table 2:** Confusion matrix of ISOFS-OSSAE technique

No. of Runs	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)
Run-1	543	94	6	6
Run-2	541	91	9	8
Run-3	542	90	10	7
Run-4	543	94	6	6
Run-5	541	93	7	8

**Table 3:** Overall classification results analysis of ISOFS-OSSAE technique

No. of Runs	Precision	Recall	F-Measure	Accuracy	Kappa
Run-1	0.9891	0.9891	0.9891	0.9815	0.8258
Run-2	0.9836	0.9854	0.9845	0.9738	0.8122
Run-3	0.9819	0.9872	0.9846	0.9738	0.8223
Run-4	0.9891	0.9891	0.9891	0.9815	0.8258
Run-5	0.9891	0.9891	0.9891	0.9815	0.8258
Average	0.9862	0.9872	0.9867	0.9775	0.8207

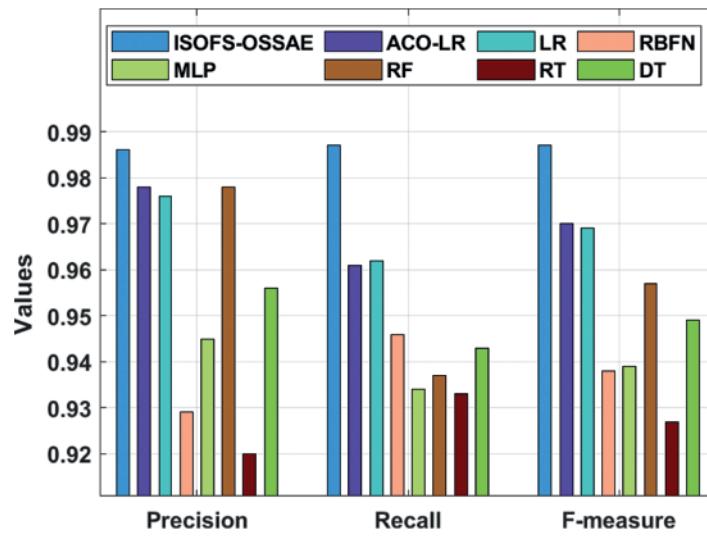
**Fig. 5** demonstrates the analysis of the average results of the ISOFS-OSSAE technique on the test dataset. The figure reported that the ISOFS-OSSAE technique has resulted in maximum average  $prec_n$ ,  $reca_t$ ,  $F_{measure}$ ,  $acc_y$ , and  $kappa$  of 0.9862, 0.9872, 0.9867, 0.9775, and 0.8207.

**Figure 5:** Average classification results analysis of ISOFS-OSSAE technique

**Tab. 4** offers a comprehensive comparative analysis of the ISOFS-OSSAE with recent techniques [25]. **Fig. 6** demonstrates the  $prec_n$ ,  $reca_l$ , and  $F_{measure}$  analysis of the ISOFS-OSSAE technique on the test dataset. The figure reported that the random tree (RT) technique has accomplished least performance with the minimal values of  $prec_n$ ,  $reca_l$ , and  $F_{measure}$ . Followed by, the radial basis function network (RBFN) technique has attained slightly increased values of  $prec_n$ ,  $reca_l$ , and  $F_{measure}$ . Next to that, the multilayer perceptron (MLP) model has accomplished moderate values of  $prec_n$ ,  $reca_l$ , and  $F_{measure}$ . In line with, the ACO with logistic regression (ACO-LR), LR, RF, and DT models have tried to obtain reasonable values of  $prec_n$ ,  $reca_l$ , and  $F_{measure}$ . However, the proposed ISOFS-OSSAE technique has resulted in maximum classification performance with the  $prec_n$ ,  $reca_l$ , and  $F_{measure}$  of 0.986, 0.987, and 0.987.

**Table 4:** A comparison study of ISOFS-OSSAE technique with recent methods

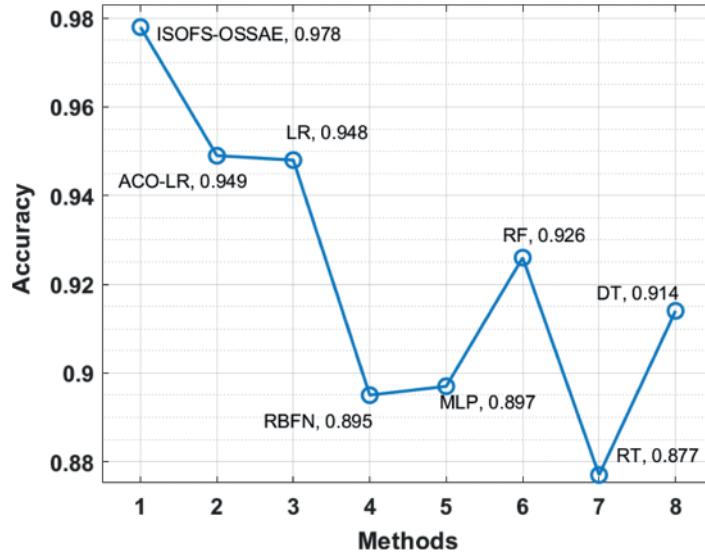
Methods	Precision	Recall	F-measure	Accuracy	Kappa
ISOFS-OSSAE	0.986	0.987	0.987	0.978	0.821
ACO-LR	0.978	0.961	0.970	0.949	0.796
LR	0.976	0.962	0.969	0.948	0.792
RBFN	0.929	0.946	0.938	0.895	0.614
MLP	0.945	0.934	0.939	0.897	0.592
RF	0.978	0.937	0.957	0.926	0.685
RT	0.920	0.933	0.927	0.877	0.542
DT	0.956	0.943	0.949	0.914	0.658



