Big Data in Education: Students at Risk as a Case Study

Ahmed B. Altamimi
Department of computer Engineering, College of Computer Science and Engineering, University of Ha'il, Saudi Arabia
altamimi.a@uoh.edu.sa (corresponding author)

Received: 11 July 2023 | Revised: 10 August 2023 and 16 August 2023 | Accepted: 21 August 2023

 DOI: https://doi.org/10.48084/etasr.6190

ABSTRACT

This paper analyzes various machine learning algorithms to predict student failure in a specific educational dataset and a specific environment. The paper handles the prediction of student failure given the students' grades, course difficulty level, and GPA, differing from most of the provided studies in the literature, where focus is given to the surrounding environment. The main aim is to early detect students at risk of academic underperformance and implement specific interventions to enhance their academic outcomes. A diverse set of eleven Machine Learning (ML) algorithms was used to analyze the dataset. The data went through preprocessing, and features were engineered to effectively capture essential information that may impact students' academic performance. A meticulous process for model selection and evaluation was utilized to compare the algorithms' performance with regard to metrics such as accuracy, precision, recall, F-score, specificity, and balanced accuracy. Our results demonstrate significant variability in the performance of the different algorithms, with Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) showing the highest overall performance, followed closely by Gradient Boosting Classifier (GBC), Neuro-Fuzzy, and Random Forest (RF). The other algorithms exhibit varying performance levels, with the Recurrent Neural Networks (RNNs) showing the weakest results in recall and F-score. Educational institutions can use the insight gained from this study to make data-driven decisions and design targeted interventions to help students at risk succeed academically. Furthermore, the methodology presented in this paper can be generalized and applied to other educational datasets for similar predictive purposes.

Keywords-machine learning algorithms; big data; accuracy; F-score; precision

I. INTRODUCTION

The concept of big data involves collecting a significant amount of data within a specific domain [1-19], spanning various areas like customer data in stores, digital library publications, and more. Big data are categorized into big media data [1] and big business data. Both industry and academia have shown a keen interest in exploring big data methodologies and their impact on different sectors due to their potential contribution to organizational profitability and productivity. Big data exhibits several defining characteristics, often referred to as the 5 Vs: volume, velocity, value, veracity, and variety [2]. Volume refers to the sheer amount of data, which can pose challenges in processing and storage. Veracity encompasses data quality, distinguishing between data in motion and data at rest. Value highlights data's significance within a system. Veracity pertains to data quality, ensuring it's free from noise. Variety pertains to data's structural diversity. For instance, humidity sensor data differ from alarm system data, which must be considered in the analysis phase.

The applications of big data are diverse. Call centers [3] gather extensive data that aligns with the 5 Vs, enhancing customer experiences and optimizing product recommendations. Information technology companies utilize logs to predict hacking attempts and network resource demands [2]. Scientists employ big data to identify valuable publications [3]. Each application involves three main stages: gathering, processing, and mining [4]. Data gathering utilizes stored data, which are then formatted for streamlined mining. Data mining algorithms like decision trees [1] and Bayesian methods [1] unravel relationships between raw data and external outcomes. Data analysis in big data presents a challenge, but it is not the only one. Various challenges and issues must be addressed in big data, such as managing large data volumes, ensuring data privacy, storing vast amounts of data, visualizing data, scheduling jobs, and maintaining fault tolerance. These challenges are distinct and require specific attention.

This paper focuses on a significant application within the education sector, specifically addressing student performance prediction. The collection of large amounts of student grade data at the end of a semester presents an opportunity for utilizing big data to enhance students' final course marks. The study aims to provide early warnings for students who might be at risk of failing a course based on their early assessment results. Predicting student failure and performance is a vital aspect of education research. Employing predictive analytics techniques enables the identification of students prone to academic underachievement, allowing timely interventions and support. This has far-reaching benefits, including improving education quality, student retention, dropout rates, and overall academic achievements. In addition to the mentioned points, the investigation into student performance prediction offers implications for various aspects of the education system [19]:

- Personalized learning: Accurate performance prediction facilitates tailored learning approaches, adapting teaching methods and materials to individual student needs. This can enhance engagement, motivation, and academic outcomes.
- Early career guidance: Performance prediction aids educators and counselors in identifying students' strengths and interests, guiding them towards suitable career paths and education decisions.
- Parental involvement: Reliable predictions encourage parental engagement in supporting their child's education, fostering collaboration between parents, students, and educators.
- Policy development and implementation: Predictive analytics informs policy-makers enabling targeted interventions to enhance education quality and resolve challenges.
- Evaluation and accountability: Performance prediction helps institutions assess program effectiveness, leading to continuous improvement and refined educational outcomes.
- Research opportunities: Exploring student performance prediction contributes to educational research in teaching methods, curriculum development, and educational technology, improving educational practices.

Investigating student failure and performance prediction extends beyond identifying students at-risk. These areas offer implications for individuals, institutions, and the society, including addressing inequality, enabling social mobility, informing policy decisions, and fostering interdisciplinary research and innovation. Accurate predictive models contribute to inclusive, equitable, and effective educational systems.

II. RELATED WORK

Educational institutions have increasingly focused on predicting student failures, leading to a surge in research using Machine Learning (ML) and data mining techniques to analyze contributing factors. The objective is to identify students at risk of failure early and intervene promptly for improved success. Various methodologies have been applied, including statistical methods, ML, and fuzzy logic. Notable studies in this field are highlighted below.

