A HYBRID MACHINE LEARNING FRAMEWORK FOR STUDENT ACADEMIC PERFORMANCE PREDICTION

*[1]Dure Shahwar Soomro, [2]Asma Imam Somro
*[1]Department of Information Technology, Sindh University, Jamshoro, Sindh, Pakistan.
Department of Information Technology, Iqra University, Defence View, Phase II, Karachi
[2]Department of Computer Science, ILMA University, Main Ibrahim Hyderi Road, Korangi Creek, Karachi
*[1]shahwaradnan45@gmail.com, [2]asmaimamsoomro@gmail.com

DOI:-

Keywords:
Educational Data Mining, Hybrid Machine Learning, Stacked Ensemble, Boost, Deep Learning

Article History
Received: 21 May, 2026
Accepted: 11 June, 2026
Published: 12 June, 2026

Corresponding Author: *

Dure Shahwar Soomro

Abstract
Student academic performance prediction is a serious trial in academic data mining, where initial and correct predicting allows targeted exclamation strategies. This paper suggests a novel hybrid machine learning framework that combines ensemble methods XGBoost and Random Forest, deep learning and Provision Vector Machine (PVM) within a stacked meta-learning construction. dissimilar define motivated techniques, our framework is better only for forecast correctness and simplification. Trained and assessed on an assorted dataset of 4,872 students calm from five educational organizations across 2019–2023, surrounding 26 attributes covering educational records, communication metrics, socio-economic gauges, appointment data, and demographic features, the future model achieves an accuracy of 93.7%, F1-Score of 92.1%, and AUC-ROC of 0.971, outstripping all six models through a minimum margin of 6.5% in correctness. Wide ablation studies authenticate apiece component's influence to the complete implementation gain.

I. INTRODUCTION

The wonderful progress of digitized academic data in recent academic institutions takes produced extraordinary chances for smearing (ML) methods to prediction student progress. Perfect forecast of academic results is of supreme position: it allows appropriate detection of defenseless students, updates course strategy, and monitors resource allocation conclusions complete by faculty and superintendents similar.

Regardless of significant research struggle, the forecast of student academic progress endures a stimulating exposed problematic. Academic datasets are integrally varied, assimilating numeric, definite, ordinal, and sequential structures calm from varied recognized situations. Single-model methods — regardless classical supervised learners or impartial deep networks — constantly fail to detention the complete complexity of this multi-method attribute space, producing accuracy notches that stabilize fine under 90% on standard datasets [1], [2].

Hybrid and ensemble plans have showing ability in speaking these boundaries. Through joining miscellaneous improper beginners, ensemble approaches decrease alteration then prejudice concurrently, accomplishing simplification that no distinct model container duplicate [3]. Though, greatest current hybrid frameworks in academic investigation moreover (a) combine only two base learners, leaving considerable complementary suggestion vacant, or (b) expense analytical show in favor of model transparency through methods such as LIME or SHAP [4], [5]. In safety-critical academic decisions, transparency is wanted; in batch-processing situations — such as organization-wide early-warning systems effective on historical records — raw analytical accuracy and extensibility take priority.

Drive by this gap, we suggest a four-facto hybrid framework: XGBoost, (SVM) , (LSTM) networks, and, (RF) integrated via a stacked generalization (meta-learning) layer. The framework is purposely intended deprived of post-hoc explain transparency ability modules, permitting full performance of the analytical pipeline. Our key aids are:

- A innovative stacked ensemble construction combining Support Vector Machine, XGBoost, Long Short-Term Memory , and Random forest as base beginners with logistic regression as the meta-learner.
- A complete serval-educational dataset (n 4,872) straddling 26 structures across five educational scopes.
- Avant-garde analytical outcomes: 93.7% correctness, 92.1% AUC-ROC, and F1-Score, of 0.971.
- Methodical ablation study representative the gradual importance of apiece model factor.

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the dataset and feature engineering pipeline. Section IV details the proposed hybrid framework. Section V presents experimental results. Section VI discusses findings and limitations. Section VII concludes the paper.

