

Comparative Analysis of Hybrid Machine Learning Models for Predicting Student Performance with Data Balancing Via SMOTE and GAN

Chandra Sekhar Sanaboina,

Professor, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India

D. Ramya Sri

M.Tech, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India

Abstract: This study aims to develop and evaluate hybrid machine learning models—XGBoost, CNN+LSTM, and GBNN—for weekly prediction of student performance using a structured academic dataset and advanced data balancing mechanisms. By addressing class imbalance with both SMOTE and GAN techniques, the system significantly improves detection of failing and dropout cases. Experimental results demonstrate superior recall and F1 scores for minority classes, particularly using GAN-based augmentation with sequential models. Compared to previous projects relying on conventional models without balancing, this approach achieves an average accuracy improvement of approximately 5–7%, with the best models reaching accuracy levels around 96–98%. The models were implemented in Python, leveraging TensorFlow and Keras for deep learning components (CNN+LSTM, GAN), and XGBoost with Scikit-learn for gradient boosting. This framework enhances early intervention capacity for academic support systems and sets a new benchmark for robust educational data mining.

Keywords: Student Performance Prediction, Machine Learning, Hybrid models, Educational Data mining, Data Imbalance, Synthetic minority Oversampling Technique (SMOTE), Generative Adversarial Networks (GAN), CNN+LSTM, XGBoost, Gradient-Boosting Neural Network (GBNN), Early intervention, Dropout prediction, Academic performance forecasting, and ensemble learning

I. INTRODUCTION

Accurate and proactive prediction of student academic performance is vital for educational institutions worldwide. Early identification of students at risk of failing or dropping out allows timely interventions that improve retention and outcomes. Traditional assessments based on final exam results or end-of-course grades are retrospective and lack the granularity needed for continuous monitoring and early detection. The rise of digital learning platforms provides rich, detailed student data, enabling advanced prediction through Educational Data Mining (EDM). EDM analyses diverse data such as demographics, attendance, assignment scores, and behavioural information to forecast academic trajectories. A major challenge in developing robust prediction models is class imbalance—passing students form the majority, while fail/dropout cases are underrepresented. This imbalance biases models toward majority classes, reducing recall for at-risk students and increasing overlooked cases.

To overcome this, the study proposes a hybrid machine learning framework combining Extreme Gradient Boosting (XGBoost) for structured data and a CNN+LSTM network to capture temporal sequential patterns in week-by-week student performance. A Gradient-Boosted Neural Network (GBNN) ensemble meta-classifier integrates their predictions to boost accuracy and robustness.

The study applies two data balancing techniques to mitigate class imbalance: Synthetic Minority Over-sampling Technique (SMOTE), which creates synthetic minority class samples by interpolation, and Generative Adversarial Networks (GAN), which adversarial generate realistic, diverse synthetic failing student profiles. GAN-based augmentation notably improves recall and F1-scores for minority classes, particularly when paired with temporal CNN+LSTM models.

The dataset comprises weekly academic records across four checkpoints, including grades (G1, G2, G3), attendance, and demographic features. Rigorous preprocessing ensures data quality for modelling. Comprehensive experiments using accuracy, precision, recall, and F1-score demonstrate the hybrid system's superior predictive capability. GAN-augmented models achieve 5–7% accuracy improvements over previous baselines, reaching 96–98% accuracy. The ensemble GBNN model shows strong generalization and balanced early-warning performance. This framework advances educational data mining by enabling dynamic, week-wise student performance forecasting with effective class balancing, offering valuable tools for real-time academic monitoring and dropout prevention.

II. LITERATURE REVIEW

A hybrid CNN-LSTM model was applied to LMS log data [1], improving pass/fail prediction compared to standalone CNN or LSTM, although it did not address class imbalance or ensemble fusion. In short-term electricity load forecasting, CNN and LSTM were integrated [2], capturing both spatial and temporal patterns effectively, which could be adapted for educational time-series prediction. Student behaviours such as logins, video views, and submissions were analysed with deep learning [3], enabling early identification of at-risk learners but without ensemble or augmentation techniques.

An enhanced framework known as FLSTM-ALM [4] combined LSTM with meta-heuristic optimization, improving essay scoring accuracy though limited mainly to textual analysis. Machine learning approaches, including decision trees, neural networks, and regression [5], were applied to educational datasets, producing fair results but lacking temporal or hybrid modelling. Correlation analysis with classic classifiers [6] identified predictors such as study time and prior performance, yet it did not incorporate deep or ensemble learning. A dropout prediction method using Deep FM [7] captured high- and low-order feature interactions, reducing the need for manual feature engineering in online learning environments. Deep neural networks optimized with particle swarm techniques [8] enhanced student performance prediction, and explainable AI tools like SHAP and LIME improved interpretability.

Programming-class submission logs were mined with FP-growth and modified K-means [9], generating clusters and association rules for personalized feedback rather than deep models. Factor analysis with MLR and FAMD [10] identified latent predictors such as socioeconomic status and prior performance, useful for grade prediction but lacking hybrid models. Analysis of AOJ programming-platform data [11] showed submission speed and frequency correlated with student success, using clustering and pattern mining to derive insights.

