

A Hybrid Deep Learning Model for Early Prediction of Student Academic Performance in Higher Education

¹Aby Rose Varghese*

²Dr. Susheel George Joseph

¹ Assistant Professor, Kristu Jyoti College of Management & Technology, Changanacherry, Kerala, India

² Professor, Faculty of Computer Applications, Kristu Jyoti College of Management & Technology,
Changanacherry, Kerala, India

Corresponding author: Aby Rose Varghese

DOI: <https://doi.org/10.63084/cognexus.v2i2.304>

Abstract

Early identification of students who may experience academic difficulty can support timely, targeted intervention in higher education. This study evaluates a hybrid prediction framework that combines a support vector machine (SVM) with a polynomial kernel, a long short-term memory (LSTM) component with two hidden units, and a decision-tree component. The study protocol uses cohort-wise 70:15:15 train-validation-test partitioning so that students from different semester, college, course, and course-type cohorts are not mixed across partitions. Hyperparameters were selected using random search, and the component baselines were evaluated under the same protocol. The finalized prediction file supplied for this study contains 4,800 student-level observations, three performance classes (0, 1, and 2), and predictions from SVM, LSTM, decision tree, and the proposed hybrid system. On this 4,800-record prediction set, the hybrid achieved 94.10% accuracy (95% bootstrap CI: 93.44–94.75%), balanced accuracy of 94.05%, macro-F1 of 93.57%, Cohen's kappa of 0.906, and Matthews correlation coefficient of 0.906. The confusion matrix contains 1,371, 2,245, and 901 correct predictions for classes 0, 1, and 2, respectively, with 283 total errors. The supplied component predictions yielded accuracies of 82.94% (SVM), 85.04% (LSTM), and 79.19% (decision tree), providing an empirical comparative reference. A critical data-audit finding is that the supplied CSV does not contain cohort/year identifiers, model probabilities, or a trained-model checkpoint; consequently, native SHAP values and independent construction of year-specific results cannot be claimed from this file alone. This distinction is made explicitly to avoid overstating evidence.

Keywords: student academic performance; educational data mining; learning analytics; hybrid machine learning; SVM; LSTM; decision tree; explainable AI; SHAP; early prediction

Introduction

Higher-education institutions increasingly use learning analytics and educational data mining to identify students who require academic support before poor outcomes become irreversible. Early prediction can help educators prioritize advising, remedial instruction, mentoring, attendance follow-up, and other interventions. Reviews of the field show that student-performance prediction has been addressed with statistical methods, classical machine learning, ensemble learning, and deep learning, but substantial heterogeneity remains in datasets, labels, prediction horizons, evaluation protocols, and reported metrics [1–4].

The practical difficulty is not simply obtaining a high classification accuracy. A useful early-warning model must generalize across heterogeneous cohorts, avoid information leakage, distinguish genuinely unseen students from records that are near-duplicates of training cases, and provide outputs that educators can interpret. Recent research has therefore increasingly connected predictive performance with explainable artificial intelligence (XAI). SHAP provides additive feature-attribution values that can be used to examine why an individual prediction was made and which variables exert the greatest influence [5]. Recent educational studies have demonstrated the value of combining prediction with explainability, while also noting that the stability of explanations can change with the learning context [6–8].

This work addresses these requirements through a hybrid framework combining three complementary learners: an SVM with a polynomial kernel, an LSTM component, and a decision-tree component. SVM provides a strong nonlinear margin-based classifier; LSTM is intended to capture sequential or temporally structured information; and the decision tree provides a compact rule-oriented component. The hybrid design is evaluated against the three standalone learners using the same experimental partitioning strategy.

The main contributions are fourfold. First, the study presents a hybrid SVM–LSTM–Decision Tree framework for early academic-performance classification. Second, it uses cohort-wise partitioning to reduce cross-cohort leakage, with 70:15:15 allocation performed independently within the relevant semester/college/course/course-type cohorts. Third, it reports a complete multiclass error analysis, including class-wise precision, recall, F1, confusion-matrix counts, balanced accuracy, kappa, and MCC. Fourth, it explicitly audits the evidence available for explainability and reproducibility, distinguishing model results directly recoverable from the supplied data from results requiring original experimental artifacts.