Authors in [18] introduced an Artificial Neural Network (ANN) algorithm for predicting at-risk students in self-paced online courses. Their model utilized demographic data, course enrollment, and early interaction data. Achieving an accuracy of 84.57%, the ANN outperformed baseline models, showing effectiveness in identifying and avoiding false positives. Key predictors included age, education level, and early interaction metrics. At the same time, authors in [20] investigated emotional wellness and engagement as indicators of academic achievement in e-learning. ML revealed their significance, offering educators insights to enhance outcomes through tailored strategies. Authors in [21] explored student-athletes' causal attributions for sport and academic achievement's impact on dropout and GPA. Stable attributions correlated with better performance and lower dropout rates. Authors in [22] focused on predicting student failure in higher education using ensemble learning algorithms. Their approach achieved 89% accuracy and proposed personalized learning paths to boost academic success. Authors in [23] developed an early detection method for identifying at-risk students based on learning performance and behavior indicators, achieving 79% accuracy. Authors in [24] employed educational data mining and ML techniques to predict academic performance, with Decision Tree (DT) and ANN models achieving 82.38% and 84.57% accuracy, respectively. Authors in [25] introduced a Bayesian network model for Massive Open Online Course (MOOC) student success prediction, outperforming alternative models. Authors in [26] created a web-based platform to forecast academic achievements, using Logistic Regression (LR) and DT algorithms to attain an accuracy rate of 85%. Authors in [27] introduced the Enhanced Binary Genetic Algorithm (EBGA) for feature selection in predicting student performance, demonstrating superiority over other methods. Authors in [28] utilized supervised learning techniques, showing ANN's high accuracy in predicting student performance. Authors in [29] proposed an incremental ensemble for distance education student performance prediction, while authors in [30] focused on predicting high school dropouts using data mining techniques. Authors in [31] presented temporal models for predicting MOOC student attrition, highlighting the importance of considering temporal dynamics for accurate predictions. Table I summarizes the related works.

These studies demonstrate the diversity of approaches that have been used to predict student failure. Selection of the appropriate method depends on various factors, including the size and nature of the dataset, the availability of resources, and the desired interpretability of the results. Overall, the findings from these studies indicate that ML can offer valuable insights into predicting student failure and help educational institutions develop targeted interventions for at-risk students. However, the described papers in the related work either focus on specific or some ML algorithms or certain datasets. Also, the used datasets consider different factors around students' quizzes and final exams while they ignore other factors that affect the performance such as attendance, assignments, topics difficulty, GPA, and student level. In Saudi Arabia, the education environment is different from other countries: students do not have that many distraction factors that affect their studies. Therefore, in our study, we developed different algorithms to predict student failure using a synthetic dataset of 9,000 student records. The dataset included attendance, assignments, quiz scores, midterm and final exam scores, GPA, course difficulty levels, and students' educational levels. A voting system is used to decide on the student's performance.

TABLE I. RELATED WORK SUMMARY

Ref.	Purpose	Algorithms/ Techniques	Results/ Key Findings
[18]	Initial detection of at-risk students in self-paced education	ANNs	Effective in predicting at-risk learners
[20]	Predict the academic performance of students based on their emotional health	Long Short Term Memory (LSTM) network, Random Forest (RF), and Gradient Boosting Classifier (GBC)	96% prediction accuracy
[21]	Relation between sport dropout, GPA, and causal attributions	Latent profile analysis	Reasonable performance
[22]	Predict student failure and enable customized educational paths	Ensemble learning algorithms	Improved prediction of student failure
[23]	Early student-at-risk detection	k-Nearest Neighbors (kNN) classifier and Linear Discriminant Analysis (LDA)	Good detection performance
[24]	Predicting the academic performance of students	RF, nearest neighbor, Support Vector Machines (SMVs), LR, Naïve Bayes, kNN	Classification accuracy of 70–75%
[25]	Predicting MOOC performance and improving performance	Bayesian Network	Improved MOOC performance prediction
[26]	Create an online academic performance prediction system	SVM, RF, kNN, ANN, and LR	Performance was not the target of the paper. They used mean square error to measure the performance
[27]	Improved feature selection to predict student performance	Enhanced Binary Genetic Algorithm (EBGA)	Improved prediction accuracy
[28]	Predict student academic performance using supervised learning	kNN, SVM, DT, NB, ANNs	Highest accuracy with ANNs (90%+)
[29]	Predict students' performance in distance education	Combinational incremental ensemble of classifiers	Outperforms single classifiers
[30]	Early dropout prediction using data mining	DT, ANNs, SVM	Effective in reducing dropout rates
[31]	Predict student dropout in MOOCs	Temporal models	Outperforms static models

III. METHODOLOGY

This section will elucidate the methodology of ML modeling to predict the failure of students in MOOCs. The section is divided into three key subsections: data collection, ML modeling, and model evaluation. Within the data handling subsection, this section will cover the selection of relevant features and data preprocessing to enhance the model's accuracy. The ML modeling subsection will explore various algorithms, including fuzzy logic, ANNs, neuro-fuzzy, LR, DTs, Recurrent Neural Networks (RNNs), kNN, Convolution Neural Networks (CNNs), and RF to determine the most suitable approach for predicting student failure. To assess the model's performance, the model evaluation subsection provides an extensive overview of metrics, such as accuracy, precision, recall, F1 score, specificity, and balanced accuracy. Additionally, cross-validation and hyperparameter tuning will be discussed to ensure that the model is not overfitting and is optimized for prediction accuracy. Figure 1 shows the different steps utilized in this paper.

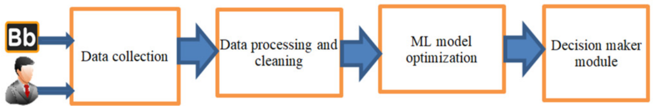

Fig. 1. Student performance prediction proposed model.

A. Data Collection

Data collection refers to acquiring data from their source and preparing them for use in a ML model. Extraction refers to the process of retrieving information from a data source using data mining techniques. First, the data are extracted from two main sources: Blackboard and the instructors. The blackboard keeps all the student's records, therefore, most data are extracted from it. However, some missing data are gained from the course instructors. Students' GPAs are automatically extracted from another system that keeps students' official records. The dataset used for this study is obtained from the blackboard courses in different years and for different colleges taught at the University of Hail, Saudi Arabia. The provided dataset includes courses from the College of Computer Science and Engineering and other colleges from 2015 to 2020. The data follow the students from their entrance to the college to the suggested time of their graduation. Since the study is related to the student's failure prediction, the focus was on the students who could not graduate on time. Therefore, almost 50% of the dataset under study has failed students' records. The data were stored in CSV format, which can be accessed on a local machine. The experiment was conducted in a Jupyter notebook, and the data were processed using Python. The total number of records is 9,000. The dataset attributes are shown in Table II. In order to carry out supervised experiments using this dataset, the target column selected to determine student performance was the Result score (pass/fail) of the course, as it is commonly preferred. The Result score is calculated based on a combination of the Midterm exam (25%), Classwork (2 quizzes, 20% weightage), Assignments (10% weightage), Attendance (10% weightage), and the Final Exam (35% weightage). The GPA score was classified into two categories: (1) Pass, for scores greater than or equal to 60, and (2) Fail, for scores less than 60.

