A Novel Hyper parameter Tuned Deep Learning Model and Optimal Feature Selection Based Student's Performance Prediction with Data Balancing

Thilagavathi.S[1], Anuratha.V[2], Elamparithi.M[3]

[1]Research Scholar, Department Of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, Email: thilagasubbu@gmail.com
[2]Associate Professor , Department Of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, Email: profanuratha@gmail.com.
[3]Associate Professor , Department Of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, Email: profelamparithi@gmail.com.

Received: 18.04.2024 Revised : 16.05.2024 Accepted: 20.05.2024

ABSTRACT
Nowadays, predicting students' performance is one of the most specific topics for learning environments, such as universities and schools, since it leads to the development of effective mechanisms that can enhance academic outcomes and avoid destruction. This paper proposes a novel hyperparameter tuned deep learning model and optimal feature selection-based student's performance prediction with data balancing. The proposed system comprises the following phases: Data collection, Dataset balancing, Preprocessing, Feature selection, and Classification. To begin, the student performance data is collected from the OULAD dataset. Then, the system uses K-Nearest Neighbor (KNN) algorithm to solve the class imbalance problem. After that, the system performs preprocessing to improve the quality of the dataset. Then, the system uses the hummingbird flight based tunicate optimization algorithm (HFTOA) to select the best features from the dataset. After that, the system uses the Hyperparameter tuned with Hard swish activation based Gated Recurrent Unit (H^2GRU) algorithm to classify the student's performance into pass or fail. The findings showed that the proposed system achieves an average result of 98.96% accuracy, 99.04% precision, 98.86% recall, 98.99% f-measure, and 98.91% AUC, which is better than the state-of-the-art methods.

Keywords: Student's Performance Prediction, Dataset balancing, Feature Selection, Deep Leaning, and Open University Learning Analytics Dataset

1. INTRODUCTION

Online education systems are becoming increasingly popular due to their flexibility and ability to reach a larger audience. However, despite these benefits, teachers find it very complex to monitor their students. In contrast to traditional educational systems, it is very hard to keep track of each student, mainly due to the large number of students per course, and the lack of face-to-face interaction. For this reason, the completion rates of online learning are notoriously lower than face-to-face learning [1]. Modern educational systems use virtual learning environments (VLEs) to support classroom activities, even in face-to-face courses [2].A VLE is defined as an educational technology that depends on web access, and it opens up the digital aspects of courses for students to complete their academic studies [3].Academic performance prediction is one of the most important aspects in online education, which is typically conducted to estimatestudents' learning performance [4, 5]. Machine learning (ML) is one of the best solutions to early predict the student's performance. Some of the commonly used ML algorithms for student's performance predictions are random forests (RF), support vector machines (SVM), logistic regression (LR), Naïve Bayes (NB), and etc. [6]. These algorithms work better, however, these traditional ML methods are typically time-consuming and labor-intensive due to the heavy preprocessing and feature engineering.

To alleviate these problems, recently, deep learning (DL) is focused by many researchers to provide successful student's performance prediction system. DL can be defined as technique to learn features directly from given different data or problems [7]. Presently, a wide range of DL algorithms is available for student academic performance classification [8].Although these predictions achieved high accuracy, the established prediction models cannot extract more representative latent features from the raw records, which affects the accuracy of the predictions [9].Also, the deep architectures are still very vulnerable to imbalanced data distributions and are affected by novel challenges such as complex data representations, the continually drifting nature of data, and learning from an extremely large number of classes [10]. To mitigate these deficiencies, this paper proposes a novel hyperparameter tuned DL model and optimal feature selection-based student's performance prediction with data balancing. The main contributions of the paper are listed as follows:

- The proposed system uses the KNN approach to balance the collected OULAD dataset, which reduces the risk of overfitting and also it improves the quality of newly created data by addressing class imbalance.
- The proposed system uses the HFTOA algorithm to select the best features from the dataset. This optimal feature selection reduces falsely selected features and it is more efficient in choosing the right features, and that the resulting model is simpler, more interpretable, and more accurate.
- The proposed system uses the H^2GRU algorithm to classify the student's performance into pass or fail, in which the hyperparameter is optimally chosen by HFTOA algorithm and the gradient vanishing problem is solved by Hard swish activation function.

