ISSN: 2581-3404 (Online) IF: 5.86 (SJIF)
IJIRTM, Volume-10, Issue-3, June-2026

Genetic Algorithm and Neural Network in Educational Data Mining

[1]Mirza Hufeza Baig [2]Aashish Kumar Tiwari
[1]Research Scholar, [2]Professor
[1]Department of Computer Science & Engineering,
[1]SAM College of Engineering and Technology, Bhopal, India.

Abstract. *Educational Data Mining (EDM) has become an effective research area for extracting meaningful knowledge from educational datasets and improving student learning outcomes. Accurate prediction of student performance plays a vital role in identifying academically weak students and supporting educational decision-making. This paper proposes a Behavior-Based Student Classification System (SCS-B) that integrates Genetic Algorithm (GA) and Back Propagation Neural Network (BP-NN) for student performance prediction and classification. Student data are collected through a structured Student Questionnaire (SQ) comprising learning techniques, personal information, student behavior, intellectual factors, and comprehensive abilities. The collected data undergo preprocessing through outlier detection and dimensionality reduction using Singular Value Decomposition (SVD). Genetic Algorithm is employed for optimal feature selection, while BP-NN is used for classification and prediction. The proposed model classifies students into four categories: Class A, Class B, Class C, and Class D based on their academic and behavioral characteristics. Experimental evaluation demonstrates that the proposed SCS-B model achieves superior classification accuracy compared with Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Multi-Tier Student Performance Evaluation Model (MTSPEM). The results confirm that integrating behavioral factors with machine learning techniques significantly enhances student performance prediction and educational decision support.*

Keywords: Educational Data Mining, Student Performance Prediction, Genetic Algorithm, Neural Network, Behavior Analysis, Machine Learning, Student Classification.

Introduction

Educational institutions continuously generate large volumes of student-related information through academic records, attendance systems, online learning platforms, and behavioral assessments. The effective analysis of these datasets can provide valuable insights into student learning patterns and academic performance. Educational Data Mining (EDM) focuses on applying machine learning and data mining techniques to educational datasets for discovering hidden patterns and improving educational quality. Traditional student evaluation systems mainly rely on examination scores and attendance records. However, academic performance is influenced by several behavioral, intellectual, and personal factors. Therefore, incorporating behavioral analysis into student performance prediction can significantly improve classification accuracy and educational decision-making. This study proposes a Behavior-Based Student Classification System (SCS-B) using Genetic Algorithm and Neural Networks for predicting

student academic performance and categorizing students according to their overall learning characteristics.

1.1 Research Objectives

The main objectives of the proposed research are:

1. To study the concepts and applications of Educational Data Mining (EDM) for student performance analysis.
2. To collect and preprocess student academic, behavioural, and personal data for effective analysis.
3. To identify the key factors influencing student academic performance.
4. To perform outlier detection and dimensionality reduction for improving data quality.
5. To apply Genetic Algorithm (GA) for selecting the most significant student performance attributes.
6. To develop a Neural Network-based classification model for student performance prediction.
7. To classify students into different performance categories based on their academic and behavioral characteristics.
8. To compare the performance of the proposed model with existing machine learning techniques such as SVM, MLP, and MTSPEM.
9. To evaluate the developed framework using Accuracy, Precision, Recall, Error Rate, and Processing Time.
10. To provide an intelligent decision-support system for identifying weak students and improving educational outcomes.

Literature Review

The literature review indicates that Educational Data Mining and Machine Learning techniques play a vital role in predicting student academic performance. Various classification and prediction models have been developed to improve educational decision-making and identify at-risk students. However, there is still a need for intelligent hybrid frameworks to achieve higher prediction accuracy and better educational outcomes.

Elbadrawy and Karypis (2016) Elbadrawy and Karypis proposed a domain-aware grade prediction and course recommendation framework. The model analyzed academic records and learning patterns to provide personalized course recommendations. The study concluded that intelligent recommendation systems enhance student success rates and academic planning.

Baker and Yacef (2017) Baker and Yacef presented a comprehensive review of Educational Data Mining methodologies and applications. The study discussed classification, clustering, prediction, and association rule mining techniques. The authors concluded that EDM plays a crucial role in improving educational quality and understanding student learning behavior.

Philipp Leitner, Khalil and Ebner (2017) The researchers conducted a detailed literature review on Learning Analytics in higher education. Their study emphasized the role of predictive analytics and machine learning techniques in identifying at-risk students and improving academic outcomes.