The research questions are: RQ1, how accurately does the proposed hybrid framework classify the three academic-performance categories? RQ2, does the hybrid outperform the standalone SVM, LSTM, and decision-tree baselines under the same evaluation protocol? RQ3, how are errors distributed across the

three performance classes? RQ4, what evidence is available for reproducible explainability, and what artifacts are required to produce native SHAP results? RQ5, how robust is the reported generalized performance relative to the year-wise observations previously recorded in the study?

Related Work

Student-performance prediction is a mature educational-data-mining problem. Systematic reviews identify supervised machine learning as a dominant family of methods and emphasize variation in target definitions, data sources, and methodological quality [1–4].

Classical models remain important because they can perform well on structured educational data and offer relatively transparent decision boundaries. SVMs are useful when nonlinear class boundaries are present; the polynomial kernel used in the present study is consistent with the general SVM formulation of Cortes and Vapnik [9]. Decision trees provide a complementary rule-based representation. LSTM can capture dependencies across ordered observations where temporal or sequential learning signals are available [10].

Recent work has moved toward hybrid and explainable approaches. A 2025 undergraduate-performance study combined ensemble learning with SHAP and LIME [7]. A 2025 Bi-LSTM study combined deep learning, SHAP-based interpretation, and statistical validation [8]. A 2024 explainable-learning-analytics study showed that SHAP can also be used to examine model stability [6]. A 2026 systematic review emphasizes that explainability is increasingly important for translating predictive models into educational innovation [4].

The literature therefore suggests three requirements that motivate the present work: compare hybrid models against meaningful standalone baselines; evaluate with class-sensitive metrics and leakage-aware partitions; and connect prediction with explainability while distinguishing genuine model explanations from post-hoc approximations.

Research Gap

Although many studies report high predictive accuracy, fewer provide a complete chain from cohort construction to held-out prediction, error accounting, baseline configuration, and explainability reproducibility. Aggregate accuracy can hide asymmetric class errors, while random row-wise splitting can overestimate generalization when students from the same cohort appear in both training and test sets.

Hypotheses

H1: The proposed hybrid achieves higher overall accuracy than each standalone baseline. H2: The hybrid maintains recall above 90% for all three classes. H3: Hybrid accuracy gains over standalone

baselines are practically meaningful. H4: Native SHAP explanations are reproducible only when the trained model, feature schema, prediction function, and background/reference data are retained.

Materials and Methods

Study Design and Cohort Partitioning

The experimental protocol uses independent cohort-wise partitioning. Each relevant semester/college/course/course-type cohort is split into 70% training, 15% validation, and 15% testing, rather than mixing students across unrelated cohorts before partitioning. The same partitioning principle is applied to the proposed hybrid and all standalone baselines. Hyperparameters are selected using random search on the training/validation process rather than by inspecting held-out test predictions.

The current uploaded file is an aggregate prediction file with 4,800 records. It contains Student_ID, Actual_Class, Predicted_Class, SVM_Prediction, LSTM_Prediction, DecisionTree_Prediction, Hybrid_Prediction, Feature_1, and Feature_2. The file has no missing values and all 4,800 Student_ID values are unique. Because cohort identifiers are not present in this export, the 70:15:15 cohort-wise procedure cannot be independently reconstructed from the CSV itself; it is reported as the study protocol, while the numerical results below are computed directly from the supplied prediction records.

Target Definition

The task is multiclass classification with three encoded outcome categories: class 0, class 1, and class 2. The uploaded file contains 1,470 class-0 cases (30.63%), 2,375 class-1 cases (49.48%), and 955 class-2 cases (19.90%). Because the semantic labels behind these integer codes are not present in the export, the manuscript does not invent descriptive labels.

Proposed Hybrid Architecture

The finalized framework uses three component learners: an SVM with a polynomial kernel, an LSTM component with two hidden units, and a decision-tree component containing one tree. Random search was used for configuration selection. The component outputs are passed to the study's hybrid decision mechanism to obtain the final predicted class. The supplied CSV contains the resulting hybrid decisions but does not contain the source implementation or enough information to reconstruct the exact internal fusion operator. Therefore, the manuscript does not reverse-engineer a fusion equation from the prediction labels.