The rest part of the paper is structured as follows: Section 2 presents the recent works related to the proposed work. Section 3 presents the brief descriptions of the proposed methodology. Section 4 covers the experimental analysis and discussions. Finally, theconclusions and future lines of research are outlined in Section 5.

2. LITERATURE SURVEY

This section surveys the recent works related to student's performance perdiction system.Hsing-Chung Chen et al. [11]developed student performance early prediction in VLE based on a deep explainable artificial intelligence. First, the system used a combination of convolutional neural network (CNN) and long short-term memory (LSTM) to extract the spatiotemporal features. Finally, the DL model was explained by visualizing and analyzing typical predictions, students' activities' maps, and feature importance. The data set was recorded from a 16-week, fully online "Digital Transformation" course at Gadjah Mada University and the system showed that the system achieved best f1-score of 0.91%.Heyul Chavez et al. [12] suggested an artificial neural network (ANN) model to predict student performance based on non-personal information. First, the system applied preprocessing to prepare the data that will be used for training. Second, the system sorted the students into pass or fail categories. Third, the system splinted the data into training and testing datasets. Finally, the system trains the model based on the corresponding dataset to predict the student performance. For model training, the system used information regarding 32,000 students collected from The Open University of the United Kingdom and the system achieved 93.81% accuracy.

Xiaoxia Jiao [13] proffered a factorization deep product neural network (FDPN) for student physical performance prediction. In FDPN network, the embedded layer was first used to convert high-dimensional feature maps into low dimensional feature vector. Then, the concatenation layer composed of three parts such as factorization machine, deep neural network, and product neural network (PNN) to express the first-order and second-order features. Finally, the prediction layer was used to predict the student performance. The experimentation was done on the OULAD dataset and the results showed that the system achieved best accuracy of 0.87%.Samina Sarwat et al. [14] presented to predict students' academic performance based on conditional generative adversarial network and deep SVM. To begin, the system collected 786 students' information from the Internet. Then, the system used an improved conditional generative adversarial network (CGAN) in combination with a deep-layer-based SVM to predict students' performance through school and home tutoring. Performance analysis showed that the system outperformed all models, with 98.2% accuracy and 97.1% specificity.Sujan Poudyal et al. [15] developed hybrid two-dimensional convolutional neural network (2D CNN) model for prediction of student academic performance. Initially, the system performed preprocessing to improve the quality of the dataset. Then, the system transformed data into a 2D format suitable for the 2D CNN architecture. Finally, the system used 2D CNN to predict the student's performance. The system used OULAD dataset to verify the effectiveness of the system and the system achieved best accuracy of 88%.

2.1 Problem Statement

The above-mentioned surveys the recent works related to student's performance prediction in VLE using DL model. All the above shows the better prediction results, but it has some deficiencies to predict the student's performance. Dataset balancing and feature selection is important to build efficient model, but none of the works focussed that. Balancing a dataset makes training a model easier because it helps prevent the model from becoming biased towards one class. And also, theOULAD data set contained many features, and not all variables were conducive to student's performance prediction. Excessive redundancy and irrelevant variables in the dataset may hinder the predictive performance of the model. But the above-mentioned works directly uses the features from the dataset for prediction. So, it decreases the prediction performance and increases the overfitting problem. Also, the above-mentioned works used DL algorithm to classify the student's performance, which shows better efficiency,but the random hyperparameter (weights, learning rate, etc.) of the network is a critical issue, because the random hyperparameter increases the processing time along with the computational complexity. This paper uses efficient feature selection with novel optimization method and DL to classify the student's performance with optimal hyperparameter tuned approach. These are briefly explained in the following section.

3. PROPOSED METHODOLOGY

Figure 1 shows the workflow diagram of the proposed novel hyperparameter tuned deep learning model and optimal feature selection-based student's performance prediction with data balancing. It mainly comprises five phases such as data collection, dataset balancing, preprocessing, feature selection, and classification. These phases are briefly explained in the following subsections.