Houbraken et al. (2017) Houbraken et al. proposed a method for discovering hidden course requirements and student competencies using academic grade data. The study highlighted the significance of educational analytics in curriculum planning and student guidance systems.

Anand et al. (2018) Anand et al. introduced a recursive clustering approach for student performance evaluation. The model classified students into different groups according to academic achievements and learning characteristics. The results indicated improved classification accuracy and educational decision support.

Bakhshinategh et al. (2018) Bakhshinategh et al. reviewed major Educational Data Mining applications developed over the previous decade. The study covered classification, clustering, prediction, and association rule mining techniques. The authors emphasized that EDM significantly contributes to intelligent educational systems and adaptive learning environments.

Park (2018) Park developed a collaborative filtering approach for personalized student performance prediction. The study utilized recommendation-based learning systems to support individualized educational guidance. The proposed method achieved higher prediction accuracy through student similarity analysis.

Gkontzis et al. (2019) Gkontzis et al. proposed a predictive analytics framework for reducing student dropout rates. The model analyzed attendance records, academic performance, and behavioral factors to identify at-risk students. The study demonstrated the effectiveness of early intervention strategies in higher education.

Alisa Bilal Zorić (2020) The study highlighted the importance of Educational Data Mining in improving teaching quality and student learning outcomes. The author concluded that predictive analytics supports academic decision-making and enables institutions to improve educational effectiveness.

Fischer, Pardos and Baker (2020) The authors investigated big data mining techniques in education and discussed opportunities and challenges associated with educational analytics. Their work emphasized the importance of machine learning and large-scale educational datasets in developing intelligent learning systems.

Said A. Salloum et al. (2020) Salloum et al. presented a comprehensive review of Educational Data Mining techniques and future research directions. The study analyzed various machine learning algorithms and concluded that integrating artificial intelligence with EDM significantly improves student performance prediction.

Ali Jaber Almalki (2021) Almalki analyzed the impact of feature selection techniques on Educational Data Mining performance. The research demonstrated that selecting relevant attributes improves classification accuracy and reduces computational complexity.

Roslan and Chen (2022) Roslan and Chen conducted a systematic literature review on student performance prediction from 2015 to 2021. The study identified Decision Tree, Random Forest, and Artificial Neural Networks as the most frequently used prediction algorithms. The authors emphasized the importance of academic records and demographic attributes in educational analytics.

Zahruddin et al. (2023) Zahruddin et al. proposed a machine learning-based educational framework for predicting student academic performance. The study utilized classification algorithms and feature extraction methods to identify performance categories. The results demonstrated improved educational quality and prediction accuracy.

Tao-Hongli (2024) Tao-Hongli developed an intelligent Educational Data Mining framework using Adaptive Sea Horse Optimization-based feature selection techniques. The proposed model improved classification accuracy and highlighted the significance of intelligent feature extraction in educational analytics. **Angeioplastis et al. (2025)**

Angeioplastis et al. proposed a data-driven Educational Data Mining framework using Moodle Learning Management System datasets. The study compared Decision Tree, Random Forest, Logistic Regression, Neural Networks, and k-Nearest Neighbor algorithms. Experimental results indicated that Neural Networks and kNN achieved superior performance in student performance prediction and personalized learning support.

2.1 Research Gap

Although numerous machine learning and Educational Data Mining models have been proposed for student performance prediction, most existing studies focus primarily on academic records and attendance data. Limited research has integrated behavioral, demographic, and psychological factors simultaneously. Furthermore, many models suffer from issues such as feature redundancy, limited dataset size, lack of interpretability, and inadequate real-time prediction capabilities. Therefore, there is a need to develop an intelligent hybrid framework that combines feature selection, clustering, classification, and machine learning techniques to achieve higher prediction accuracy and support educational decision-making.

Research Methodology

The proposed methodology consists of the following stages:

1. Student Questionnaire (SQ) Based Data Collection
2. Outlier Detection using Singular Value Decomposition
3. Dimensionality Reduction
4. Genetic Algorithm-Based Feature Selection
5. Neural Network Training and Classification
6. Student Performance Prediction

The questionnaire dataset includes:

- Learning Techniques
- Personal Information
- Student Behaviour
- Intellectual Factors
- Comprehensive Ability

After preprocessing, significant features are selected through Genetic Algorithm and provided to the Back Propagation Neural Network for training and classification