**Figure 6:** Comparative  $prec_n$ ,  $reca_l$ , and  $F_{measure}$  analysis of ISOFS-OSSAE technique

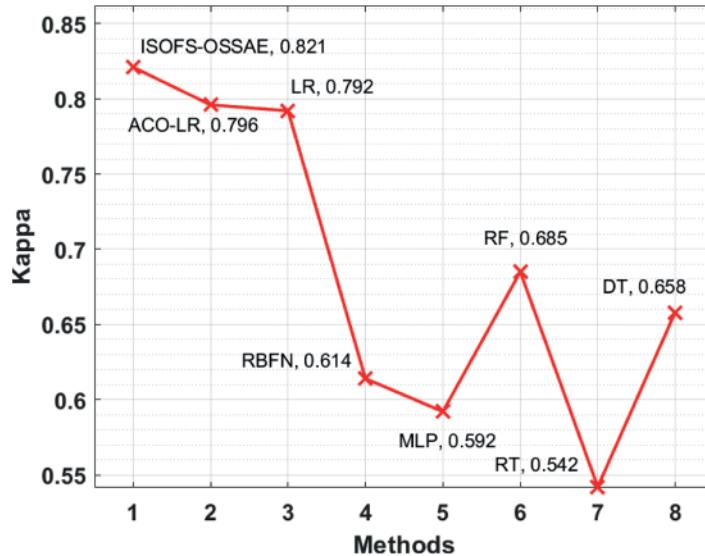
**Fig. 7** showcases the  $acc_y$  analysis of the ISOFS-OSSAE technique with recent approaches. The figure reported that the RBFN, MLP, and RT models have attained lower  $acc_y$  of 0.895, 0.897, and 0.877 respectively. Followed by, the DT and RF models have reached slightly improved  $acc_y$  of 0.914 and 0.926 respectively. Eventually, the LR and ACO-LR models have achieved reasonable  $acc_y$  of 0.948

and 0.949 respectively. However, the proposed ISOFS-OSSAE technique has resulted in maximum outcome with the  $acc_y$  of 0.978.



**Figure 7:** Comparative  $acc_y$  analysis of ISOFS-OSSAE technique

Fig. 8 displays the  $kappa$  analysis of the ISOFS-OSSAE technique with recent approaches. The figure stated that the RBFN, MLP, and RT models have reached inferior  $kappa$  of 0.614, 0.592, and 0.542 respectively. Also, the DT and RF models have got slightly enhanced  $kappa$  of 0.658 and 0.685 respectively. Ultimately, the LR and ACO-LR models have realized sensible  $kappa$  of 0.792 and 0.769 respectively. However, the proposed ISOFS-OSSAE technique has resulted in supreme outcome with the  $acc_y$  of 0.821. From the aforementioned figures and tables, it is ensured that the ISOFS-OSSAE technique is an effective tool for data mining and pattern recognition.



**Figure 8:** Comparative  $kappa$  analysis of ISOFS-OSSAE technique

#### 4 Conclusion

In this study, a novel ISOFS-OSSAE technique is derived aims to mine the educational data and derive decisions based on the feature selection and classification process. The proposed ISOFS-OSSAE model involves the design of the ISOFS technique to choose an optimal subset of features. Followed by, the OSSAE model-based classifier is derived in which the parameter tuning of the SSAE model is done by the use of the SSO algorithm. To showcase the enhanced outcomes of the ISOFS-OSSAE model, a wide range of experiments were taken place on a benchmark dataset from the UCI repository. The simulation results pointed out the improved classification performance of the ISOFS-OSSAE model over the recent state of art approaches with the higher accuracy of 0.978. Therefore, the ISOFS-OSSAE model can be utilized as an effective tool to mine data and recognize patterns. In the future, the classification performance can be improved by the utilization of clustering approaches.