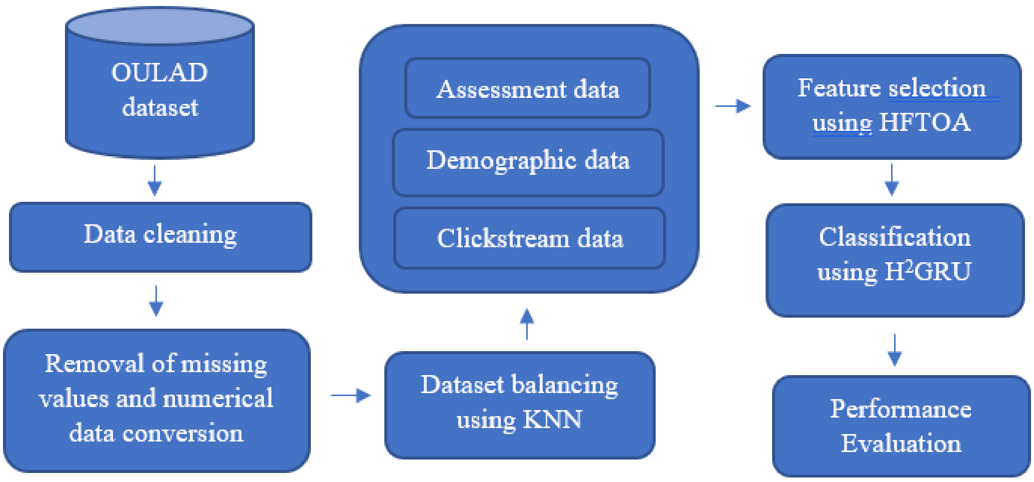

Figure 1. Workflow of the proposed methodology

3.1 Data Collection

The collection of educational data is the first step in reflecting the value of educational data mining. The student data is collected from the publicly available dataset of OULAD, which is access through https://analyse.kmi.open.ac.uk/open_dataset. It is considered to be one of the most comprehensive international open datasets in terms of e-learning data diversity, including student demographic data and interaction data between students and VLE. The role of Open University in developing this dataset is to support research in the field of learning analysis by collecting and analyzing learner data to provide personalized guidance and optimize learning resources. The dataset contains 7 course modules, 22 courses, e-learning behavior data and learning performance data of 32,593 students.Each constitutes comprises student assessment information, demographics information, course information, student's interaction information, mutual information, and assessment performance. Also, it contains some important attributes such as gender, index of multiple deprivation band, highest education level, age, region, and disability. The OULAD dataset comprises four classes such as distinction, pass, fail, and withdrawal. Our proposed approach can train and validate history courses' information and the proposed work developed a new model to identify participants' outcomes based on their demographics, assessment stream, and the clickstream. It combined 'distinction' labels and 'pass' labels into 'pass' labels and ignored the 'withdrawal' instances.

3.2 Dataset Balancing

The dataset used in this study is the imbalanced dataset.Imbalanced data refers to those types of datasets where the target class has an uneven distribution of observations. The overfitting problem can happen in the case of using the original imbalanced dataset as the model can overfit the majority class.To handle the imbalance dataset problem efficiently a K-Nearest neighbor (KNN) technique is used, which helps to prevent the model from becoming biased towards one class. KNN algorithm is a method for classifying objects based on closest training examples in the feature space. KNN is a type of instance-based learning, where the function is only approximated locally and all computation is deferred until classification.The algorithm remains impartial to the size of the class, thus avoiding any bias towards a particular size.The KNN algorithm basically requires no training, just simply stores the feature vectors and class labels of the training data. Given a new instance for testing, its class is predicted as majority class label from its k-nearest neighbors. The process involved in the proposed KNN is explained as follows:

- To begin, specify a positive integer $\vec{k}$.
- Then, calculate the Euclidean distance between pairs of cases.
- If $\vec{k}=1$ is chosen, classify the case into the group of its nearest neighbour.
- If $\left(\vec{k}\geq 1\right)$, classify the case into the group of the majority k-closest neighbours

KNN assigns a new data point to the class that has the majority of its nearest neighbours. By adjusting the value of $\vec{k}$, the number of nearest neighbours considered.