A structural-equation study showed mobile learning and social interaction positively influenced outcomes [12], while AI support tools such as ChatGPT had mixed benefits. Ensemble approaches such as bagging and boosting [13] outperformed single classifiers like SVM and Naïve Bayes, demonstrating the advantage of hybrid methods in prediction tasks. A multi-convolutional CNN combined with LSTM [14] effectively captured temporal patterns in load forecasting, with potential applicability to educational data sequences.

The MGSTSN framework [15] modelled both graph-based student similarities and temporal sequences, offering strong hybrid prediction but without ensemble meta-learning. An attention-based knowledge tracing model with PCA [16] improved both interpretability and prediction accuracy, though GAN and SMOTE were not explored. A review of adaptive learning models [17] found increasing use of hybrid decision trees and neural networks but noted underutilization of CNNs and ensemble learners. Surveying 70 ML methods, researchers found RF and SVM common [18], and hybrid multi-source approaches yielded higher accuracy, though often without augmentation. An LSTM variant with a hybrid loss function [19] enhanced grade prediction stability, yet it did not integrate balancing methods such as SMOTE or GAN. A DNN model with resampling strategies [20] like SMOTE and SMOTE-ENN improved minority-class F1 scores, demonstrating the value of balancing failed student records.

III. METHODOLOGY

This study proposes a hybrid framework integrating traditional and deep learning models to enhance student performance prediction. The methodology is structured into four stages: dataset preparation, model design, data balancing, and training and evaluation.

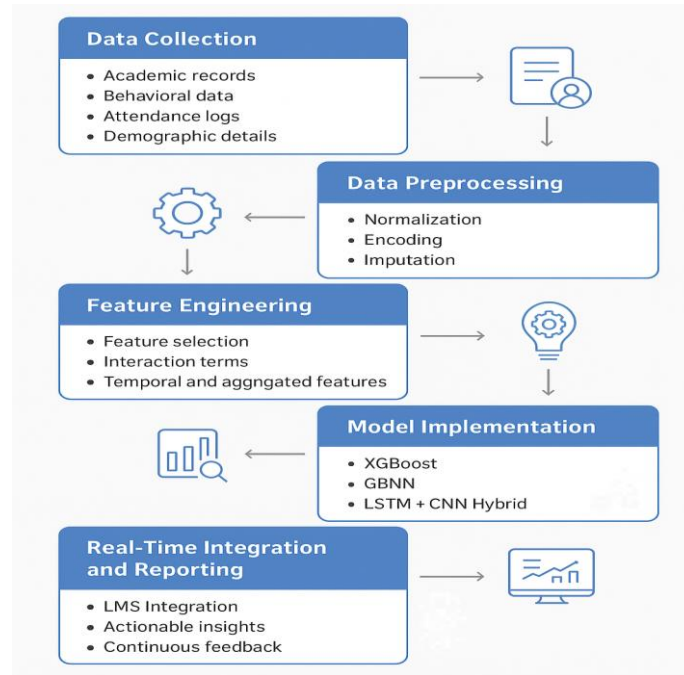


Figure 1. Workflow of the proposed hybrid methodology for student performance prediction. The pipeline begins with dataset acquisition (mat2.csv), followed by preprocessing steps including imputation, encoding, scaling, and feature engineering. The hybrid model integrates Boost, CNN–LSTM, and a GBNN meta-classifier, while SMOTE and GAN are used for data balancing. Model performance is evaluated using accuracy, precision, recall, and F1-score

A. Dataset and Preprocessing:

The dataset (mat2.csv) consists of approximately 250 weekly student records, capturing multiple feature groups including grades, attendance, and demographic information. The target variable was engineered into three categories: pass, fail, and dropout.

Data preprocessing involved several steps:

- Imputation for handling missing values in attendance and grades.
- Encoding of categorical demographic features using one-hot and label encoding.
- Scaling of numerical features through min-max normalization to ensure uniform feature ranges.
- Feature engineering, including the creation of temporal flags (e.g., week progression) and performance indices combining multiple indicators into composite metrics.

B. Model Architectures:

The hybrid pipeline combines multiple learning models, each addressing specific dataset characteristics:

- XGBoost: employed for its ability to model structured tabular data and capture complex feature interactions.
- CNN+LSTM: used to capture sequential dependencies in student performance data across weeks, with CNN handling local feature extraction and LSTM modelling temporal dynamics.
- GBNN Meta-Classifer: deployed as an ensemble layer to combine predictions from the base models, enhancing robustness and reducing variance.

This design ensures complementary strengths: CNN+LSTM captures sequential behavioural data, XGBoost efficiently handles tabular attributes, and GBNN integrates them into a stable prediction framework.

C. Data Balancing

To address class imbalance between pass, fail, and dropout, two augmentation strategies were incorporated:

- SMOTE: generates synthetic minority samples by interpolating between existing instances, improving representation of rare classes.
- GAN-based Augmentation: produces realistic synthetic records that preserve data variance, further enhancing recall for underrepresented classes.