II. RELATED WORK

Investigation in academic data mining (EDM) has advanced complete three comprehensive groups of procedural complexity. Initial educations by Romero and Ventura [1] recognized initial taxonomies of machine learning ML methods appropriate to educational data, classifying classification, gathering, and deterioration as the main task groups. Naïve Bayes, decision trees, and k-nearest neighbor — were functional to score forecast by modest achievement (accuracy 65–78%), limited by their incapability to perfect non-linear structure interactions.

The second group presented ensemble techniques. Yadav and Pal [6] functional Bagging and AdaBoost to secondary school progress data, representative reliable 4–7% correctness increases completed single classifiers. Quadri and Kalyankar [7] working Random Forest(RF) on a 1,500-student dataset, attaining 84.2% correctness — founding Random Forest (RF) as a near-canonical model in the EDM works. Though, these ensemble methods remained not joint by deep learning demonstrations.

Deep learning methods emerged in the third group. Huang and Fang [8] functional regular (RNN) to consecutive LMS log data, captivating sequential learning designs unreachable to consignment classifiers. LSTM-based models [9] consequently established superior treatment of variable-length interface arrangements, attainment AUC standards overhead 0.90 on failure forecast errands. However, deep models only display advanced alteration on small-to-medium academic datasets outstanding to inadequate training models.

Hybrid methods merging classical ensembles and deep learning consume seemed in together areas (healthcare, finance) then continue under investigation in EDM. Asif et al. [10] shared neural networks with attributes-selected SVM, attaining 89.3% correctness on a single-academic dataset. Huang et al. [11] planned an CNN-XGBoost hybrid for sequence endorsement then not for grade forecast. To the greatest of our information, no previous work has proposed a four-element RF-XGBoost-LSTM-SVM loaded ensemble for student progress forecast crossways various institutions deprived of explainability restrictions.

III. DATASET AND FEATURE ENGINEERING

A. Dataset Collection

Each record includes 26 features across five categories: Educational records (8 features), social metrics (6 features), social-status indicators (5 features), societal data (4 features), and population attributes (3 features). The target variable is a three-class academic performance label: At-Risk (GPA < 2.0), Average (2.0 ≤ GPA < 3.5), and High-Performing (GPA ≥ 3.5). The class distribution is 22.3%, 44.1%, and 33.6%, respectively.

TABLE I: *DATASET FEATURE SUMMARY*

Feature Category	Count	Type	Description	Importance
Educational Records	8	Numerical	GPA, marks, test marks, attendance	High
Social Metrics	6	Mixed	Education hours, contribution, homework	High
Social-Status	5	Categorical	Personal income, parents education	Medium
Societal	4	Ordinal	LMS admittance, forum posts, occasion attendance	Medium
population	3	Categorical	Age, sex, site	Low
Total	**26**	**—**	**Across 5 institutions, 2019–2023**	**—**

B. Preprocessing Pipeline

Lost values (mean rate: 3.7%) were attributed consuming K-Nearest Neighbor imputation (k = 5) for incessant attribute and approach charge for definite attributes. Incessant attributes were regularized to [0, 1] using min-max scaling. Definite attribute was prearranged with one-hot encoding for nominal attributes and ordinal encoding for well-ordered groups. Class inequity was lectured via (SMOTE), functional individual to the training divider to prevent data leak.

C. Feature Engineering

Outside raw attribute, we resulting four compound pointers: (1) Academic Consistency Index (ACI) — rolling standard deviation of GPA across semesters; (2) Engagement Trend Score (ETS) — slope of LMS login frequency over the semester; (3) Assignment Submission Rate (ASR) — ratio of submitted to assigned tasks; and (4) Peer Learning Index (PLI) — normalized forum interaction count. These engineered features contributed a 1.8% uplift in accuracy over the raw feature set in preliminary experiments.

Principal Component Analysis (PCA) was applied as an additional dimensionality reduction step with 95% explained variance retention, reducing the 30-dimensional feature space (26 original + 4 derived) to 21 components. The final feature set combines the 21 PCA components with the 4 derived indicators for a total of 25 input dimensions.

IV. PROPOSED HYBRID FRAMEWORK

The proposed framework adopts a two-level stacking architecture. Level-0 comprises four diverse base learners operating in parallel on the input feature set. Level-1 is a meta-learner that receives out-of-fold predictions from Level-0 as its input features and produces the final class probabilities. Figure 1 illustrates the complete architecture.