3.3 Preprocessing

Preprocessing the dataset is an extremely important task especially given the nature of OULAD.Student assessment data preprocessing involves two steps to prepare the data for analysis. The sequence of steps comprises of data cleaning and transformation. Data cleaning includes eliminating or revising any blunders, irregularities, or missing qualities in the Data. This step is essential to guarantee that the Data is precise and complete. Data transformation is the process of transforming data from one representation to another. When the original data type does not meet the requirements of the model input, data transformation is required, including data size transformation and type transformation. In our model, to calculate the similarity of the data, all input data should be numerical.

3.4 Feature Selection

In order to decrease the dimensionality of the features and increase the generalizability, operational effectiveness, and interpretability of the model, feature selection chooses the pertinent features that contribute relatively highly to the model's training. This paper uses a hummingbird flight based tunicate optimization algorithm (HFTOA). The TOA is a meta-heuristic algorithm inspired by the swarm foraging behavior of tunicate animals in the ocean, which simulates the jet propulsion and swarm behavior of tunicate animals in the foraging process.Jet propulsion is the propulsion of something in one direction, produced by ejecting a jet of fluid in the obverse direction. Swarm intelligence deals with natural and artificial systems. They depend on the collective behavior between individuals of the swarm.So the proposed system uses Hummingbird flight strategy (HFS) is constructed to replace the original random factor to coordinate the algorithm's local exploitation and global exploration performance, which effectively improves the ability of the algorithm to escape extreme values and fast convergence. Thus, this improvisation included in the conventional TOA is termed as HFTOA. The steps involved in the HFTOA is explained as follows:

The HFTOA starts with a population of randomly generated tunicates based on the design variables' allowable boundaries, as shown in the equation below:

$$\dddot{T}_L=\dddot{T}_L^{\min}+Rand\times\left(\dddot{T}_L^{\max}-\dddot{T}_L^{\min}\right) \quad (1)$$

Where, $\dddot{T}_L$ indicates the position of each tunicate, $Rand$ refers a random number within range [0, 1], $\dddot{T}_L^{\min}$ and $\dddot{T}_L^{\max}$ signifies design variables' lower and upper bounds, respectively. After that, compute the fitness of each individual. Classification accuracy is the important criteria to design a fitness function. The Equation (2) maximizes the accuracy in every iteration with certain selected features, which alternatively decreases the error rate.It is mathematically expressed as follows:

$$Fitness=\max\left(Accuracy\right) \quad (2)$$

$$Accuracy=\frac{\hat{P}_{TP}+\hat{P}_{TN}}{\hat{P}_{TS}} \quad (3)$$

Where, $\hat{P}_{TP}$ refers the true positive, $\hat{P}_{TN}$ indicates the true negative, and $\hat{P}_{TS}$ denotes the total number of samples. Then, the tunicatesadjust their location during the iterations by the following formula:

$$\dddot{T}_L(\vec{y}+1)=\frac{\dddot{T}_L(\vec{y})+\dddot{T}_L(\vec{y}+1)}{2+\breve{\alpha}_1} \quad (4)$$

Where, $\breve{\alpha}_1$ refers a random number within range [0, 1], $\vec{y}$ refers a current iteration, and $\dddot{T}_L(\vec{y})$ denotes to the updated position of the tunicate with respect to the position of the food source based on equation (5).

$$\dddot{T}_L(\vec{y})=\begin{cases} F_S+\vec{\kappa}\times\left|F_S-Rand\times\dddot{T}_L\right|, & if \;\; Rand\geq 0.5 \\ F_S-\vec{\kappa}\times\left|F_S-Rand\times\dddot{T}_L\right|, & if \;\; Rand<0.5 \end{cases} \quad (5)$$

Where, F_S refers the food source, which is represented by the population's optimal tunicate position and $\vec{\kappa}$ refers the HFS, which enhances the search ability. HFS have main concept of it do not fly in large groups in confined spaces and simultaneous position changes in all directions are not realistic. It is mathematically defined as follows:

$$\vec{\kappa}=\left|2.Rand\,.\bar{\rho}-\bar{\rho}\right| \quad (6)$$

$$\bar{\rho}=\exp\left(2-2\frac{\vec{y}}{\vec{Y}}\right) \quad (7)$$

Where, $\vec{y}$ refers a current iteration and $\vec{Y}$ indicates the maximum number of iterations.Until the maximum iterations satisfied the process will be repeated. Once the optimized features were getting the algorithm will terminated.