B. Data Preprocessing and Data Cleaning

Data preprocessing involves extracting the relevant information from the target dataset that is required for the ML model. Records with inconsistent values were removed from the data set, such as samples with high grades but very low attendance time. This restriction is applied because the students are allowed only 25% absences from the semester lectures. As a result, the data of 9,000 students were evaluated in this study.

TABLE II. DATASET ATTRIBUTES AND THEIR RANGES

Attribute	Description	Type	Value
Student_ID	Student unique identifier	Numeric	1-9000
College	Student collage	String	NA
Department	Student department	String	NA
Program	Program of study within the department	String	NA
GPA	Student GPA	Numeric	1-4
Credit	Course number of credit hours	Categorical	1-5
Teacher_ID	Teacher identifier	String	NA
Level	1-8, represents the course semester	Categorical	1-4
Difficulty	The difficulty of the course		1-5
Assignments	Classwork grades, including all assignments	Numeric	10
Quiz_1	The first Quiz	Numeric	10
Quiz_2	The second Quiz	Numeric	10
Attendance	Students attendance	Numeric	10
Midterm	Course midterm grade	Numeric	25
Final	Course final grade	Numeric	35
Results	Pass or fail	Categorical	0-1

This study focuses on identifying the features that are indicative of student failure and using them to predict such events. The extracted dataset consists of time series data with no primary keys. However, to map the data to the ML model, it is necessary to group the data based on student identity, which involves assigning each student as the key. Previous research on Educational Data Mining (EDM) datasets followed a similar approach and performed grouping as a preprocessing step to structure the raw data. In this study, a similar grouping was performed. Once the data are grouped based on student identity, the learning progression data for each student can be extracted from the target dataset and stored as a preprocessed data table. Learning progression refers to the process of acquiring and mastering knowledge, skills, or competencies over a period of time. Individuals go through a sequence of educational and developmental stages as they gain a deeper understanding and mastery of a subject or skill. Learning progression typically involves the acquisition of increasingly complex concepts and skills, building on what was previously learned, and applying it to new and more challenging contexts. Learning progression can be measured and evaluated using various methods, such as classwork, midterm, final, or results, to track a learner's progress and identify areas of strength and weakness. In order to identify shifts in a student's learning progression, we analyze the topics that the student has mastered on a semester basis. This is achieved by computing the discrepancy between the student's progress on a given semester and their cumulative progress up until that the last semester, resulting in a list of topics the student has mastered in each semester. These fluctuations in a student's learning progression are then used to extract various features that our machine-learning model will utilize.

After obtaining the preprocessed data, we must proceed to clean them before utilizing them in the modeling process. Although the preprocessed data contain all the required information, they are not structured optimally for modeling as they are in a raw form. Moreover, most of the attributes in the preprocessed data are of varying data types, which makes it necessary to standardize the data type to a consistent numerical format for modeling. Through data cleaning, the preprocessed data are transformed into a cleaned dataset that is suitable for input of the ML model. This data-cleaning process has been described in [19]. Also, we normalized the course data based on the Z-score procedure. The Z-score procedure, also known as standardization, is a statistical method used to transform a distribution of raw scores (i.e. scores on a test or other measure) into a standard normal distribution, where the mean is 0 and the standard deviation is 1. This procedure is often used in data analysis and research to compare scores across different populations, measures, or time points. The Z-score of a raw score is calculated by subtracting the mean of the distribution from the raw score and dividing the result by the standard deviation of the distribution.

$$Z = (X - \mu) / \sigma \tag{3}$$

where Z is the Z-score, X is the raw score, μ is the mean of the distribution, and σ is the standard deviation of the distribution.

Once the Z-score is calculated for each raw score in the distribution, the transformed scores can be plotted on a standard normal distribution, allowing comparisons between different distributions. The Z-score procedure has several important uses, such as identifying outliers, comparing scores on different measures or tests, and determining the proportion of scores that fall within a certain range of values. However, it is important to note that the Z-score procedure assumes that the distribution is normal or approximately normal and may not be appropriate for non-normal distributions.

We utilized the sklearn.feature_selection algorithm from Python to check on the best features that represent the dataset. It turns out that the best features that represent the data are (in sequence): Final, Midterm, Quiz_2, Assignments, Quiz_1, Credit, and Level. Therefore, the GPA and difficulty seem to have no effect on the prediction process, see Figure 2.

C. Machine Learning Model Optimization

We utilized various ML models previously reviewed in the background section: LR, GBC, ET, DT, RF, ANNs, RNNs, CNNs, kNN, FL, and NF. The proposed model is depicted in Figure 3, where the cleaned data are fed into each ML model, and each model is optimized for the best performance based on the input data. The accuracy output of each model is then utilized in the voting module. The voting module calculates the average accuracy of the ML models producing either 0 or 1. If the average accuracy of 0 is greater than the average accuracy of 1, it is determined that the probability of a student failing is higher than the probability that he/she will pass, and vice versa.

Feature	Percentage
Final	52.66%
Midterm	36.55%
Assignments	8.30%
Level	1.11%
Quiz_1	0.63%
Credit	0.57%
Quiz_2	0.17%
GPA	0.00%
Difficulty	0.00%

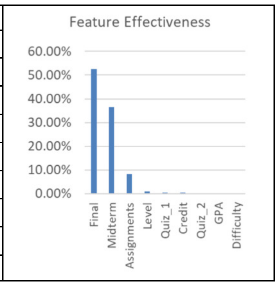

Fig. 2. Feature effectiveness.

D. Performance Metrics

To evaluate the proposed ML models, standard criteria were utilized. Table III presents a comprehensive overview of the key performance metrics and their corresponding equations used to evaluate the classification models, particularly in binary classification tasks. It includes basic terms such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which are the building blocks for calculating more advanced metrics. The table also outlines essential evaluation metrics, such as accuracy, precision, F-score, recall/sensitivity, specificity, and balanced accuracy. These metrics provide valuable insight into the performance of a classification model, helping identify its strengths and weaknesses and guiding further improvements in model development.