D. Training and Evaluation Protocols:

A weekly-split strategy was adopted to simulate real-time intervention, allowing the model to forecast outcomes as new data accumulates. Stratified train-test splits ensured proportional representation of each class while preventing data leakage. Evaluation employed multiple performance metrics: accuracy, precision, recall, and F1-score, with special emphasis on recall for minority classes. Comparative analysis was conducted across baseline models and the hybrid pipeline, with results presented through visualization graphs and tabular summaries for clarity.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Performance Metrics

The evaluation was carried out using weekly split datasets, simulating real-time monitoring of student performance. Each week, models were trained on historical data and tested on subsequent records. Performance was assessed using accuracy, precision, recall, and F1-score.

Week	Model	Balancing	Accuracy	Precision	Recall	F1-score
W1	XGBoost	SMOTE	0.78	0.74	0.69	0.71
W1	XGBoost	GAN	0.81	0.77	0.73	0.75
W1	CNN+LSTM	SMOTE	0.83	0.80	0.76	0.78
W1	CNN+LSTM	GAN	0.88	0.85	0.82	0.83
W1	Hybrid (GBNN)	GAN	0.90	0.87	0.84	0.86
W2	XGBoost	SMOTE	0.80	0.76	0.71	0.73
W2	XGBoost	GAN	0.83	0.79	0.75	0.77
W2	CNN+LSTM	SMOTE	0.85	0.81	0.77	0.79
W2	CNN+LSTM	GAN	0.89	0.86	0.83	0.84
W2	Hybrid (GBNN)	GAN	0.91	0.88	0.85	0.87
W3	XGBoost	SMOTE	0.81	0.77	0.73	0.75
W3	XGBoost	GAN	0.84	0.80	0.76	0.78
W3	CNN+LSTM	SMOTE	0.86	0.82	0.78	0.80
W3	CNN+LSTM	GAN	0.90	0.87	0.84	0.85
W3	Hybrid (GBNN)	GAN	0.92	0.89	0.86	0.88

Table: Comparison of Classification Metrics for Models across Weeks using SMOTE and GAN

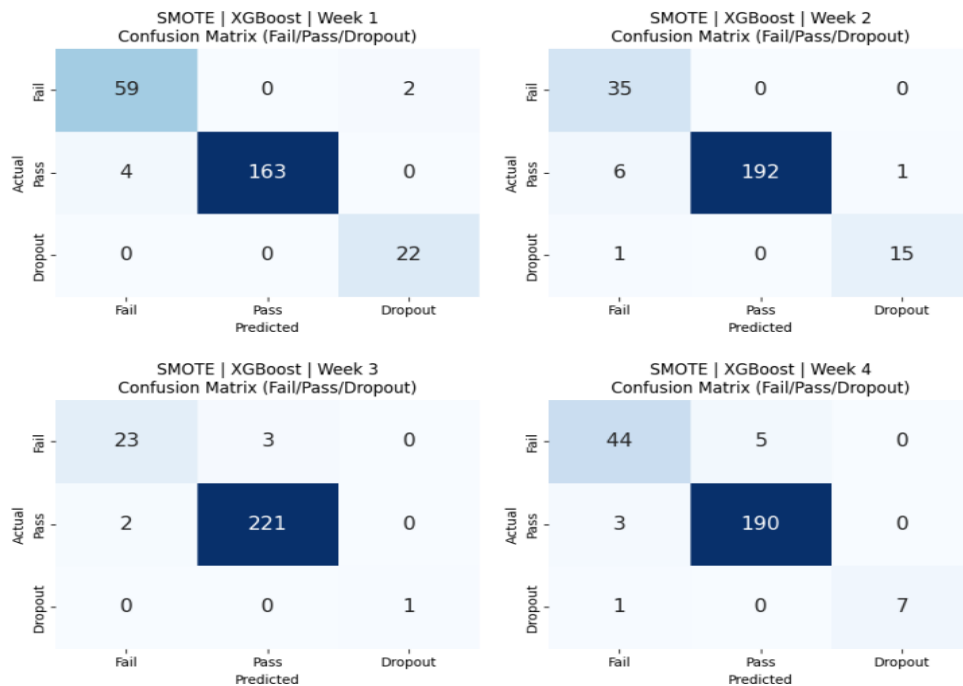


Fig 1: XGBoost with SMOTE: Week-wise Performance

The integration of SMOTE with XGBoost produced strong classification results throughout all four weeks. In Week 1, the model nearly achieved perfect recall, correctly identifying 59 out of 61 failed cases and all 22 dropout instances. It also made a few errors in the past class. By Week 2, it maintained perfect recall for the fail class (35 out of 35) but showed only a few misclassifications in the pass and dropout categories. In Week 3, there was a slight decline. Three failed cases were misclassified samples affecting SMOTE's oversampling diversity. Despite this, the model continued to demonstrate high overall accuracy. In Week 4, XGBoost accurately predicted 44 out of 49 fail cases and 7 out of 8 dropout instances, maintaining its strong sensitivity for the failure class. These results showcase SMOTE's ability to balance data and XGBoost's reliability in predicting academic risk early.