3.5 Classification

At last, the classification is performed by using the Hyperparameter tuned with Hard swish activation based Gated Recurrent Unit (H^2GRU) from the optimally selected features, which classifies the student's performance into two classes such as pass or fail.GRU is a class of recurrent neural network designed to increase the speed performance when massive numbers of data are concerned. The core principle of a GRU is to minimize network process time complexity and preserve long-term dependencies in text sequences.This approach consists of only two gates: the update gate and the reset gate. The update gate regulates the flow of data in time steps. The reset gate determines how much previous data is transferred and how much is lost.No protected cell state is used; hence the gates are applied directly to the hidden state. This difference makes the GRU more computationally efficient than the other algorithms, however, the computations in the classical GRU are associated with random weights to compute the output. This randomly chosen weight values increases the computational complexity and takes more iterationsduring training, which requires high amount of time. So the proposed system uses HFTOA, which optimally chosen weights for the corresponding input and reduce the complexity of the system. In addition, the classical GRU uses sigmoid and tanh activation functions to solve the gradient vanishing problem occur in the network. These activation function works better, nevertheless, these two activation functions not converged quickly and thus takes much more time when compared to models trained on the other activation functions. Hence, the proposed system uses the hard swish activation function instead of sigmoid and tanh activation function to avoid the overfitting issues and solves the gradient vanishing problem. Thus, the optimal hyperparameter tuning and new activation function included in the classical GRU is termed as H^2GRU. The structure of classical GRU is shown in figure 2.

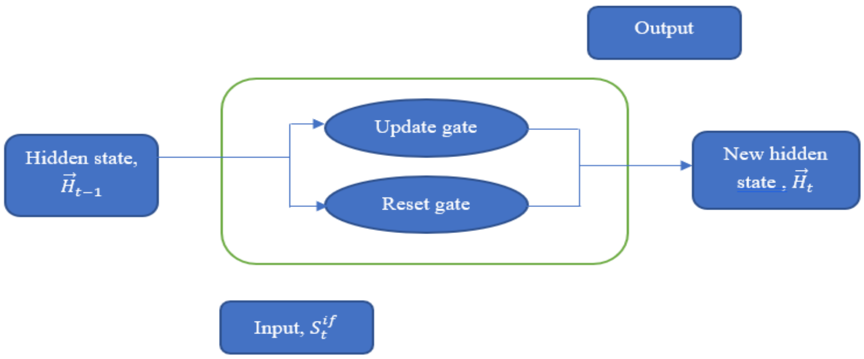

Figure 2. Structure of GRU

Figure 2 initially takes the input as the optimally selected features S_t^{if} with time step t, then it sends to the GRU. It mainly comprises two gates such as update gate $(\vec{u})$ and reset gate $(\vec{r})$. The following calculation can be used to express each relationship in the GRU:

$$\vec{u}_t = \mu^* \left(\ddot{\omega}_{\vec{u}} \left[\vec{H}_{t-1}, S_t^{if}\right]\right) \tag{8}$$

$$\vec{r}_t = \mu^* \left(\ddot{\omega}_{\vec{r}} \left[\vec{H}_{t-1}, S_t^{if}\right]\right) \tag{9}$$

$$\vec{c}_t = \mu^* \left(\ddot{\omega}_{S^{if}} \left[\vec{r}_t, \vec{H}_{t-1}, S_t^{if}\right]\right) \tag{10}$$

$$\vec{H}_t = (1-\vec{u}_t)\,\vec{H}_{t-1} + \vec{u}_t\,\vec{c}_t \tag{11}$$

Where, $\ddot{\omega}$ refersto the weight matrix of update gate, reset gate, and corresponding input, respectively, which is optimally chosen by HFTOA algorithm, $\vec{c}_t$ indicates the candidate hidden layer, $\vec{H}_t$ indicates the forward hidden state output, μ^* indicates the Hard swish activation function.Hard swish is a new activation function, which replaces the sigmoid and tanh activation function with a piecewise linear function that is much easier to compute. It is mathematically formulated as follows:

$$\mu^* = 2 * S^{if} * \max\left(0, \min\left(1, \left(\hat{\zeta}\; S^{if} * 0.2 + 0.5\right)\right)\right) \tag{12}$$

Where, $\hat{\zeta}$ indicates the trainable parameter. Finally, the output from the GRU is fed into the SoftMax classifier to predict the student's performance, where, 0 indicates the failure and 1 indicates the pass, respectively.