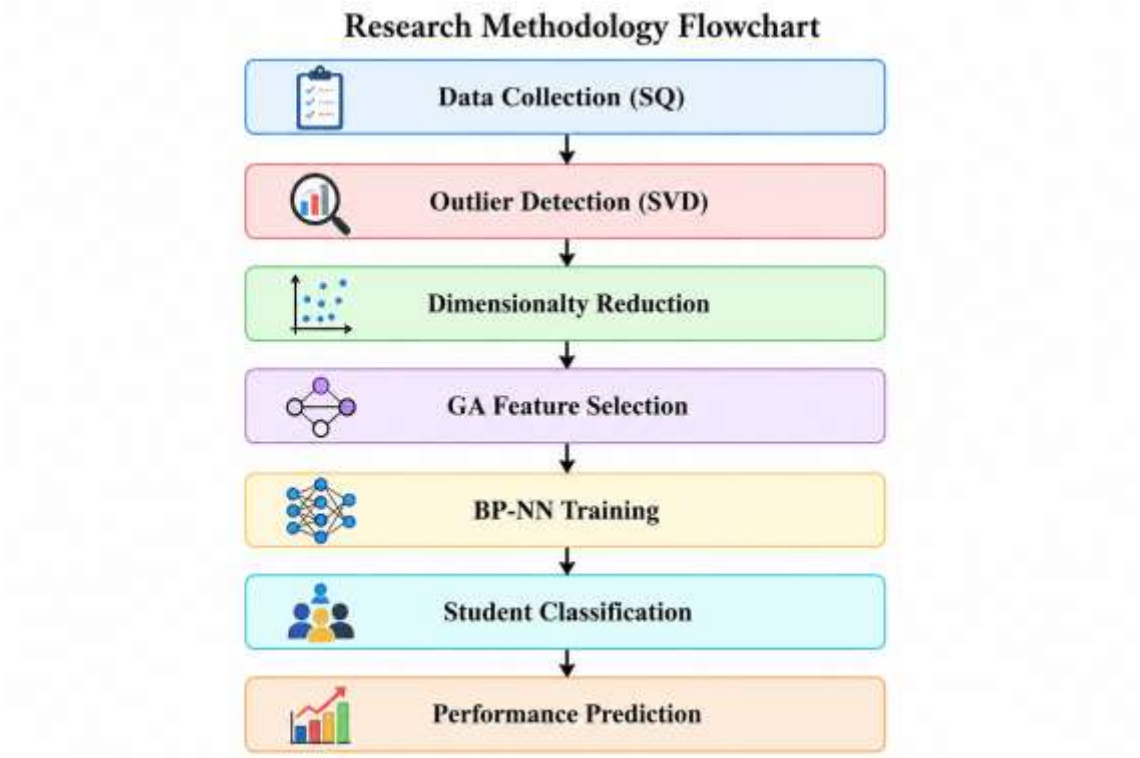


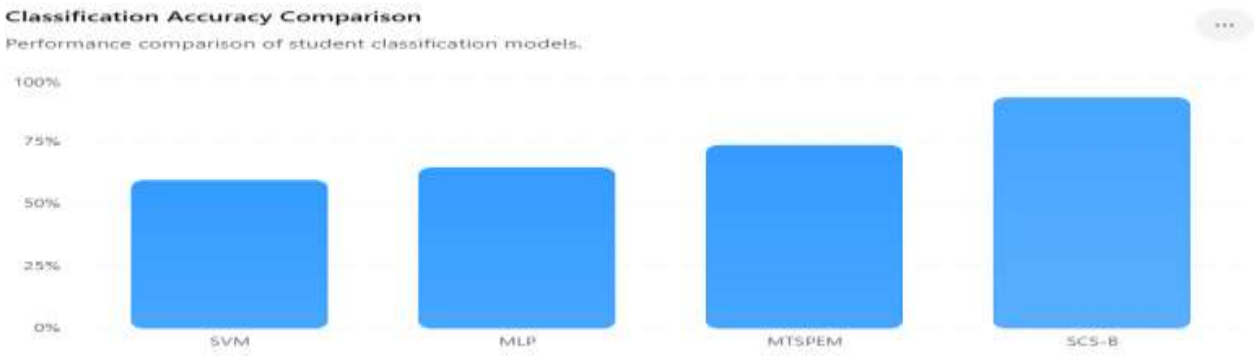
Figure 1: Research Methodology Flowchart for Student Performance Prediction using Genetic Algorithm and Back Propagation Neural Network (BP-NN).

Experimental Results

The proposed SCS-B model was evaluated using 200 student samples and compared with existing models such as SVM, MLP, and MTSPEM. The obtained results demonstrated superior performance of the proposed framework.