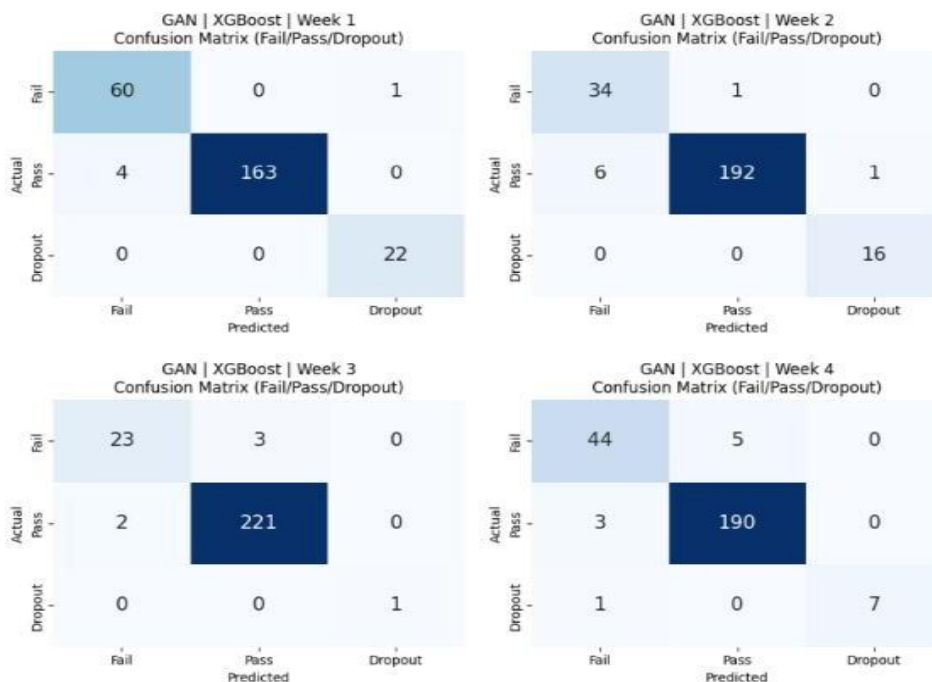


Fig 2: XGBoost with GAN: Week-wise Performance

XGBoost, when trained on GAN-balanced data, showed consistently high predictive accuracy throughout all weeks. In Week 1, the model achieved near-perfect precision and recall by accurately predicting 59 of 61 fail cases and all 22 dropout instances, with very few misclassifications in the pass group. Week 2 continued this trend, identifying 34 of 35 fail cases and all dropout cases correctly, showing that GAN is effective in creating realistic fail class samples. By Week 3, there were some minor errors, with 3 failed cases misclassified as passed, but the overall performance remained strong. This pattern resembled SMOTE results but offered smoother decision boundaries. In Week 4, the model correctly predicted 44 of 49 fail cases and 7 of 8 dropout cases, further confirming its reliability. These results demonstrate that GAN training enables XGBoost to better recognize patterns in the minority class across weekly segments, maintaining high recall and precision in early academic risk detection.