TABLE III. PERFORMANCE MEASURE CRITERIA

Term	Equation
TP	Count of true positives
TN	Count of true negatives
FP	Count of false positives
FN	Count of false negatives
Accuracy	Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision	Precision = TP / (TP + FP)
F-score	F-score = 2 × (Precision * Recall) / (Precision + Recall)
Recall/Sensitivity	Recall = TP / (TP + FN)
Specificity	Specificity = TN / (TN + FP)
Balanced Accuracy	Balanced Accuracy = (Recall + Specificity) / 2

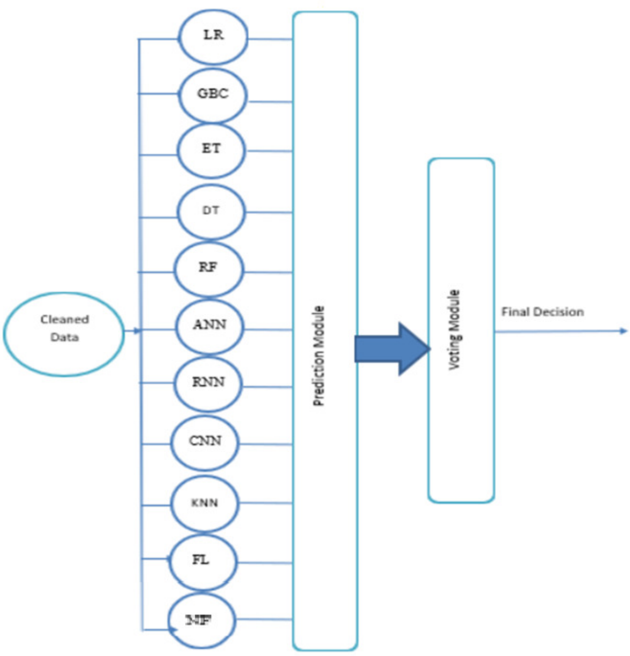

Fig. 3. Details of the proposed model.

IV. EXPERIMENTAL RESULTS

In this section, the implementation of the proposed methods is described, and the results are explained. The first model implemented is Fuzzy Logic with the membership functions presented in Figure 4.

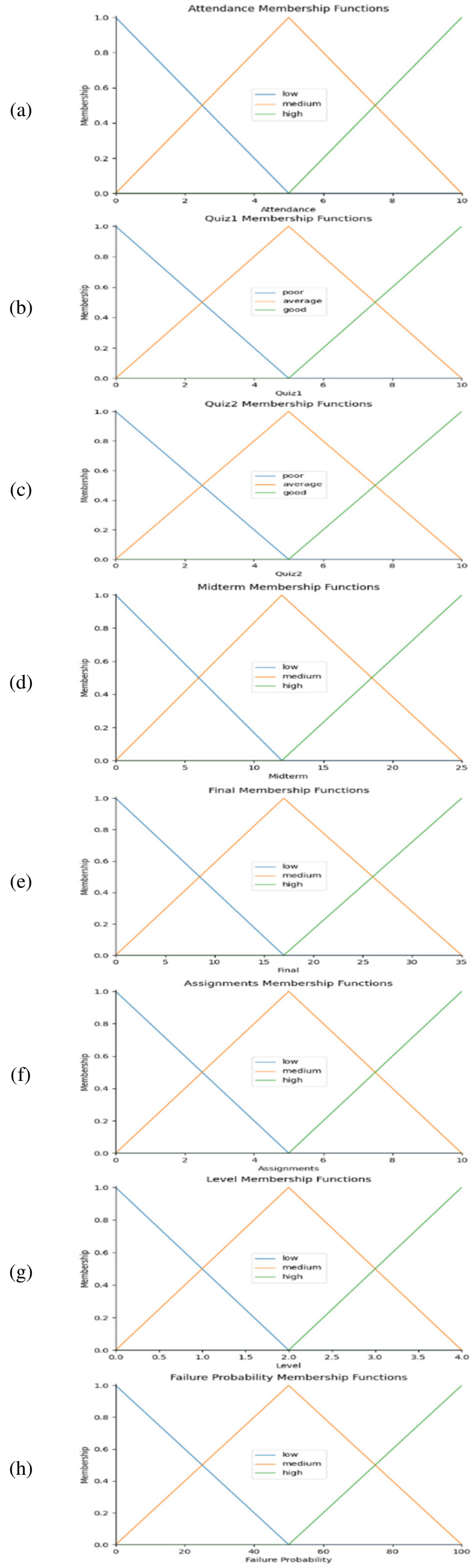

Fig. 4. Fuzzy logic membership functions: (a) Attendance, (b) Quiz 1, (c) Quiz 2, (d) Midterm, (e) Final, (f) Assignment, (g) Level, and (h) Failure probability.

In Figure 5, the performance of the Fuzzy logic classification algorithm is analyzed with including accuracy, precision, recall, F-score, specificity, and balanced accuracy. The Fuzzy algorithm attained an accuracy of 95.4%. Precision, indicating positive result accuracy, reached 88.4%, while recall, highlighting positive instance identification, was 86.4%. The F-score at 0.874 is a comprehensive performance measure. Notably, the algorithm excelled in imbalanced data with specificity at 0.974 (97.4% negative identification) and balanced accuracy at 0.919, which considers both recall and specificity. Overall, the Fuzzy algorithm showed strong classification performance with high accuracy, precision, recall, F-score, and demonstrated robustness on imbalanced data. Another implemented model, Neuro-fuzzy, displayed similar fuzzy membership functions. In Figure 6, Neuro-fuzzy exhibited high performance in accuracy, precision, recall, F-score, specificity, and balanced accuracy. Its accuracy at 0.982 signifies a high proportion of correct classifications. Precision indicated an 84.8% accurate prediction of positive outcomes. A recall of 0.971 demonstrated successful identification of positive cases. The F-score of 0.905 balances precision and recall. The algorithm's specificity at 0.983 ensured accurate negative case identification. With a balanced accuracy of 0.977, it performed well across both classes. In summary, Neuro-fuzzy proved a promising classification approach, achieving notable accuracy, precision, recall, F-score, specificity, and balanced accuracy.