Table 1: Classification Accuracy Comparison.

Model	**Accuracy (%)**
SVM	59.2
MLP	64.3
MTSPEM	73.3
SCS-B (Proposed)	92.4



Graph 2: Classification Accuracy.

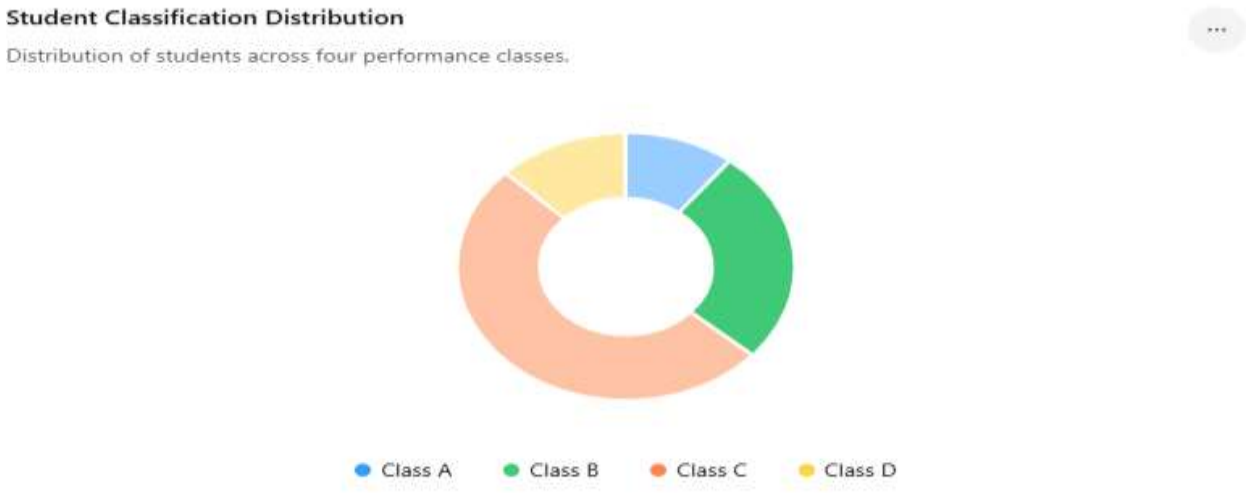
The classification accuracy comparison demonstrates that the proposed SCS-B model significantly outperforms the existing models. SVM achieved an accuracy of 59.2%, while MLP and MTSPEM obtained 64.3% and 73.3%, respectively. The proposed SCS-B framework achieved the highest accuracy of 92.4%, indicating the effectiveness of integrating Genetic Algorithm and Neural Network techniques for student performance prediction.

Table 2: Processing Time Comparison.

Model	**Processing Time (ms)**
SVM	68.61
MLP	120.0
MTSPEM	46.3
SCS-B (Proposed)	23.7

Graph 3: Processing Time Analysis.

Processing time is an important parameter for evaluating model efficiency. The proposed SCS-B model required only 23.7 ms for processing student records, which is significantly lower than SVM (68.61 ms), MLP (120 ms), and MTSPEM (46.3 ms). The reduced execution time indicates that the proposed framework is computationally efficient and suitable for large educational datasets



Graph 4: Student Classification Distribution.

The final classification results show that 51% of students belong to Class C, indicating average academic and behavioral performance. Class B contains 25% of students, while Class A includes only 11% of students with excellent academic and behavioral characteristics. Approximately 13% of students fall into Class D, representing students who require special academic attention and support. These findings demonstrate the capability of the proposed SCS-B model to effectively categorize students based on their overall performance. The experimental results confirm that the proposed Behavior-Based Student Classification System (SCS-B) provides superior classification accuracy with lower processing time compared to traditional machine learning models. The integration of Genetic Algorithm for feature selection and Back Propagation Neural Network for classification improves prediction performance and supports educational decision-making. The framework can assist institutions in identifying weak students early and implementing targeted interventions to improve academic outcomes

Conclusion

This research presented a Behaviour-Based Student Classification System (SCS-B) for predicting and classifying student academic performance using Educational Data Mining techniques. The proposed framework integrates Genetic Algorithm (GA) for feature selection and Back Propagation Neural Network (BP-NN) for classification, enabling effective analysis of academic, behavioural, and personal attributes of students. Experimental results demonstrated that the proposed model achieved higher classification accuracy and lower processing time compared to existing approaches such as SVM, MLP, and MTSPEM. The classification of students into four performance categories (Class A, B, C, and D)

provides valuable insights for educators to identify weak students and implement appropriate academic interventions. Overall, the proposed framework contributes to improving educational decision-making, student monitoring, and academic success through intelligent data-driven analysis.