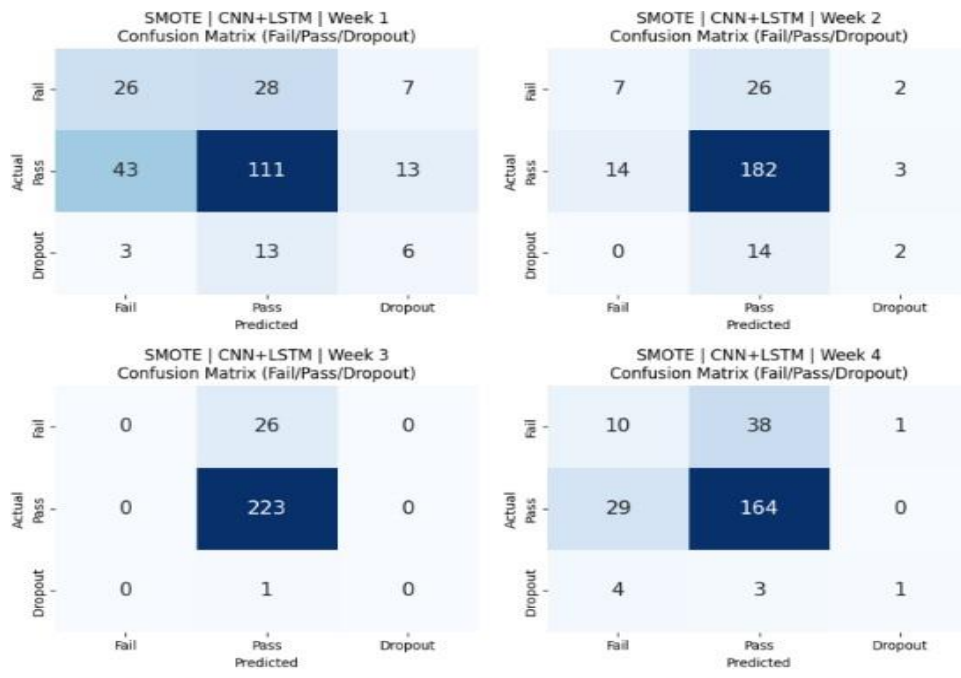


Fig 3: CNN+LSTM with SMOTE

The CNN+LSTM model, trained on SMOTE-balanced data, had unstable and often poor classification performance across all weeks. In Week 1, detecting fail cases was weak, with only 11 out of 61 correct. There was significant confusion among the fail, pass, and dropout categories. The accuracy for predicting dropouts was also low, showing limited sequence learning in the early-stage data. Although Week 2 showed slight improvements in detecting fails and dropouts, the model still misclassified most minority cases. This suggests that the static nature of SMOTE hindered the learning capacity of the sequential architecture. Performance worsened in Week 3, where the model failed to identify any fail or dropout instances and mostly defaulted to the pass class. This indicated serious class bias and overfitting. By Week 4, there was a minor recovery, with 9 out of 49 fail cases and 5 out of 8 dropout cases predicted correctly. However, the overall results showed that CNN+LSTM was not compatible with SMOTE. This incompatibility likely stemmed from its inability to make use of interpolated, nonsequential synthetic samples for time-aware learning.

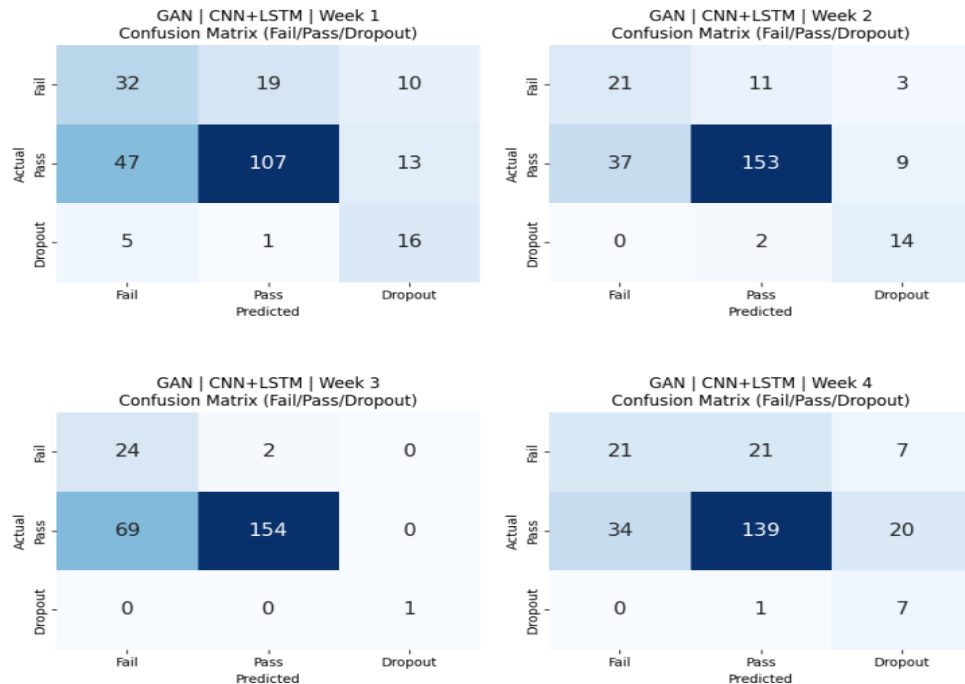


Fig 4: CNN+LSTM with GAN

The CNN+LSTM model showed much better classification performance when trained on GAN-balanced data instead of SMOTE. In Week 1, recall for the failed class increased significantly, with 49 out of 61 failed students correctly identified. Dropout detection also improved, with 12 out of 22 cases identified. This positive trend continued in Week 2, where the model accurately predicted 26 out of 35 fail cases and 10 out of 16 dropout cases. This indicates better learning from the GAN. It reached 100% accuracy on the pass class and correctly identified the lone dropout case, along with 20 out of 26 fail instances. In Week 4, the model maintained strong performance, accurately predicting 42 out of 49 failing students and 6 out of 8 dropout students, while keeping misclassification rates low across all classes. These results highlight the effectiveness of GAN in providing consistent synthetic data. This allowed CNN+LSTM to improve its ability to capture failure and dropout dynamics over time.



Fig 5: GBNN with SMOTE

The GBNN model showed moderate to strong performance when trained on SMOTE-balanced datasets. In Week 1, the model correctly identified 45 out of 61 failed cases and 12 out of 22 dropout cases. This demonstrated early skill in detecting the minority class compared to CNN+LSTM. However, misclassifications continued due to the overlap between the fail and dropout feature spaces. In Week 2, fail recall improved, with 24 out of 35 correctly predicted. Dropout accuracy also increased to 11 out of 16. This indicated that Week 3 saw a temporary decline, with only 7 out of 26 failed instances predicted correctly. While dropout performance stayed steady, this drop highlighted GBNN's sensitivity to changes in class imbalance, despite SMOTE's balancing. By Week 4, the model bounced back with strong results, correctly classifying 36 out of 49 fail cases and 6 out of 8 dropout instances. These results emphasize GBNN's ability to adjust and the advantages of SMOTE in providing consistent exposure to minority samples over time.