4. RESULTS AND DISCUSSION

This section covers the performance of the proposed novel hyperparameter tuned deep learning model and optimal feature selection-based student's performance prediction with data balancing is analyzed with the existing approaches in terms of some basic evaluation metrics. The performance of the system is implemented in the working platform of PYTHON with windows 10 OS.

4.1 Performance Analysis

In this section, the outcomes of the proposedH^{2}GRU model are investigated against the existing GRU, Recurrent Neural Network (RNN), Multi-Layer Perceptron (MLP), and Artificial neural Network (ANN) methods. This investigation is done based on accuracy, precision, recall, f-measure, and area under curve (AUC) metrics. This analysis could be shown in the following figures and table.

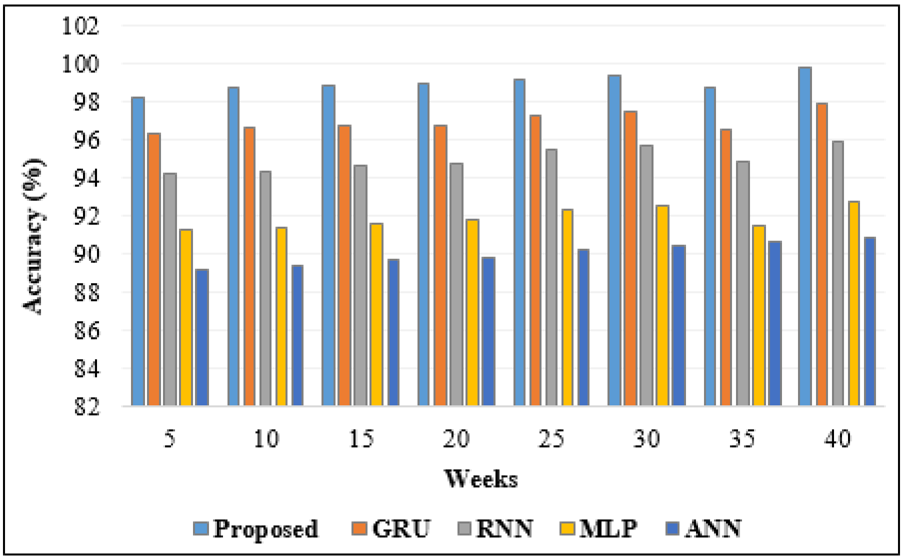

Figure 3. Accuracy analysis

Figure 3 illustrates the efficiency of the proposed method and the existing methods in terms of accuracy metric for week 5 to 40. The classification accuracy is defined as the ratio of true positive to true negative events to the total number of events obtained by the classification method.For week 5, the existing methods achieves accuracy of 96.34%, 94.21%, 91.23%, and 89.12%, which is lower hen compared to the proposed one, because, the proposed one achieves higher accuracy of 98.23%. When, increasing the weeks from 5 to 40, it maximally reaches 99.78% accuracy. This clearly shows that the proposed one achieves superior outcomes than the existing methods.

Table 1. Precision, Recall, and F-measure analysis

Metrics	Weeks	Proposed	GRU	RNN	MLP	ANN
Precision	5	98.34	9.42	94.29	91.32	89.23
	10	98.83	96.74	94.42	91.41	89.41
	15	98.94	96.82	94.72	91.63	89.73
	20	98.98	96.85	94.79	91.83	89.95
	25	99.23	97.34	95.52	92.42	90.32
	30	99.41	97.51	95.76	92.65	90.56
	35	98.74	96.65	94.94	91.53	90.73
	40	99.82	97.93	95.95	92.86	90.97
Recall	5	98.12	96.26	94.15	91.14	89.03
	10	98.65	96.54	94.24	91.25	89.25
	15	98.74	96.65	94.56	91.46	89.52
	20	98.87	96.69	94.67	91.67	89.78
	25	99.05	97.14	95.38	92.23	90.16
	30	99.23	97.36	95.56	92.47	90.35
	35	98.56	96.47	95.79	91.38	90.59
	40	99.65	97.75	95.81	92.71	90.79
F-measure	5	98.26	96.37	94.26	91.26	89.16
	10	98.79	96.69	94.39	91.39	89.39
	15	98.87	96.78	94.68	91.59	89.71
	20	98.96	96.81	94.75	91.81	89.86
	25	99.15	97.28	95.49	92.38	90.27
	30	99.9	97.46	95.71	92.61	90.49
	35	98.71	96.59	94.92	91.49	90.71
	40	99.84	97.89	95.94	92.82	90.94