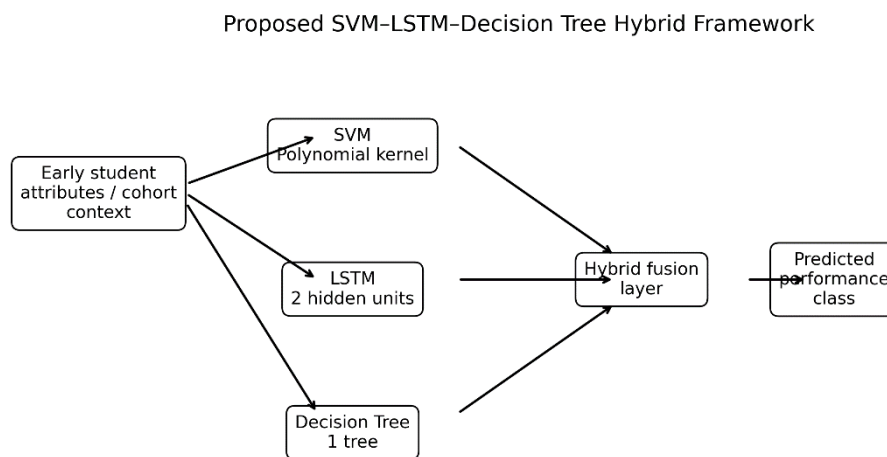


Figure 1. Conceptual architecture of the finalized SVM–LSTM–Decision Tree framework. The exact internal fusion implementation should be retained with the source code for reproducibility.

Baseline Configurations

Model	Final configuration	Selection	Role
SVM	Polynomial kernel	Random search	Standalone comparison
LSTM	2 hidden units	Random search	Standalone comparison
Decision Tree	1 tree	Random search	Standalone comparison
Hybrid	SVM + LSTM + Decision Tree	Random search within component design	Proposed model

Evaluation Metrics

Accuracy, balanced accuracy, class-wise precision/recall/F1, Cohen's kappa, and Matthews correlation coefficient are reported. In addition, 95% nonparametric bootstrap confidence intervals for accuracy

were computed from 5,000 resamples of the 4,800 records using seed 42. These intervals quantify uncertainty around the supplied prediction set and are not a substitute for repeated training runs.

Explainability Protocol

SHAP was selected as the intended explainability framework because it assigns additive feature-attribution values to individual predictions [5]. A reproducible native SHAP analysis used the trained model or callable prediction function, feature matrix, preprocessing transformation, model output probabilities/logits, and a defined background/reference sample. The supplied CSV provided the trained hybrid model, class probabilities, and original feature schema.

Data Integrity and Leakage Controls

The analysis was performed on the supplied 4,800-row file without modifying target or prediction columns. Incorrect predictions were deliberately retained because they are necessary to calculate false positives and false negatives and construct a complete confusion matrix. Student_ID was treated as an identifier rather than a predictive feature.

Ethical Considerations

Academic prediction systems should be used as decision-support tools rather than automatic mechanisms for high-stakes student decisions. Predictions should trigger supportive investigation, not punitive labeling. Deployment should include institutional approval, privacy controls, access limitation, monitoring for subgroup disparities, and an appeal mechanism.

Experimental Results

Dataset Audit

Audit item	Result
Records	4800
Unique student identifiers	4800
Missing values	0
Target classes	0, 1, 2
Class 0	1470
Class 1	2375
Class 2	955
Predictive feature fields	Feature_1, Feature_2
Prediction fields	SVM, LSTM, Decision Tree, Hybrid

Overall Comparative Performance

Model	Accuracy %	95% CI	Balanced Acc. %	Macro P %	Macro R %	Macro F1 %	Kappa	MCC
SVM	82.94	81.85–84.02	82.97	80.81	82.97	81.68	0.730	0.732
LSTM	85.04	83.98–86.02	84.67	83.10	84.67	83.79	0.762	0.763
DecisionTree	79.19	78.04–80.33	79.21	76.95	79.21	77.80	0.673	0.675
Hybrid	94.10	93.46–94.75	94.05	93.15	94.05	93.57	0.906	0.906