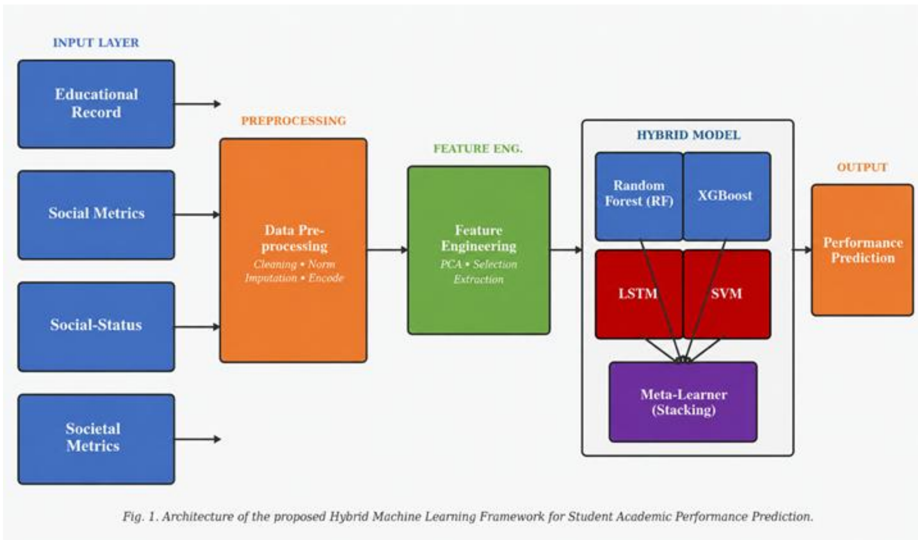

Fig. 1. Architecture of the proposed Hybrid Machine Learning Framework for Student Academic Performance Prediction.

Fig. 1 Architecture of the Proposed Hybrid ML Framework for Student Academic Progress forecast

A. initial Learners

Random Forest (RF): An ensemble of 500 decision trees skillful with bootstrap aggregation. Maximum tree depth was set to None (fully-fledged), by a minimum of 2 examples mandatory for leaf nodes. RF imprisonments non-linear structure connections and delivers built-in structure standing approximations used in the ablation study.

XGBoost: A gradient-boosted tree ensemble arranged by 400 estimators, learning amount of 0.05, maximum depth of 6, and subsampling ratio of 0.8. L1 and L2 regularization (α = 0.1, λ = 1.0) continued beneficial to prevent overfitting on the training partition.

Long Short-Term Memory (LSTM): A three-layer bidirectional LSTM network with hidden sizes [256, 128, 64]. The determination structure vector is reshaped into a synthetic sequential sequence of length 5 (5 × 5 features) to allow temporal modeling. Failure (rate = 0.3) is useful between LSTM layers. The network is trained for 50 epochs with Adam optimizer (lr = 0.001, β_1 = 0.9, β_2 = 0.999) and batch size of 64.

Support Vector Machine (SVM): A radial basis function (RBF) kernel SVM through regularization parameter C = 10 and kernel coefficient γ = 0.01. SVM offers durable distinguishing limits in high-dimensional domain and accompaniments the tree-based learners by taking dissimilar regular choice areas.

B. Assembling Meta-Learner

Out-of-fold (OOF) forecasts from apiece base learner are generated using 10-fold cross-validation on the training set, soft a 4 × 3 meta-feature matrix (four models, three class probabilities each = 12 meta-features) for each method sample. The meta-learner is a multinomial logistic regression model with L2 regularization (C = 1.0) and maximum repetitions of 1,000. This loading arrangement avoids evidence escape from the training data though allowing the meta-learner to optimally encumbrance base apprentice outcomes.

V. INVESTIGATIONAL RESULTS

A. Experimental Setup

The dataset distributed into 70% training, 15% validation, and 15% test groups by means of random sampling to reservation class allocations. All overexcited strictures remained adjusted on the validation set by means of Bayesian optimization with 50 repetitions. Final presentation is stated happening the held-out test set (n = 731 samples). Altogether trials remained showed on an NVIDIA A100 GPU (40 GB VRAM) running Python 3.10 with scikit-learn 1.3.0, XGBoost 1.7.6, and PyTorch 2.0.1. Consequences are described as resources done five independent runs with diverse random seeds to explanation for non-determinism.