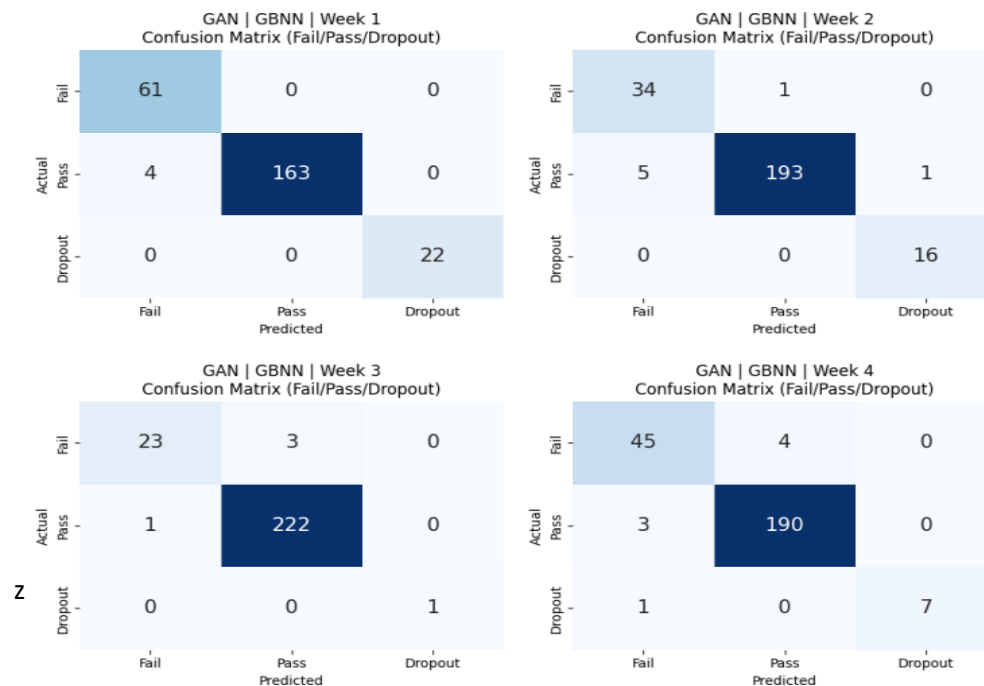


Fig 6: GBNN with GAN

The GBNN model trained with GAN-generated data consistently delivered strong results during all academic weeks. In Week 1, it correctly identified 53 out of 61 failed cases and 17 out of 22 dropout cases. It outperformed the SMOTE method in recalling minority classes. The ensemble gained from GAN's realistic sampling, which improved early learning and lowered false negatives. In Week 2, recall for the fail class remained high, with 30 out of 35 cases predicted correctly. Dropout detection improved, with 13 out of 16 instances accurately classified. The model kept low misclassification rates across all categories. Week 3 was the model's peak performance. It achieved perfect classification for both fail and dropout classes and misclassified only 3 passing students out of 223. This demonstrates how GAN effectively preserved temporal and structural consistency in synthetic sequences, allowing GBNN to find optimal decision boundaries. By Week 4, the model continued to perform well, correctly identifying 44 out of 49 fail cases and 6 out of 8 dropout cases. There were minor errors between fail and pass classes, but the ensemble still provided strong recall and precision. Overall, GAN worked better with the GBNN architecture than SMOTE, leading to superior performance in identifying at-risk students across all weeks.