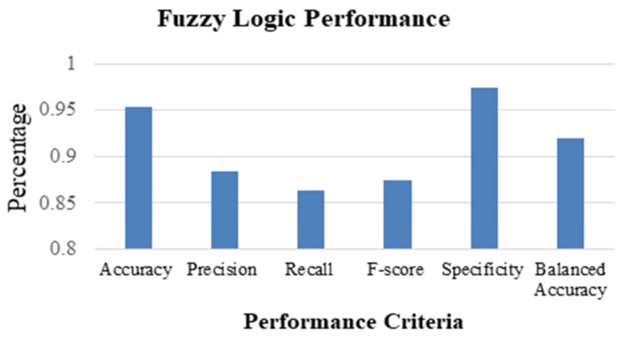

Fig. 5. Fuzzy logic performance.

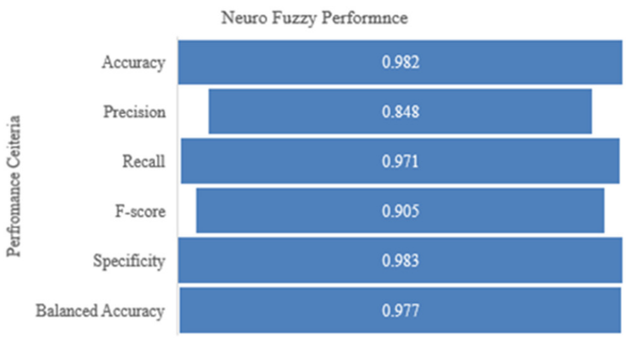

Fig. 6. Neuro-fuzzy logic.

In terms of algorithm performance, the LR algorithm achieved an accuracy of 91%, indicating generally accurate predictions. Its precision of 0.577 suggests moderate ability to correctly identify positive cases, while a recall of 0.174 indicates difficulty in identifying positive cases. The F-score, at 0.267, highlights relatively poor performance in identifying positive cases. However, the algorithm's specificity of 0.987 demonstrates its ability to identify negative cases. The balanced accuracy of 0.5805 reflects a moderate overall performance, indicating that the LR algorithm excels in identifying negative cases but struggles with positive cases.

On the other hand, the GBC algorithm produced impressive results, achieving an accuracy of 98.6%. The precision score of 0.983 signifies high accuracy in positive predictions. The recall score of 83.8% indicates effective identification of the positive cases. The F-score, reaching 0.905, showcases strong performance in predicting both positive and negative cases. With a specificity score of 0.999, the algorithm excelled in identifying negative cases. The balanced accuracy of 0.9185 provides an overall measure of performance across both positive and negative classes, underscoring the algorithm's remarkable capabilities.

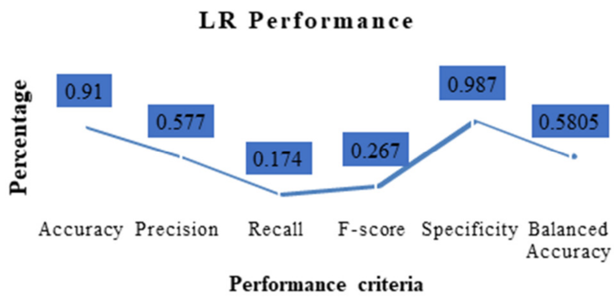

Fig. 7. Linear regression performance.

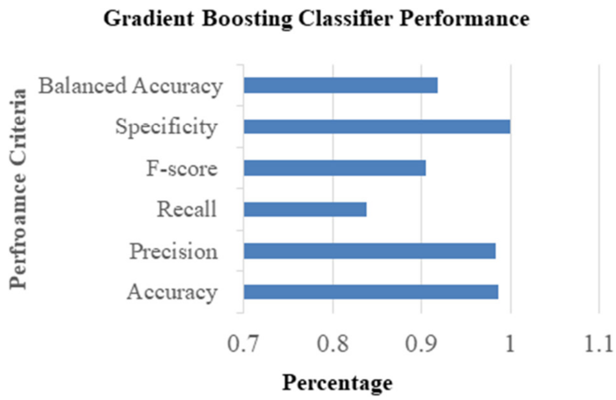

Fig. 8. Gradient boosting classifier performance.

In Figure 9, the performance of the Extra Trees (ET) algorithm is illustrated, showcasing strong results. The algorithm achieved an impressive accuracy of 0.971, indicating overall effectiveness. The precision score of 0.9 and the recall score of 0.682 highlight the model's ability to identify positive cases. Despite a slightly lower F-score of 0.776, the high specificity score of 0.994 signifies accurate identification of the negative cases. The balanced accuracy score of 0.838 further reinforces the algorithm's proficiency. In conclusion, the ET algorithm exhibits satisfying performance.

Figure 10 displays results of the DT algorithm, which achieved 96.2% accuracy, showcasing its effectiveness in accurate instance classification. With a precision of 86.8% and specificity of 98.3%, the model minimizes false-positive and false-negative errors. The recall score of 81.3% indicates a substantial identification of positive cases but leaves room for improvement. The F-score of 84% signifies a balanced trade-off between precision and recall. The balanced accuracy of 89.8% underscores the algorithm's equilibrium in recognizing positive and negative cases. Overall, the DT algorithm displays strong performance in this classification task, with potential for further enhancement.

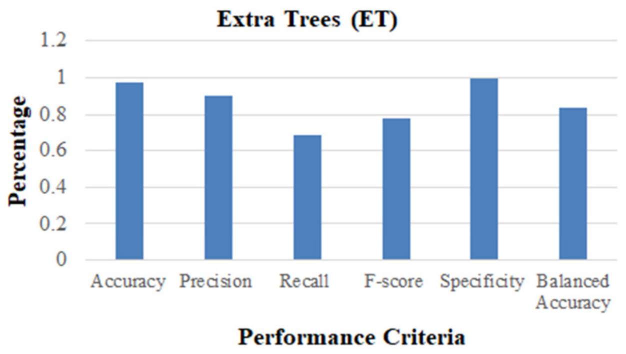

Fig. 9. Extra trees performance.