**Funding Statement:** The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP 1/279/42). [www.kku.edu.sa](http://www.kku.edu.sa). The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

#### References

- [1] R. J. A. Cabrera, C. A. P. Legaspi, E. J. G. Papa, R. D. Samonte and D. D. Acula, "HeMatic: An automated leukemia detector with separation of overlapping blood cells through image processing and genetic algorithm," in *2017 Int. Conf. on Applied System Innovation, ICASI 2017. Proc.: IEEE, Sapporo, Japan*, pp. 985–987, 2017.
- [2] M. Kumar, A. J. Singh and D. Handa, "Literature survey on student's performance prediction in education using data mining techniques," *International Journal of Education and Management Engineering (IJEME)*, vol. 7, no. 6, pp. 40–49, 2017.
- [3] C. Romero and S. Ventura, "Data mining in education: Data mining in education," *WIREs Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.
- [4] Y. V. Paredes, R. F. Siegle, I. H. Hsiao and S. D. Craig, "Educational data mining and learning analytics for improving online learning environments," in *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, North America, vol. 64, no. 1, pp. 500–504, 2020.
- [5] N. Padhy, "The survey of data mining applications and feature scope," *International Journal of Computer Science, Engineering and Information Technology*, vol. 2, no. 3, pp. 43–58, 2012.
- [6] I. Almuniri and A. M. Said, "School's performance evaluation based on data mining," *International Journal of Engineering and Information Systems*, vol. 1, no. 9, pp. 56–62, 2017.
- [7] C. Jalota and R. Agrawal, "Analysis of educational data mining using classification," in *2019 Int. Conf. on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, pp. 243–247, 2019.
- [8] T. S. Kumar, "Data mining based marketing decision support system using hybrid machine learning algorithm," *Journal of Artificial Intelligence and Capsule Networks*, vol. 2, no. 3, pp. 185–193, 2020.
- [9] S. Saeed, A. Shaikh, M. A. Memon and S. M. R. Naqvi, "Impact of data mining techniques to analyze health care data," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 4, pp. 682–690, 2018.
- [10] S. Dwivedi and V. S. K. Roshni, "Recommender system for big data in education," in *2017 5th National Conf. on E-Learning & E-Learning Technologies (EELTECH)*, Hyderabad, India, pp. 1–4, 2017.
- [11] M. Serik, G. Nurbekova and J. Kultan, "Big data technology in education," *Bulletin of the Karaganda University Pedagogy Series*, vol. 100, no. 4, pp. 8–15, 2020.

- [12] W. Villegas-Ch, M. R. Cañizares and X. P. Pacheco, "Improvement of an online education model with the integration of machine learning and data analysis in an lms," *Applied Sciences*, vol. 10, no. 15, pp. 5371, 2020.
- [13] B. K. Francis and S. S. Babu, "Predicting academic performance of students using a hybrid data mining approach," *Journal of Medical Systems*, vol. 43, no. 6, pp. 162, 2019.
- [14] V. Mhetre and M. Nagar, "Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA," in *2017 Int. Conf. on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 475–479, 2017.
- [15] N. M. Aslam, I. U. Khan, L. H. Alamri and R. S. Almuslim, "An improved early student's academic performance prediction using deep learning," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 12, pp. 108, 2021.
- [16] D. Shah, D. Patel, J. Adesara, P. Hingu and M. Shah, "Exploiting the capabilities of blockchain and machine learning in education," *Augmented Human Research*, vol. 6, no. 1, pp. 1, 2021.
- [17] R. Malhotra and M. Khanna, "Mining the impact of object oriented metrics for change prediction using machine learning and search-based techniques," in *2015 Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, India, pp. 228–234, 2015.
- [18] B. K. Yousafzai, M. Hayat and S. Afzal, "Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student," *Education and Information Technologies*, vol. 25, no. 6, pp. 4677–4697, 2020.
- [19] S. Shadravan, H. R. Naji and V. K. Bardsiri, "The sailfish optimizer: A novel nature-inspired metaheuristic algorithm for solving constrained engineering optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 80, pp. 20–34, 2019.
- [20] L. L. Li, Q. Shen, M. L. Tseng and S. Luo, "Power system hybrid dynamic economic emission dispatch with wind energy based on improved sailfish algorithm," *Journal of Cleaner Production*, vol. 316, pp. 128318, Sep. 2021.
- [21] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [22] Z. Han, M. M. Hossain, Y. Wang, J. Li and C. Xu, "Combustion stability monitoring through flame imaging and stacked sparse autoencoder based deep neural network," *Applied Energy*, vol. 259, pp. 114159, 2020.
- [23] M. Neshat, G. Sepidnam and M. Sargolzaei, "Swallow swarm optimization algorithm: A new method to optimization," *Neural Computing and Applications*, vol. 23, no. 2, pp. 429–454, 2013.
- [24] UCI Machine Learning Repository: Student Performance Data Set, 2019. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/student+performance>.
- [25] A. M. Mesleh and G. Kanaan, "Support vector machine text classification system: Using ant colony optimization based feature subset selection," in *2008 Int. Conf. on Computer Engineering & Systems*, Cairo, Egypt, pp. 143–148, 2008.