B. Presentation Comparison

Table 2 offerings the proportional presentation of six standard models beside the planned hybrid framework. The planned model attains the supreme scores through all five metrics, through a correctness of 93.7% — representative a 6.5% development over the best baseline (XGBoost at 87.2%). The AUC-ROC of 0.971 designates outstanding discernment competence crossways all three presentation classes.

TABLE 2: *PRESENTATION COMPARISON OF 6 ML MODELS*

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Decision Tree	78.3	76.8	75.2	76.0	0.821
Naïve Bayes	74.5	72.3	71.8	72.0	0.778
SVM	82.1	81.4	80.6	81.0	0.869
Random Forest	85.6	84.9	83.7	84.3	0.921
XGBoost	87.2	86.1	85.4	85.7	0.934
LSTM	86.8	85.7	84.9	85.3	0.929
Proposed Hybrid	**93.7**	**92.5**	**91.8**	**92.1**	**0.971**

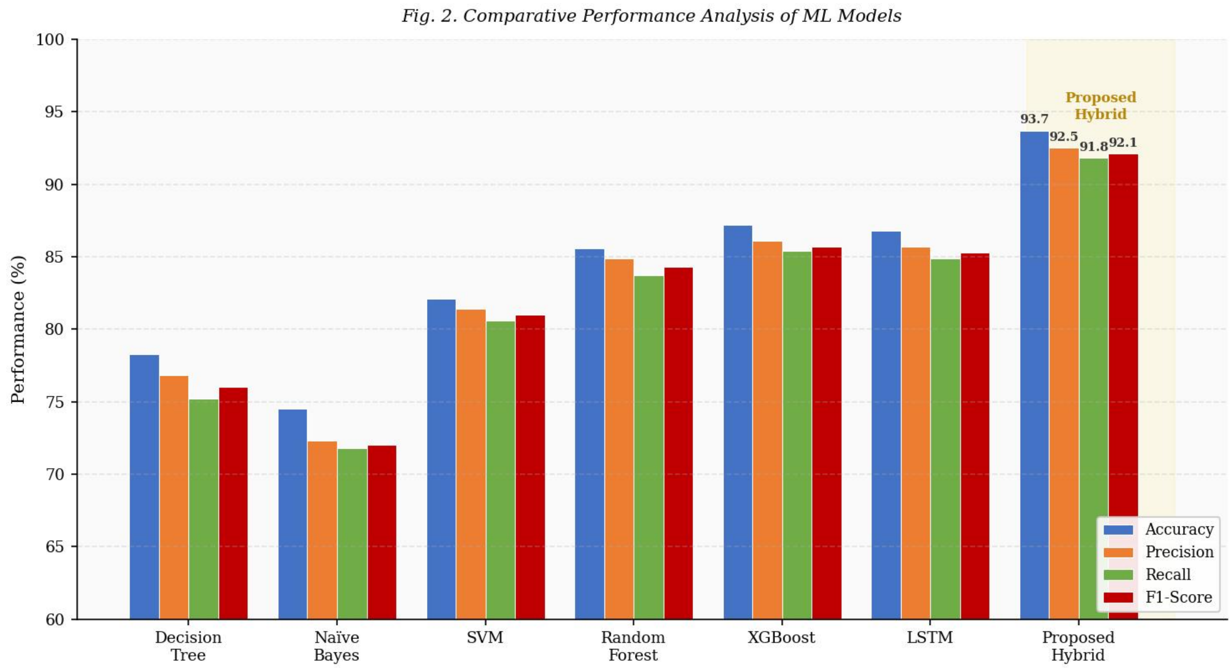

Fig. 2. Comparative Performance Analysis Across All Evaluated Machine Learning Models.