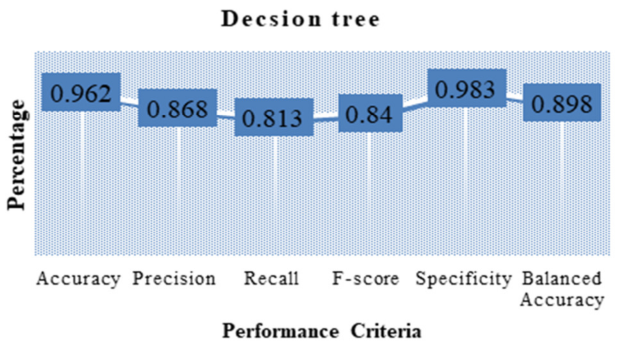

Fig. 10. Decision tree performance.

RF in Figure 11 demonstrates strong performance. It achieved an accuracy of 98.1%, indicating that it can correctly classify a high percentage of instances. With a precision of 85.3% and specificity of 98.7%, the algorithm effectively minimizes false-positive errors while correctly identifying negative cases. The recall of 90.4% shows that the model successfully identifies most of the positive cases, and the F-score of 87.8% highlights a good balance between precision and recall. The balanced accuracy of 94.55% further emphasizes that the algorithm performs well in identifying both positive and negative cases, making it a reliable choice for this classification problem.

The ANN algorithm in Figure 12 demonstrates outstanding performance. It achieves an accuracy of 99.9%, precision of 99.3%, recall of 99.3%, F-score of 99.3%, specificity of 99.9%, and balanced accuracy of 99.6%. These impressive results across all metrics indicate that ANNs are highly effective in identifying positive and negative cases while minimizing false positives and false negatives. The almost perfect scores in each category suggest that the ANN algorithm is well-suited for this type of problem and provides a reliable and balanced solution.

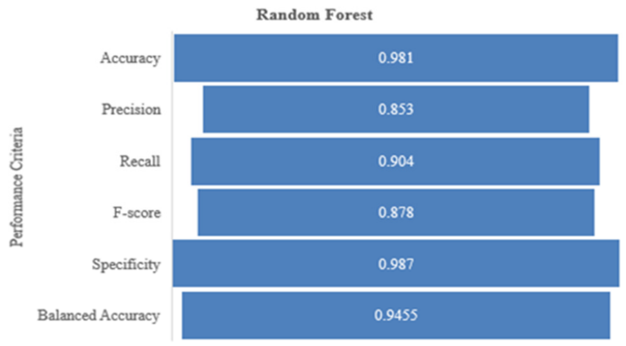

Fig. 11. Random forest performance.

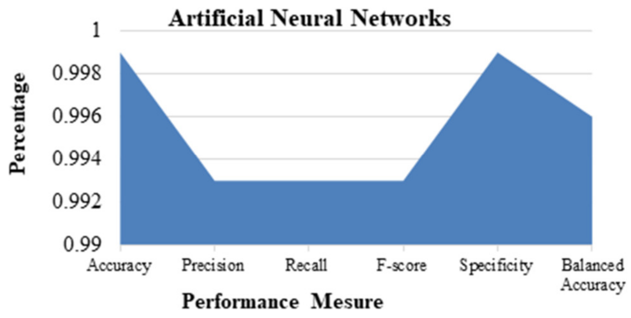

Fig. 12. Artificial neural network performance.

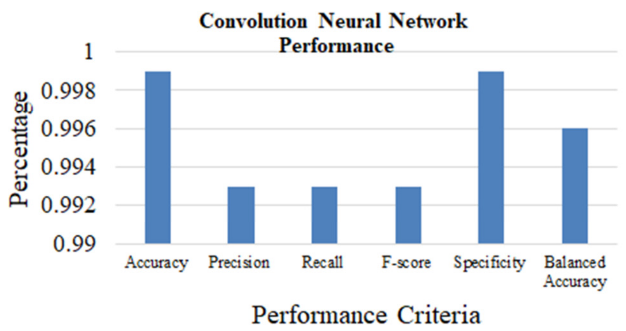

Fig. 13. Convolution neural network performance.

In Figure 13, the CNN demonstrates exceptional performance in a classification task with 1800 samples. The CNN achieves an impressive accuracy of 99.9%, along with a precision, recall, and F-score of 99.3%. These values highlight the CNN's effectiveness in accurately identifying positive and negative cases while minimizing errors. The model's specificity of 99.9% and balanced accuracy of 99.6% further underscore its ability to classify both classes accurately, making it a top-performing classifier.

Figure 14 presents the performance of the kNN algorithm. It attains an accuracy of 96.1%, indicating a substantial number of correct classifications. However, the recall of 59.2% and F-score of 70.6% suggest that the kNN struggles to identify positive cases compared to its strong specificity of 99.3%, which excels at identifying negative cases. With a precision of 87.5%, the algorithm keeps false-positive errors low. The balanced accuracy of 79.25% suggests a somewhat imbalanced performance, leaning toward specificity. There's potential for improvement in the kNN algorithm, particularly in identifying positive cases.

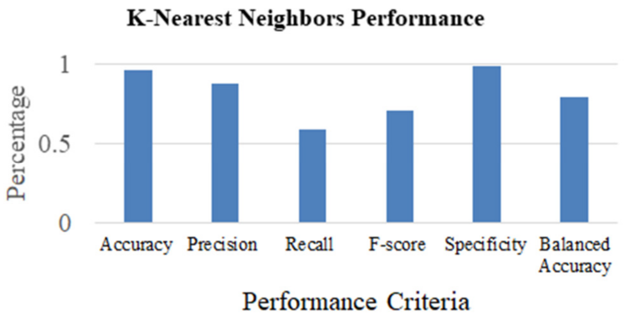

Fig. 14. kNN performance.

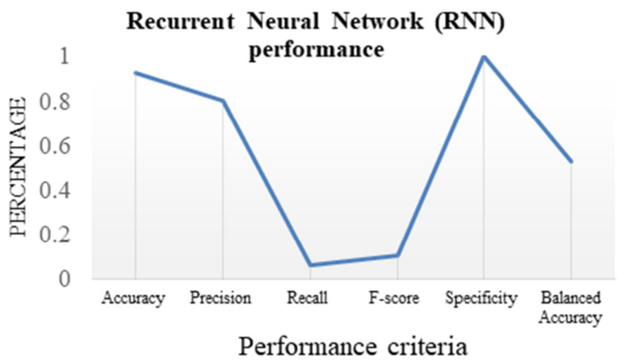

Fig. 15. Recurrent neural network performance.