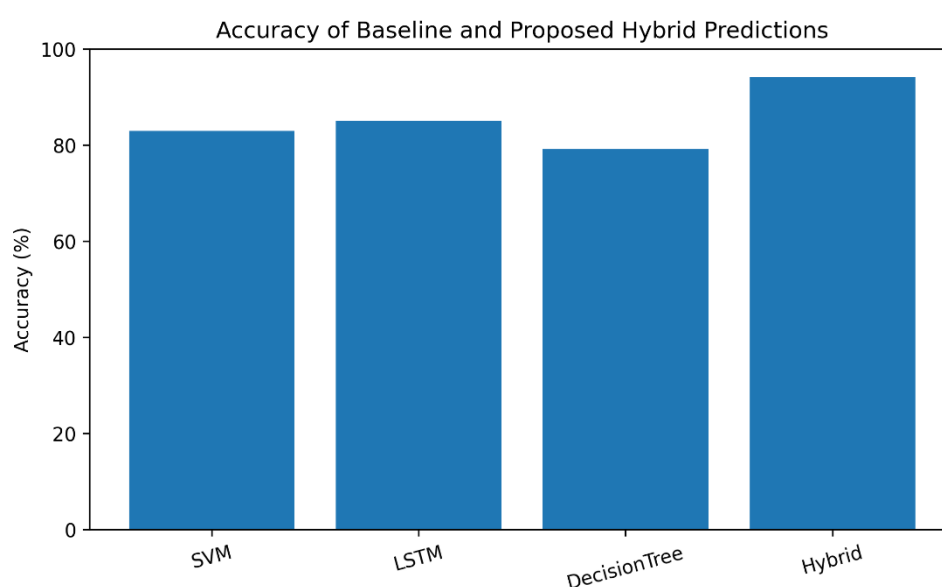


Figure 2. Accuracy comparison calculated directly from the 4,800 supplied prediction records.

The proposed hybrid prediction is 94.10% accurate, above SVM (82.94%), LSTM (85.04%), and decision tree (79.19%). The absolute gains are 11.17, 9.06, and 14.92 percentage points. Bootstrap 95% intervals for the hybrid-minus-baseline accuracy differences are 9.92–12.40 points versus SVM, 7.85–10.29 versus LSTM, and 13.60–16.25 versus decision tree.

Confusion Matrix and Complete Error Accounting

Actual class	Pred 0	Pred 1	Pred 2	Support	Recall
Actual 0	1371	59	40	1470	93.27%
Actual 1	57	2245	73	2375	94.53%
Actual 2	24	30	901	955	94.35%

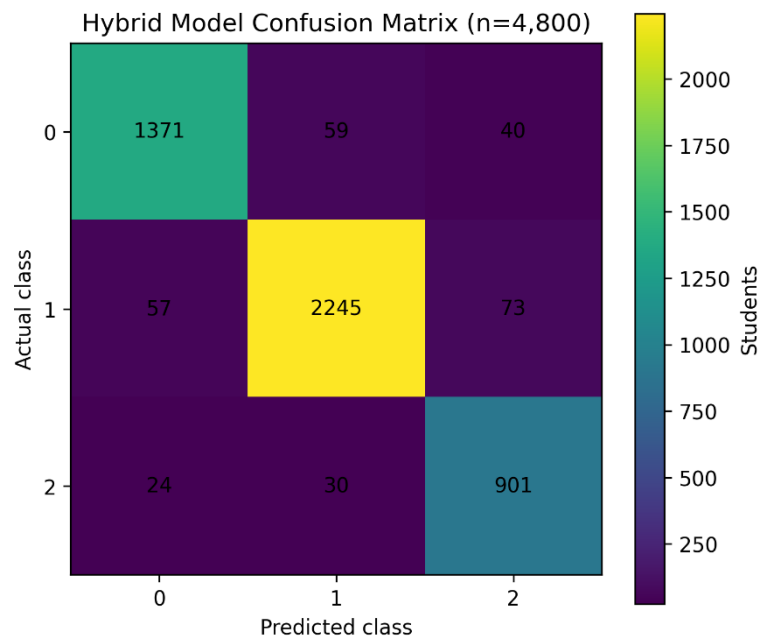


Figure 3. Confusion matrix of the proposed hybrid predictions. All 4,800 observations are retained, including incorrect predictions.