Next, table 1 indicates the performance of the proposed and existing methods in terms of precision, recall, and f-measure metrics. The ratio of true positives to all other true positives and false positives is known as precision. The ratio of true positives to all other true positives and false negatives is known as recall. The harmonic mean of precision and recall is known as the F-Measure. Herein also, the proposed one offer better outcomes of 99.42% precision, 99.65% recall, and 99.84% f-measure for 40 weeks, which is better outcomes when compared to the existing approaches. Similarly, when concerning the other weeks, the proposed one achieves better outcomes than the existing methods. Thus, it concludes that the proposed one achieves high-level outcome than the existing methods. Figure 4demonstrates the outcomes of the proposed method and the existing methods in terms of AUC metric. The AUC is the area under the Receiver Operating Characteristics (ROC) curve that indicates the trade-off between the true positive rate (TPR) and false positive rate (FAR).
In this figure, the proposed one achieves better AUC of 98.21%, which is 1.9%, 4.04%, 7.02%, and 9.13% better than the existing methods for week 5. Similarly, for the remaining week 10 to 40, the proposed one achieves better efficiency than the existing methods. Thus, the overall analysisshows that the proposed one achieves superior outcomes than the existing methods for all the weeks. The reason is that the proposed system initially balances the dataset by using the KNN approach, which improves the prediction rate and avoids the misclassification results. Next, the proposed system performed preprocessing to improve the quality of the dataset and achieved clear results. Then, the system performed feature selection using theHFTOAmethod, which is used to make the process more accurate and it also increases the prediction power of the algorithms by selecting the most critical variables and eliminating the redundant and irrelevant ones. Finally, the proposed system uses the H^2GRU algorithm for student's performance prediction, in which the hyperparameter is optimally chosen by the HFTOA approach, which boosts the classification performance and achieves optimal accuracy. Thus, these incorporations in the proposed work boosts the outcomes of the proposed work.