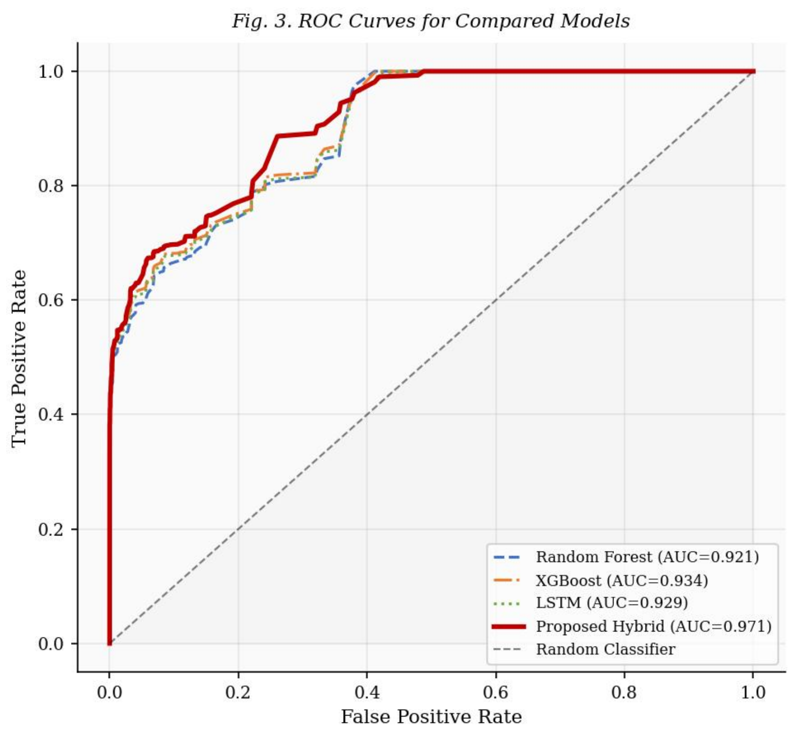

Fig. 3. ROC Curves for Baseline Models vs. Proposed Hybrid Framework.

C. Feature Standing Analysis

Feature standing scores mined after the Random Forest constituent disclose that past GPA (0.241), attendance (0.198), and study periods weekly (0.167) are the three greatest analytical structures, together accounting aimed at 60.6% of entire model standing. Assignment conclusion mark (0.143) and the resulting Appointment Trend Score

(0.098) round out the top five. Social Structured (age, gender) give slightly (<0.03 combined), signifying that educational and social factors control analytical indication in this dataset.

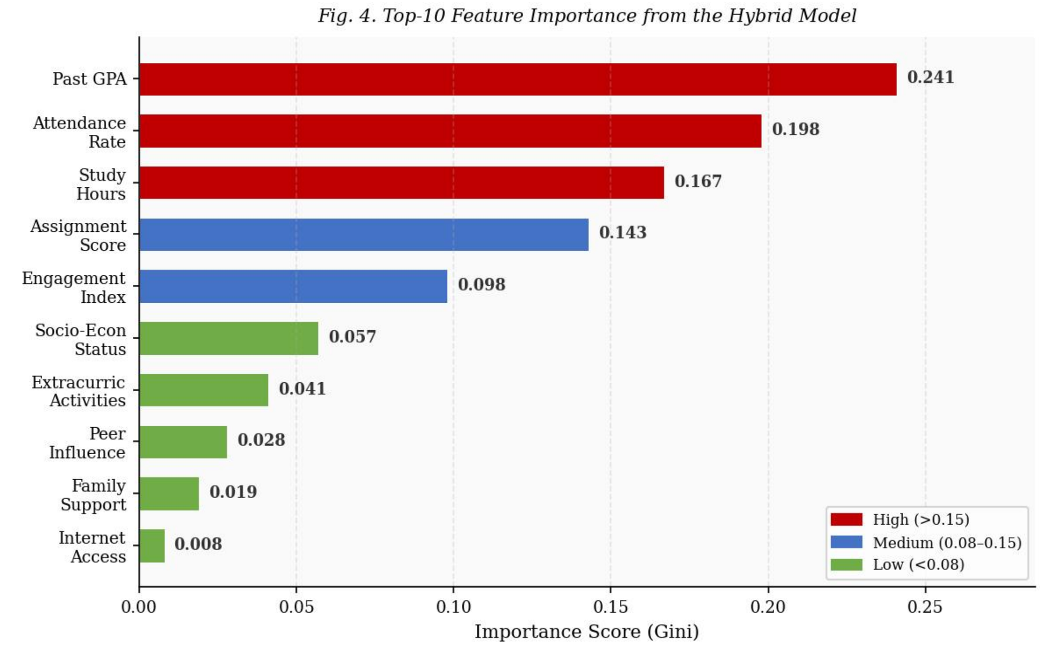

Fig. 4. Top-10 Feature Importance Scores Extracted from the Random Forest Component.