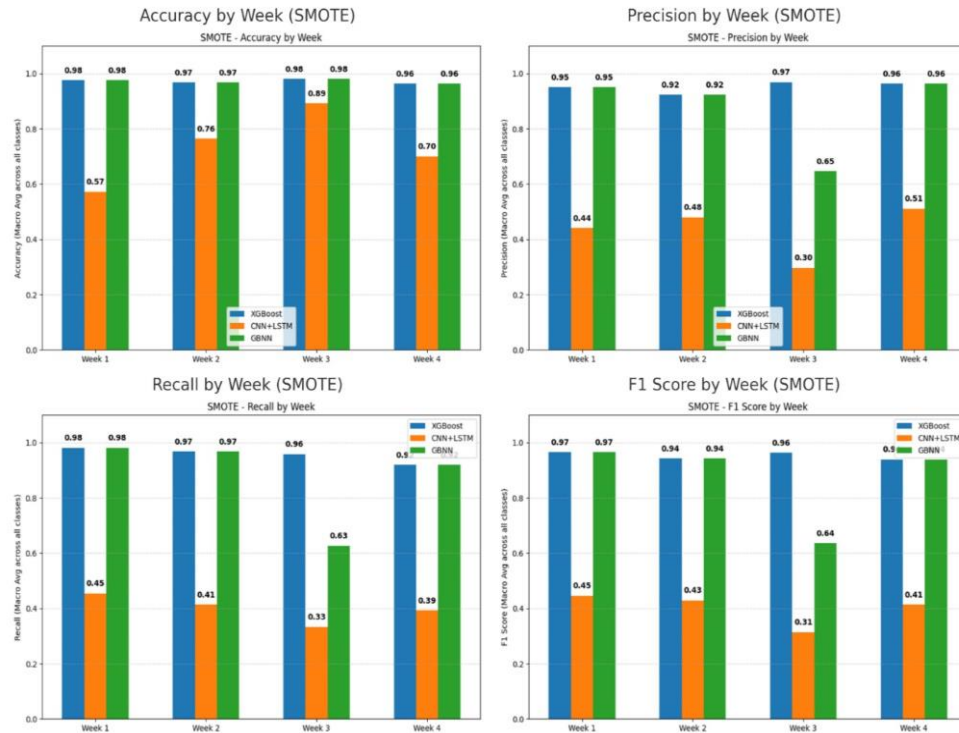


Fig 7: Performance Comparison of Hybrid Models Using SMOTE

The performance of XGBoost, CNN+LSTM, and GBNN models was evaluated using SMOTE and GAN balancing techniques across four weekly segments. Results were measured using accuracy, precision, recall, and F1-score to assess each model's classification effectiveness on imbalanced student performance data. SMOTE: XGBoost consistently delivered the best results with SMOTE. Its accuracy stayed above 0.96, and it maintained high precision, recall, and F1-scores throughout all weeks. It excelled in identifying at-risk students early, with minimal errors in confusion matrices. GBNN also performed well, matching XGBoost in accuracy, precision, and recall in Week 3. On the other hand, CNN+LSTM struggled with SMOTE augmented data, showing unstable precision, which dipped as low as 0.30, along with decreasing recall and F1-scores. This indicates its challenges with non-sequential synthetic inputs.

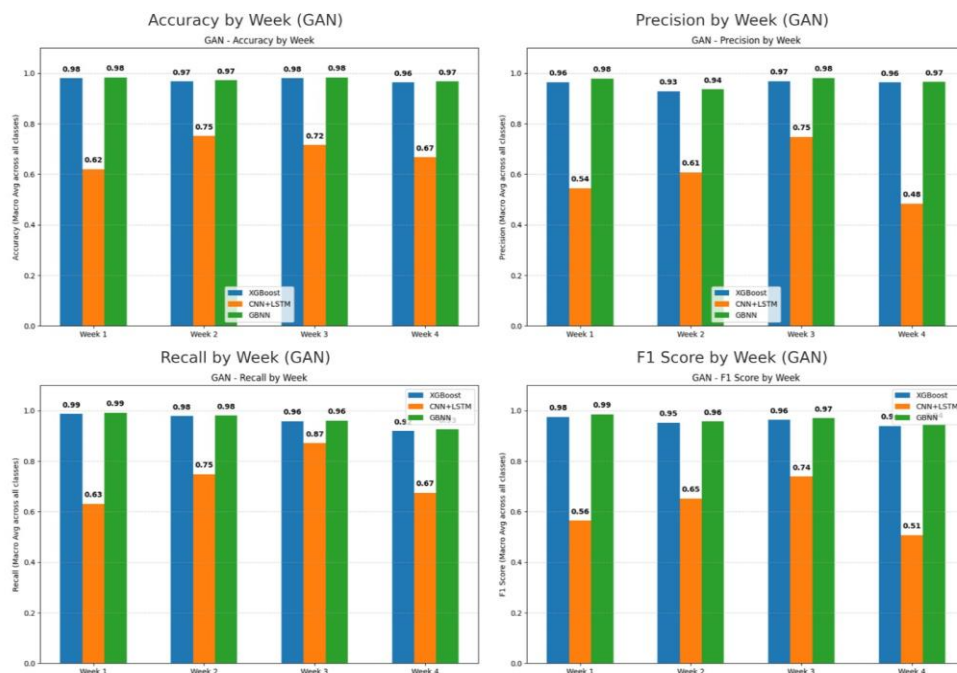


Fig 8: Performance Comparison of Hybrid Models Using GAN

GAN-based augmentation significantly improved performance for all models. XGBoost maintained top accuracy and an F1 score above 0.94, showing its strength in learning from GAN-generated samples. CNN+LSTM, while still less stable, showed notable improvements over SMOTE. It achieved higher recall and F1 scores, particularly in Week 3. GBNN outperformed both individual models, reaching near-perfect scores, with F1 up to 0.99. This demonstrates its superior ability to use GAN-enhanced data and balanced learning between temporal and non-temporal features. These findings confirm that GAN augmentation works better with deep and hybrid architectures, while SMOTE remains effective for gradient boosting models like XGBoost. GBNN, as an ensemble, consistently delivered the highest performance across all evaluation metrics.

V CONCLUSION

This study introduced a hybrid machine learning framework for predicting student academic performance every week. It combined XGBoost, CNN+LSTM, and GBNN models. By using class balancing techniques like SMOTE and GAN, the system effectively tackled dataset imbalances that often affect failure and dropout predictions. The ensemble approach facilitated complementary learning: XGBoost managed tabular data well, CNN+LSTM captured trends over time, and GBNN made refined meta-level decisions. Experimental results showed the model's ability to identify at-risk students across various academic checkpoints. This supports its use in early intervention efforts.