Future Scope

1. The proposed model can be extended using Deep Learning techniques such as CNN, RNN, and LSTM for improved prediction accuracy.
2. Real-time student performance monitoring systems can be developed by integrating Learning Management System (LMS) data.
3. Additional behavioral, psychological, and socio-economic factors may be incorporated to enhance prediction performance.
4. Hybrid optimization algorithms such as PSO, GWO, and ACO can be integrated with Neural Networks for better feature selection.
5. Explainable Artificial Intelligence (XAI) techniques can be applied to improve model transparency and interpretability.
6. The framework can be tested on larger and multi-institutional datasets to improve generalization capability. Cloud-based and web-based educational analytics platforms can be developed for practical deployment.
7. Personalized learning recommendation systems can be integrated to support individual student learning needs.
8. The proposed framework can be extended for dropout prediction and student retention analysis.
9. Future research may focus on developing intelligent decision-support systems for smart learning environments and digital education platforms.

References

1. A. Elbadrawy and G. Karypis, “Domain-Aware Grade Prediction and Top-n Course Recommendation,” in Proceedings of the 10th ACM Conference on Recommender Systems, 2016, pp. 183–190.
2. R. S. Baker and K. Yacef, “The State of Educational Data Mining in 2009: A Review and Future Visions,” Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2017.
3. P. Leitner, M. Khalil, and M. Ebner, “Learning Analytics in Higher Education—A Literature Review,” in Learning Analytics: Fundaments, Applications and Trends, Springer, 2017, pp. 1–23.
4. G. Houbraken, K. Hardeman, H. Vanthienen, and B. Baesens, “Predicting Student Performance in Higher Education Using Data Mining Techniques,” Expert Systems with Applications, vol. 82, pp. 67–77, 2017.
5. V. Anand, S. Sharma, and R. Kumar, “Student Performance Evaluation Using Recursive Clustering Techniques,” International Journal of Computer Applications, vol. 182, no. 12, pp. 15–20, 2018.
6. B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, “Educational Data Mining Applications and Tasks: A Survey of the Last 10 Years,” Education and Information Technologies, vol. 23, no. 1, pp. 537–553, 2018.
7. Y. Park, “A Collaborative Filtering Approach to Student Performance Prediction,” Educational Technology Research and Development, vol. 66, no. 4, pp. 857–879, 2018.

8. A. F. Gkontzis, T. Kotsiantis, and C. Pierrakeas, "Predicting Student Dropout in Higher Education Through Data Mining Techniques," International Journal of Learning Analytics and Artificial Intelligence for Education, vol. 1, no. 1, pp. 1–12, 2019.
9. A. B. Zorić, "Educational Data Mining and Learning Analytics in Higher Education," TEM Journal, vol. 9, no. 3, pp. 1091–1097, 2020.
10. C. Fischer, Z. A. Pardos, and R. S. Baker, "Mining Big Data in Education: Affordances and Challenges," Review of Research in Education, vol. 44, no. 1, pp. 130–160, 2020.
11. S. A. Salloum, A. Q. M. Alhamad, M. Al-Emran, A. A. Monem, and K. Shaalan, "Using Machine Learning Techniques for Student Performance Prediction: A Systematic Review," IEEE Access, vol. 8, pp. 199828–199839, 2020.
12. A. J. Almalki, "Educational Data Mining and Feature Selection Techniques for Student Performance Prediction," Journal of Information Technology Education: Research, vol. 20, pp. 101–120, 2021.
13. N. Roslan and L. Chen, "Educational Data Mining for Student Performance Prediction: A Systematic Literature Review," Computers and Education: Artificial Intelligence, vol. 3, 100024, 2022.
14. M. Zahruddin, A. Rahman, and S. Hasan, "Machine Learning Approaches for Academic Performance Prediction in Higher Education," International Journal of Advanced Computer Science and Applications, vol. 14, no. 4, pp. 245–253, 2023.
15. Tao Hongli, "An Intelligent Educational Data Mining Framework for Student Performance Prediction Using Adaptive Optimization Techniques," Expert Systems with Applications, vol. 238, 2024.
16. Angeioplastis, P., et al., "Machine Learning-Based Student Performance Prediction in Learning Management Systems: A Comparative Study," Computers & Education, vol. 205, 2025.