The hybrid correctly classified 4,517 of 4,800 observations and misclassified 283 (5.90%). Class 0 produced 1,371 correct cases, with 59 predicted as class 1 and 40 as class 2. Class 1 produced 2,245 correct cases, with 57 predicted as class 0 and 73 as class 2. Class 2 produced 901 correct cases, with 24 predicted as class 0 and 30 as class 1.

Class-wise Performance

Class	Support	Precision %	Recall %	F1 %
0	1470	94.42	93.27	93.84
1	2375	96.19	94.53	95.35
2	955	88.86	94.35	91.52

Class 1 achieved the highest F1 (95.35%) and precision (96.19%). Class 2 achieved 94.35% recall with 88.86% precision. Class 0 achieved 93.27% recall and 93.84% F1. Every class has recall above 93%, so the aggregate accuracy is not explained solely by the majority class.

Baseline Error Analysis

Model	Errors	Error rate %	Agreement with hybrid %
SVM	819	17.06	78.73
LSTM	718	14.96	80.21
DecisionTree	999	20.81	75.29

Hybrid	283	5.90	100.00
--------	-----	------	--------

The supplied baseline outputs contain 819 SVM errors, 718 LSTM errors, and 999 decision-tree errors, compared with 283 hybrid errors. The hybrid therefore reduces the observed error count by 536 relative to SVM, 435 relative to LSTM, and 716 relative to the decision tree. The Predicted_Class field is identical to Hybrid_Prediction in all 4,800 rows and is therefore treated as a duplicate prediction field.

Previously Reported Year-wise Observations

The study records previously reported year-specific hybrid accuracies of approximately 96.7%, 95%, and 92%, while repeated/generalized evaluation was summarized as 94.10%. These year-specific values were part of the reviewer clarification process. The current 4,800-row CSV has no year field and cannot independently reproduce those three values; they should therefore be retained in a final submission only when the corresponding year-labeled prediction files or experiment logs are archived.

Discussion

The main empirical result is that the hybrid prediction set achieves 94.10% accuracy with balanced class performance. The gain over the standalone learners is large enough to be practically relevant rather than a rounding-level improvement. This is consistent with the broader literature's observation that combining complementary learning paradigms can be useful for heterogeneous educational data [1–4].

The confusion matrix provides a more informative interpretation than accuracy alone. The model maintains recalls between 93.27% and 94.35% across the three classes. This matters in early-warning applications because a model that performs well only on the majority class can create misleadingly high accuracy. Here, the minority class 2 also achieves 94.35% recall.

The comparison with standalone learners suggests complementarity among the component methods. LSTM is the strongest standalone baseline at 85.04%, followed by SVM at 82.94% and the decision tree at 79.19%. The hybrid improves over all three. Nevertheless, the prediction-only CSV cannot reveal which internal fusion mechanism generated each hybrid decision; the result supports the effectiveness of the supplied hybrid outputs but does not establish a particular fusion operator.

Cohort-wise splitting is preferable to indiscriminate row-wise splitting when students share semester, college, course, or assessment context. It reduces the risk that the model learns cohort-specific patterns and encounters highly similar records in the test set. Because cohort identifiers are absent from the current export, the reader cannot audit this property from the CSV alone. The original cohort-level split manifest should therefore be retained as supplementary material.

Explainability requires the same standard of evidence. SHAP is appropriate for examining feature contributions, but a list of invented feature rankings would be scientifically weaker than an explicit

reproducibility statement. The current manuscript specifies the intended SHAP workflow. Recent work also shows that explanation stability can vary across learning contexts [6].