The RNN performance, as presented in Figure 15, achieves an overall accuracy of 92.4%. While demonstrating a strong specificity of 99.9%, indicating proficient identification of negative cases, it struggles with positive case detection, exhibiting a low recall of 5.6%. Despite the 80% precision, the F-score is only 10.5%, highlighting an imbalance between precision and recall. The balanced accuracy of 52.75% underscores the algorithm's uneven performance between positive and negative case classification. In essence, the RNN algorithm's performance in this context is suboptimal, particularly evident in its challenge to identify positive cases, as evidenced by the low recall and F-score values.

A. Discussion

The given results summarized in Table IV present the performance metrics of 12 different algorithms. ANNs and CNNs exhibit the highest performance in almost all the metrics among all the algorithms. They both demonstrate an impressive 99.9% accuracy, 99.3% precision and recall, 99.3% F-score, 99.9% specificity, and 99.6% balanced accuracy. GBC ranks third in accuracy with 98.6%. It also features high precision and specificity, however, its recall is lower, leading to an F-score of 90.5% and a balanced accuracy of 91.85%. The Neuro Fuzzy algorithm has the second-highest accuracy (98.2%) and balanced accuracy (97.7%). It also shows good recall (97.1%) and F-score (90.5%), although its precision (84.8%) and specificity (98.3%) are lower compared to the top-performing models. The RF algorithm demonstrates strong performance with an accuracy of 98.1%. DT, kNN, and RNNs have relatively lower performance metrics than the top-performing models.

The ANN and CNN algorithms demonstrate the best overall performance among all the models. The GBC, Neuro-Fuzzy, and RF algorithms also perform well, but not as strongly as the ANN and CNN models. The other algorithms exhibit varying performance levels, with the LR and RNN algorithms showing the weakest results.

TABLE IV. RESULT SUMMARY

Algorithm	TP	TN	FP	FN	Accuracy	Precision	Recall	F-score	Specificity	Balanced Accuracy
Fuzzy	344	1701	45	54	0.954	0.884	0.864	0.874	0.974	0.919
Neuro Fuzzy	168	1765	30	5	0.982	0.848	0.971	0.905	0.983	0.977
LR	30	1636	22	142	0.91	0.577	0.174	0.267	0.987	0.5805
GBC	119	1656	2	23	0.986	0.983	0.838	0.905	0.999	0.9185
ET	90	1658	10	42	0.971	0.9	0.682	0.776	0.994	0.838
DT	178	1554	27	41	0.962	0.868	0.813	0.84	0.983	0.898
RF	122	1644	21	13	0.981	0.853	0.904	0.878	0.987	0.9455
ANN	141	1657	1	1	0.999	0.993	0.993	0.993	0.999	0.996
CNN	141	1657	1	1	0.999	0.993	0.993	0.993	0.999	0.996
kNN	84	1646	12	58	0.961	0.875	0.592	0.706	0.993	0.7925
RNN	8	1656	2	134	0.924	0.8	0.056	0.105	0.999	0.5275

V. CONCLUSION

The current paper addresses the challenge of predicting student performance using various data such as assignment grades, quizzes, midterms, finals, and other parameters. The focus is on preemptively identifying student failure. Multiple machine learning algorithms are proposed, including Fuzzy, Neuro-Fuzzy, LR, GBC, ET, DT, RF, ANN, CNN, kNN, and RNN. These algorithms' performances are evaluated using metrics like Accuracy, Precision, Recall, F-score, Specificity, and Balanced Accuracy. The findings reveal that ANN and CNN outperform the other models, with GBC, Neuro-Fuzzy, and RF also showing strong but less consistent results. The other algorithms exhibit varying levels of effectiveness, with LR and RNN performing poorly. The paper's innovation lies in automating accurate student performance assessment, potentially saving instructors' time and ensuring fairness. For student performance evaluation, the recommendation is to prioritize the use of ANNs and CNNs due to their superior performance compared

REFERENCES

[1] I. Guellil and K. Boukhalfa, "Social big data mining: A survey focused on opinion mining and sentiments analysis," in *2015 12th International Symposium on Programming and Systems (ISPS)*, Algiers, Algeria, Apr. 2015, https://doi.org/10.1109/ISPS.2015.7244976.

[2] S. Sharma and V. Mangat, "Technology and Trends to Handle Big Data: Survey," in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, Haryana, India, Oct. 2015, pp. 266–271, https://doi.org/10.1109/ACCT.2015.121.

[3] F. Xia, W. Wang, T. M. Bekele, and H. Liu, "Big Scholarly Data: A Survey," *IEEE Transactions on Big Data*, vol. 3, no. 1, pp. 18–35, Mar. 2017, https://doi.org/10.1109/TBDATA.2016.2641460.

[4] J. Zhang, X. Yao, G. Han, and Y. Gui, "A survey of recent technologies and challenges in big data utilizations," in *2015 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, Korea, Jul. 2015, pp. 497–499, https://doi.org/10.1109/ICTC.2015.7354594.

[5] J. V. Gautam, H. B. Prajapati, V. K. Dabhi, and S. Chaudhary, "A survey on job scheduling algorithms in Big data processing," in *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, Mar. 2015, pp. 1–11, https://doi.org/10.1109/ICECCT.2015.7226035.

[6] A. Fahad *et al.*, "A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 3, pp. 267–279, Sep. 2014, https://doi.org/10.1109/TETC.2014.2330519.

[7] S. Gole and B. Tidke, "A survey of big data in social media using data mining techniques," in *2015 International Conference on Advanced Computing and Communication Systems*, Coimbatore, India, Jan. 2015, https://doi.org/10.1109/ICACCS.2015.7324059.

[8] J. Wang, Y. Wu, N. Yen, S. Guo, and Z. Cheng, "Big Data Analytics for Emergency Communication Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1758–1778, 2016, https://doi.org/10.1109/COMST.2016.2540004.

[9] S. Yu, M. Liu, W. Dou, X. Liu, and S. Zhou, "Networking for Big Data: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 531–549, 2017, https://doi.org/10.1109/COMST.2016.2610963.