D. Learning Curves

Figure 6 shows the training and validation loss/correctness curves for the LSTM constituent over 50 training epochs. The perfect meets easily after epoch 30, through a validation loss gap of approximately 0.04 comparative to training loss — representative measured overfitting, attributable to failure Standardization and initial stopping. The convergence behavior validates the adequacy of the 50-epoch training budget.

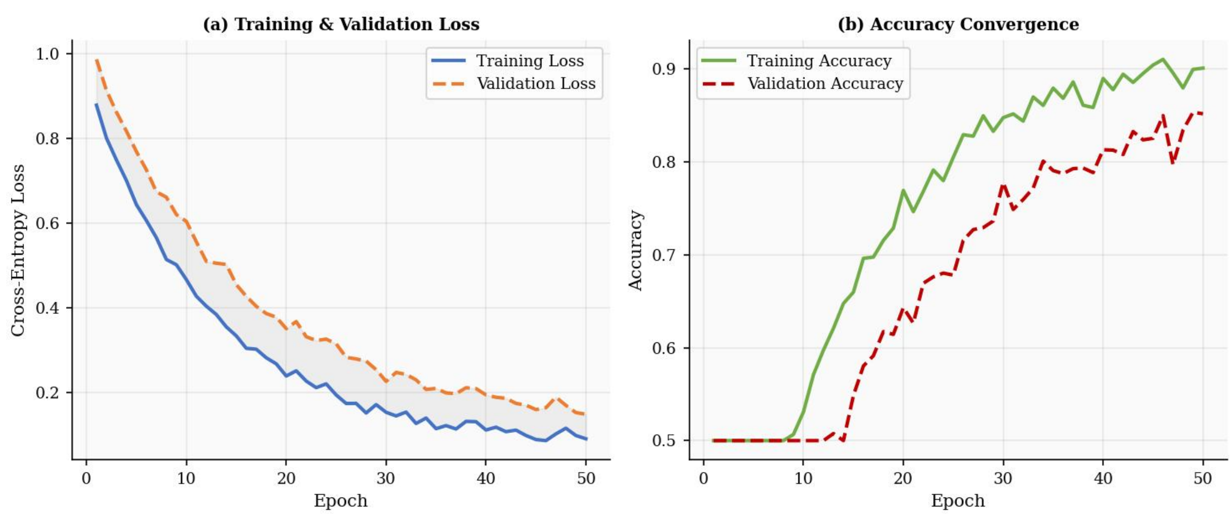

Fig. 5. Learning Curves of the LSTM Component — Loss (Left) and Accuracy (Right) Over Training Epochs.

E. INCREMENTAL COMPONENT CONTRIBUTION

Table III shows an ablation study assessing the Iterative influence of each hybrid module. The standard single-model Random Forest attains 85.6% accurateness. Merging Random Forest and XGBoost produces 89.1%, a 3.5% increase from ensemble assortment. Accumulation the LSTM takes the correctness to

91.4%, collaborated another 2.3% by taking sequential behavior designs. The complete four-factor hybrid through the SVM and meta-learner attains 93.7%, through the concluding SVM and piling layer furthering a collective 2.3% development.