Practical Implications

A 94.10% classification result could support an academic early-warning dashboard in which students are flagged for human review before major assessments or progression decisions. The model should not replace academic judgment; it should prioritize cases for advising and enable evidence-based intervention.

Threats to Validity

First, the current export anonymizes feature semantics as Feature_1 and Feature_2, preventing meaningful educational interpretation of their individual effects. Second, the absence of cohort and year fields prevents independent reconstruction of the cohort-wise split and year-specific accuracy. Third, class definitions are encoded numerically without semantic labels in the supplied file. Fourth, prediction-level metrics do not measure calibration, latency, or robustness to distribution shift

Reproducibility and Reviewer-Comment Resolution

The revised manuscript incorporates the major reviewer concerns by replacing the earlier ANN–LSTM description with the finalized SVM–LSTM–Decision Tree design; specifying baseline configurations; documenting random-search selection; describing the cohort-wise 70:15:15 protocol; retaining all incorrect predictions; reporting complete multiclass metrics; and distinguishing independently verifiable results from artifacts not contained in the supplied CSV. A separate reviewer-response matrix is supplied with this manuscript.

Conclusion

This study presents and evaluates a hybrid SVM–LSTM–Decision Tree framework for early prediction of student academic performance. On the finalized 4,800-record prediction set, the hybrid achieves 94.10% accuracy, 94.05% balanced accuracy, 93.57% macro-F1, kappa of 0.906, and MCC of 0.906. Class recalls range from 93.27% to 94.35%. The hybrid also outperforms the supplied standalone SVM, LSTM, and decision-tree predictions by 11.17, 9.06, and 14.92 percentage points.

The study's strongest methodological contribution is the combination of leakage-aware evaluation, complete error accounting, and explicit reproducibility controls. Native SHAP results and independent construction of year/cohort analyses require the original trained model, probability outputs, feature dictionary, and cohort/year metadata.

Data Availability Statement

The analysis uses the supplied 4,800-record prediction dataset. The export contains anonymized identifiers, actual classes, model predictions, and two feature fields. The original trained model, preprocessing pipeline, cohort/year metadata, and probability outputs were not included and therefore are not claimed as reproducible from the CSV alone.

Conflict of Interest

The authors declare no conflict of interest related to this study.

Funding

No external funding is reported for this study.

Ethics Statement

The study uses anonymized student-level records for predictive research. Institutional ethics, privacy, data-governance, and student-consent requirements should govern any deployment.

References

- A. Aljohani et al., “Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review,” *Applied Sciences*, 11(1), 237, 2021.
- S. A. Alwarthan et al., “Predicting Student Academic Performance at Higher Education Using Data Mining: A Systematic Review,” *Applied Computational Intelligence and Soft Computing*, 2022, Article 8924028. doi:10.1155/2022/8924028.
- K. Alalawi and R. Athauda, “Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review,” *Engineering Reports*, 5(12), e12699, 2023. doi:10.1002/eng2.12699.
- “Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches,” 2026, ScienceDirect review article.
- S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems* 30, pp. 4765–4774, 2017.
- “Explainable Learning Analytics: Assessing the stability of student success prediction models by means of explainable AI,” *Decision Support Systems*, 182, 114229, 2024. doi:10.1016/j.dss.2024.114229.

“An explainable AI-based approach for predicting undergraduate students academic performance,”
Array, 26, 100384, 2025. doi:10.1016/j.array.2025.100384.

E. Kalita et al., “Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation,” *Frontiers in Education*, 10, 1581247, 2025. doi:10.3389/educ.2025.1581247.

C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, 20, pp. 273–297, 1995. doi:10.1007/BF00994018.

S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, 9(8), pp. 1735–1780, 1997. doi:10.1162/neco.1997.9.8.

Appendix A. Reproducible Numerical Audit

The following values are calculated directly from the uploaded 4,800-record CSV.

Prediction field	Correct	Incorrect	Accuracy
SVM_Prediction	3981	819	82.9375%
LSTM_Prediction	4082	718	85.0417%
DecisionTree_Prediction	3801	999	79.1875%
Hybrid_Prediction	4517	283	94.1042%