[10] D. Ramesh, P. Suraj, and L. Saini, "Big data analytics in healthcare: A survey approach," in *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, Durgapur, India, Jan. 2016, https://doi.org/10.1109/MicroCom.2016.7522520.

[11] M. Pandey, R. Litoriya, and P. Pandey, "Mobile applications in context of big data: A survey," in *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, Indore, India, Mar. 2016, https://doi.org/10.1109/CDAN.2016.7570942.

[12] M. Saberi, A. Karduck, O. K. Hussain, and E. Chang, "Challenges in Efficient Customer Recognition in Contact Centre: State-of-the-Art Survey by Focusing on Big Data Techniques Applicability," in *2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, Ostrava, Czech Republic, Sep. 2016, pp. 548–554, https://doi.org/10.1109/INCoS.2016.136.

[13] Y. Hou, J. Xu, Y. Huang, and X. Ma, "A big data application to predict depression in the university based on the reading habits," in *2016 3rd International Conference on Systems and Informatics (ICSAI)*, Shanghai, China, Aug. 2016, pp. 1085–1089, https://doi.org/10.1109/ICSAI.2016.7811112.

[14] R. M. Mathew and R. Gunasundari, "A Cluster-based Undersampling Technique for Multiclass Skewed Datasets", *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 3, pp. 10785–10790, Jun. 2023.

[15] A. B. Rashid, R. R. R. Ikram, Y. Thamilarasan, L. Salahuddin, N. F. A. Yusof, and Z. B. Rashid, "A Student Learning Style Auto-Detection Model in a Learning Management System," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 11000–11005, Jun. 2023, https://doi.org/10.48084/etasr.5751.

[16] S. Joseph, N. Mduma, and D. Nyambo, "A Deep Learning Model for Predicting Stock Prices in Tanzania," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10517–10522, Apr. 2023, https://doi.org/10.48084/etasr.5710.

[17] B. Veloso, M. A. Barbosa, H. Faria, F. S. Marcondes, D. Durães, and P. Novais, "A Systematic Review on Student Failure Prediction," in *Methodologies and Intelligent Systems for Technology Enhanced Learning, Workshops, 12th International Conference*, 2023, pp. 43–52, https://doi.org/10.1007/978-3-031-20257-5_5.

[18] H. Waheed, S.-U. Hassan, R. Nawaz, N. R. Aljohani, G. Chen, and D. Gasevic, "Early prediction of learners at risk in self-paced education: A neural network approach," *Expert Systems with Applications*, vol. 213, Mar. 2023, Art. no. 118868, https://doi.org/10.1016/j.eswa.2022.118868.

[19] Nurmalitasari, Z. Awang Long, and M. F. Mohd Noor, "The Predictive Learning Analytics for Student Dropout Using Data Mining Technique: A Systematic Literature Review," in *Advances in Technology Transfer Through IoT and IT Solutions*, A. Ismail, F. N. Zulkipli, Z. Awang Long, and A. Öchsner, Eds. Springer Nature Switzerland, 2023, pp. 9–17.

[20] A. Kukkar, R. Mohana, A. Sharma, and A. Nayyar, "Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms," *Education and Information Technologies*, vol. 28, no. 8, pp. 9655–9684, Aug. 2023, https://doi.org/10.1007/s10639-022-11573-9.

[21] M. Saarinen, R. Bertram, K. Aunola, J. Pankkonen, and T. V. Ryba, "Student Athletes' Causal Attributions for Sport and School Achievement in Relation to Sport Dropout and Grade Point Average," *Journal of Sport & Exercise Psychology*, vol. 45, no. 1, pp. 15–25, Feb. 2023, https://doi.org/10.1123/jsep.2022-0115.

[22] L. K. Smirani, H. A. Yamani, L. J. Menzli, and J. A. Boulahia, "Using Ensemble Learning Algorithms to Predict Student Failure and Enabling Customized Educational Paths," *Scientific Programming*, vol. 2022, Apr. 2022, Art. no. e3805235, https://doi.org/10.1155/2022/3805235.

[23] T. A. Kustitskaya, A. A. Kytmanov, and M. V. Noskov, "Early Student-at-Risk Detection by Current Learning Performance and Learning Behavior Indicators," *Cybernetics and Information Technologies*, vol. 22, no. 1, pp. 117–133, Mar. 2022, https://doi.org/10.2478/cait-2022-0008.

[24] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, Mar. 2022, Art. no. 11, https://doi.org/10.1186/s40561-022-00192-z.

[25] J. Hao, J. Gan, and L. Zhu, "MOOC performance prediction and personal performance improvement via Bayesian network," *Education and Information Technologies*, vol. 27, no. 5, pp. 7303–7326, Jun. 2022, https://doi.org/10.1007/s10639-022-10926-8.

[26] D. Alboaneen, M. Almelihi, R. Alsubaie, R. Alghamdi, L. Alshehri, and R. Alharthi, "Development of a Web-Based Prediction System for Students' Academic Performance," *Data*, vol. 7, no. 2, Feb. 2022, Art. no. 21, https://doi.org/10.3390/data7020021.

[27] S. S. Shreem, H. Turabieh, S. Al Azwari, and F. Baothman, "Enhanced binary genetic algorithm as a feature selection to predict student performance," *Soft Computing*, vol. 26, no. 4, pp. 1811–1823, Feb. 2022, https://doi.org/10.1007/s00500-021-06424-7.

[28] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, "Student Academic Performance Prediction using Supervised Learning Techniques," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 14, pp. 92–104, Jul. 2019, https://doi.org/10.3991/ijet.v14i14.10310.

[29] S. Kotsiantis, K. Patriarcheas, and M. Xenos, "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 529–535, Aug. 2010, https://doi.org/10.1016/j.knosys.2010.03.010.

[30] C. Márquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. Mousa Fardoun, and S. Ventura, "Early dropout prediction using data mining: a case study with high school students," *Expert Systems*, vol. 33, no. 1, pp. 107–124, 2016, https://doi.org/10.1111/exsy.12135.

[31] M. Fei and D.-Y. Yeung, "Temporal Models for Predicting Student Dropout in Massive Open Online Courses," in *2015 IEEE International*

Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, Aug. 2015, pp. 256–263, https://doi.org/10.1109/ICDMW.2015.174.