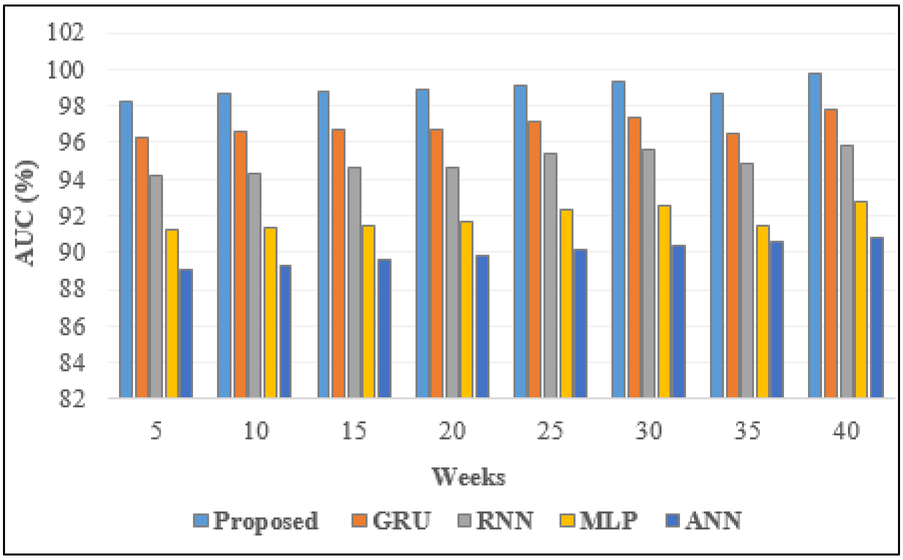

Figure 4. AUC analysis

5. CONCLUSION

This study uses a novel hyperparameter tuned deep learning model and optimal feature selection-based student's performance prediction with data balancing. To test and evaluate the performance of the system, it used OULAD dataset. In the experimental analysis, the performance of the proposed method is weighted against the existing methods such as GRU, RNN, MLP, and ANN methods. This evaluation is performed with the help of accuracy, precision, recall, f-measure, and AUC methods. The evaluation is made up of 5 to 40 weeks data. For 5 weeks, the proposed one achieves 98.23% accuracy, 98.34% precision, 98.12% recall, 98.26% f-measure, and 98.21% AUC, which is better than the existing methods. Similarly, for considering the other weeks, the proposed one achieves high-level outcomes than the existing methods. Thus, the overall experimental shows that the proposed one not only have satisfactory predictive performance, but also have the lowest misprediction cost. In future, this work will be prolonged to perform dimensionality reduction to reduce the dimension of the dataset and achieves full percentage results.

REFERENCES

[1] Gámez-Granados, J. C., Esteban, A., Rodriguez-Lozano, F. J., & Zafra, A. (2023). An algorithm based on fuzzy ordinal classification to predict students' academic performance. Applied Intelligence, 53(22), 27537-27559.

[2] Queiroga, E. M., Enríquez, C. R., Cechinel, C., Casas, A. P., Paragarino, V. R., Bencke, L. R., & Ramos, V. F. C. (2021). Using virtual learning environment data for the development of institutional educational policies. Applied Sciences, 11(15), 6811.

[3] Bagunaid, W., Chilamkurti, N., & Veeraraghavan, P. (2022). AISAR: Artificial Intelligence-Based Student Assessment and Recommendation System for E-Learning in Big Data. Sustainability, 14(17), 10551.

[4] Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial intelligence-enabled prediction model of student academic performance in online engineering education. Artificial Intelligence Review, 55(8), 6321-6344.

[5] Alboaneen, D., Almelihi, M., Alsubaie, R., Alghamdi, R., Alshehri, L., & Alharthi, R. (2022). Development of a web-based prediction system for students' academic performance. Data, 7(2), 21.

[6] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 11.

[7] Li, S., & Liu, T. (2021). Performance prediction for higher education students using deep learning. Complexity, 2021, 1-10.

[8] Abuzinadah, N., Umer, M., Ishaq, A., Al Hejaili, A., Alsubai, S., Eshmawi, A. A., ... & Ashraf, I. (2023). Role of convolutional features and machine learning for predicting student academic performance from MOODLE data. Plos one, 18(11), e0293061.

[9] Wen, X., & Juan, H. (2023). Early prediction of students' performance using a deep neural network based on online learning activity sequence. Applied Sciences, 13(15), 8933.

[10] Dablain, D., Krawczyk, B., & Chawla, N. V. (2022). DeepSMOTE: Fusing deep learning and SMOTE for imbalanced data. IEEE Transactions on Neural Networks and Learning Systems.
[11] Chen, H. C., Prasetyo, E., Tseng, S. S., Putra, K. T., Kusumawardani, S. S., & Weng, C. E. (2022). Week-Wise Student Performance Early Prediction in Virtual Learning Environment Using a Deep Explainable Artificial Intelligence. Applied Sciences, 12(4), 1885.
[12] Chen, H. C., Prasetyo, E., Tseng, S. S., Putra, K. T., Kusumawardani, S. S., & Weng, C. E. (2022). Week-Wise Student Performance Early Prediction in Virtual Learning Environment Using a Deep Explainable Artificial Intelligence. Applied Sciences, 12(4), 1885.
[13] Jiao, X. (2022). A Factorization Deep Product Neural Network for Student Physical Performance Prediction. Computational Intelligence and Neuroscience, 2022.
[14] Sarwat, S., Ullah, N., Sadiq, S., Saleem, R., Umer, M., Eshmawi, A. A., ... & Ashraf, I. (2022). Predicting Students' Academic Performance with Conditional Generative Adversarial Network and Deep SVM. Sensors, 22(13), 4834.
[15] Poudyal, S., Mohammadi-Aragh, M. J., & Ball, J. E. (2022). Prediction of student academic performance using a hybrid 2D CNN model. Electronics, 11(7), 1005.