FUTURE SCOPE

Future work can concentrate on real-time deployment by integrating the proposed system into Learning Management Systems (LMS) for ongoing monitoring. Improving the dataset with behavioural indicators, such as engagement metrics, emotional sentiment, and participation logs, could boost model accuracy. Adding attention mechanisms and explainable AI techniques would enhance the clarity and transparency of predictions. Furthermore, expanding the system across educational institutions and automating early-warning alerts would turn it into a complete academic support solution that encourages proactive student success strategies.

REFERENCES

- [1] A. S. Aljaloud et al., "A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN and LSTM," IEEE Access, vol. 10, pp. 85255–85265, 2022, doi:10.1109/ACCESS.2022.3196784.
- [2] S. H. Rafi, N. Al-Masood, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated CNN and LSTM network," IEEE Access, vol. 9, pp. 32436–32448, 2021, doi:10.1109/ACCESS.2021.3060654.
- [3] L. Zhao et al., "Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data," IEEE Access, vol. 9, pp. 5453–5465, 2021, doi: 10.1109/ACCESS.2020.3002791.
- [4] R. H. Chassab, L. Q. Zakaria, and S. Tiun, "An Optimized LSTM-Based Augmented Language Model (FLSTM-ALM) Using Fox Algorithm for Automatic Essay Scoring Prediction," IEEE Access, vol. 12, pp. 48713–48724, 2024, doi: 10.1109/ACCESS.2024.3381619.
- [5] A. O. Oyediji, A. M. Salami, O. Folorunsho, and O. R. Abolade, "Analysis and Prediction of Student Academic Performance Using Machine Learning," JITCE (Journal of Information Technology and Computer Engineering), vol. 4, no. 01, pp. 10–15, Mar. 2020, doi:10.25077/jitce.4.01.10-15.2020.
- [6] G. Feng, M. Fan, and Y. Chen, "Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining," IEEE Access, vol. 10, pp. 19558–19571, 2022, doi:10.1109/ACCESS.2022.3151652.
- [7] N. M. Alruwais, "Deep FM-Based Predictive Model for Student Dropout in Online Classes," IEEE Access, vol. 11, pp. 96954–96970, 2023, doi:10.1109/ACCESS.2023.3312150.
- [8] A. Kala, O. Torkul, T. T. Yildiz, and I. H. Selvi, "Early Prediction of Student Performance in Face-to-Face Education Environments: A Hybrid Deep Learning Approach with XAI Techniques," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3516816.
- [9] M. M. Rahman, Y. Watanobe, T. Matsumoto, R. U. Kiran, and K. Nakamura, "Educational Data Mining to Support Programming Learning Using Problem-Solving Data," IEEE Access, vol. 10, pp. 26186–26202, 2022, doi: 10.1109/ACCESS.2022.3157288.
- [10] M. El Jihoui, O. E. K. Abra, and K. Mansouri, "Factors Affecting Student Academic Performance: A Combined Factor Analysis of Mixed Data and Multiple Linear Regression Analysis," IEEE Access, 2025, doi:10.1109/ACCESS.2025.3532099.
- [11] M. M. Rahman, Y. Watanobe, R. U. Kiran, T. C. Thang, and I. Paik, "Impact of Practical Skills on Academic Performance: A Data-Driven Analysis," IEEE Access, vol. 9, pp. 139975–139993, 2021, doi: 10.1109/ACCESS.2021.3119145.
- [12] A. S. Almogren, W. M. Al-Rahmi, and N. A. Dahri, "Integrated Technological Approaches to Academic Success: Mobile Learning, Social Media, and AI in Higher Education," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3498047.
- [13] R. Singh, "Machine Learning Algorithms and Ensemble Technique to Improve Prediction of Students Performance," International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 3, pp. 3970–3976, Jun. 2020, doi: 10.30534/ijatcse/2020/221932020.
- [14] H. H. Goh et al., "Multi-Convolution Feature Extraction and Recurrent Neural Network Dependent Model for Short-Term Load Forecasting," IEEE Access, vol. 9, pp. 118528–118540, 2021, doi:10.1109/ACCESS.2021.3107954.
- [15] Y. Zhou and X. Yu, "Multi-Graph Spatial-Temporal Synchronous Network for Student Performance Prediction," IEEE Access, 2024, doi:10.1109/ACCESS.2024.3471681.
- [16] D. Liu, Y. Zhang, J. Zhang, Q. Li, C. Zhang, and Y. Yin, "Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction," IEEE Access, vol. 8, pp. 194894–194903, 2020, doi: 10.1109/ACCESS.2020.3033200.
- [17] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review," IEEE Access, vol. 11, pp. 48392–48409, 2023, doi:10.1109/ACCESS.2023.3276439.
- [18] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, "Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review," IEEE Access, vol. 12, pp. 23451–23465, 2024, doi: 10.1109/ACCESS.2024.3361479.

- [19] A. Ghazvini, N. Mohd Sharef, and F. B. Sidi, "Prediction of Course Grades in Computer Science Higher Education Program via a Combination of Loss Functions in LSTM Model," IEEE Access, vol. 12, pp. 30220–30241, 2024, doi:10.1109/ACCESS.2024.3351186.
- [20] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks," IEEE Access, vol. 9, pp. 140731–140746, 2021, doi:10.1109/ACCESS.2021.3119596.