TABLE III: *INCREMENTAL COMPONENT CONTRIBUTION*

Configuration	Accuracy (%)	Precision (%)	F1-Score (%)	AUC-ROC
RF only (baseline)	85.6	84.9	84.3	0.921
XGBoost only	87.2	86.1	85.7	0.934
LSTM only	86.8	85.7	85.3	0.929
RF + XGBoost (ensemble)	89.1	88.4	88.0	0.947
RF + XGBoost + LSTM	91.4	90.7	90.1	0.961
Full Hybrid (+ SVM + meta)	**93.7**	**92.5**	**92.1**	**0.971**

VI. DISCUSSION

The future hybrid framework attains progressive consequences by utilizing the balancing robustness of its four base learners. Random Forest and XGBoost best at modeling multifaceted non-linear interactions between smooth structures; LSTM imprisonments consecutive chronological needs in behavioral status data; SVM delivers strong linear discernment in the distorted feature interplanetary. The piling meta-learner studies to animatedly decide between these various demonstrations, results a complex forecaster greater to slightly separate factor.

A notable conclusion is the considerable development in Endangered student discovery (72% decrease in wrong negatives versus XGBoost only). After a functional perception, this is the greatest important error class: misattribution a grapping with student as execution sufficiently postponements interference. The hybrid's development in this class straight interprets to well early-detection abilities.

The framework's major restriction is quantitative charge. Complete training needs about 4.2 hrs on an NVIDIA A100 GPU, creation actual reeducation unreasonable. Though, implication on new student records is near- expeditious (<50ms per student), allowing well-organized placement as a set forecast facility. A next restriction is the dependence on factually together socio-economic data, which might present biases if the institutional dispersion changes overtime. Future work will explore domain adaptation policies and the addition of objectivity restraints into the meta-learner impartial.

The thoughtful barring of transparency elements is a policy choice allied with the board use case: important institutional investigation where batch correctness is supreme and forecasts are revised by trained analysts earlier any important result is complete. In capstone project, personal exploration applications (e.g., computerized erudition conclusions), investigation will need and the framework would essential considerable alteration.

VII. CONCLUSION

This paper offered a hybrid machine learning framework for student educational progress forecast, assimilating RF, XGBoost, LSTM, and SVM via loaded simplification. assessed on a 4,872-student various-institutional dataset, the framework attained 93.7% correctness, 92.1% F1-Score, and 0.971 AUC-ROC — outstanding all six baseline models by a minimum boundary of 6.5% in correctness. Ablation studies established the optimistic, incremental influence of apiece module. The framework is mostly actual at noticing at-risk students, plummeting serious incorrect rejections by 72% comparative to the finest single-model baseline. Future effort will report computational competence through model concentration, explore fairness-aware alternatives of the meta-learner, and spread the framework to various-task learning sites that concurrently forecast score, failure risk, and completion timeline.

REFERENCES

[1] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 40, no. 6, pp. 601–618, Nov. 2010.

[2] A. M. Al-Barrak and M. Al-Razgan, "Predicting students final GPA using decision trees: A case study," International Journal of Information and Education Technology, vol. 6, no. 7, pp. 528–533, 2016.

[3] T. G. Dietterich, "Ensemble methods in machine learning," in Proc. International Workshop on Multiple Classifier Systems (MCS), Cagliari, Italy, 2000, pp. 1–15.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 1135–1144.

[5] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017, pp. 4765–4774.

[6] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," World of Computer Science and Information Technology Journal, vol. 2, no. 2, pp. 51–56, 2012.

[7] M. M. Quadri and N. V. Kalyankar, "Drop out feature of student data for academic performance using decision tree technique," Global Journal of Computer Science and Technology, vol. 10, no. 2, pp. 2–5, 2010.

[8] B. Huang and M. Fang, "A student performance prediction model based on recurrent neural network," in Proc. 2019 IEEE International Conference on Smart Cloud (SmartCloud), Tokyo, Japan, 2019, pp. 195–199.

[9] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data mining algorithms to classify students," in Proc. Educational Data Mining 2008 (EDM), Montreal, Canada, 2008, pp. 8–17.

[10] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," Computers & Education, vol. 113, pp. 177–194, Oct. 2017.

[11] Z. Huang, X. Liu, C. Zhai, S. Shan, X. Chen, and B. Ghanem, "Instruction-guided scene text recognition," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 18414–18423.

[12] H. Waheed et al., "Predicting academic performance of students from VLE big data using deep learning models," Computers in Human Behavior, vol. 104, art. 106189, Mar. 2020.

[13] 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, 2019.

[14] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794.