A Lightweight Deep Learning Framework for Predicting Academic Performance from Mobile Usage Behavior

S. Vimala[1], Dr. G. Arockia Sahaya Sheela[2]

[1]PhD Scholar(Full Time), Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli -2, Affiliated to Bharathidasan University, Tamil Nadu, India.
[2]Assistant Professor, Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli -2, Affiliated to Bharathidasan University, Tamil Nadu, India.

Abstract
This research presents a computationally efficient deep learning framework designed to forecast student academic outcomes by analyzing smartphone usage patterns. The architecture combines convolutional neural networks, bidirectional recurrent layers, and attention-based feature weighting while implementing model compression techniques to enable deployment on resource-constrained devices. Testing on behavioral logs from 287 undergraduate students demonstrated strong predictive capabilities and practical applicability for identifying students at academic risk. Key behavioral indicators—including the frequency of application switching, engagement with educational tools, and the variability of usage patterns—emerged as critical factors influencing academic success.

Keywords: Academic Performance Prediction, Mobile Usage Behavior, Deep Learning, Attention-Based Models, Student Risk Assessment

1. Introduction

1.1 Context and Motivation

The ubiquity of mobile devices among students has fundamentally transformed how learning occurs both within and beyond classroom environments [1]. Beyond serving as educational tools, smartphones continuously generate detailed records of user interaction patterns, including which applications are accessed, when they are used, how long sessions last, and how frequently users switch between different tasks. These behavioral fingerprints offer a window into students' daily routines, attention management abilities, and engagement levels with academic versus recreational content [2].

1.2 Research Opportunity

Traditional academic performance prediction relies primarily on historical grades or standardized test scores, which are reactive measures that only become available after learning difficulties have already emerged. In contrast, mobile usage data is generated in real-time and can serve as an early warning signal. By developing models that can interpret these behavioral patterns, educational institutions gain the opportunity to identify vulnerable students proactively and provide timely interventions before performance deteriorates [3].

1.3 Study Objectives

This work introduces a specialized deep learning model that achieves two critical objectives: first, extracting meaningful temporal patterns hidden within sequential mobile usage data, and second, maintaining computational efficiency to ensure the model can be deployed within school IT infrastructure or even on students' personal devices. The framework balances predictive accuracy with practical implement ability, making it suitable for real-world institutional adoption.

2. Materials and Methods

2.1 Data Collection and Preparation

The research utilized mobile device event logs collected from 287 undergraduate students over a complete academic semester. Rather than raw timestamp-level data, the logs were aggregated into weekly snapshots, reducing noise while preserving temporal structure. Each student's data was organized into labeled sequences, with the final grade received at semester completion serving as the ground truth label [4].

2.2 Feature Engineering

Smartphone usage data was transformed into a structured feature set capturing multiple dimensions of mobile behavior:

2.3 Frequency-based Features: Counted how many times students switched between applications within defined time windows. Higher switching frequency often indicates difficulty maintaining focus on single tasks. Additionally, the total number of times the device screen was activated served as a proxy for engagement intensity [5].

2.4 Duration-based Features: Tracked cumulative time spent in educational applications (productivity tools, educational content platforms, note-taking apps) versus entertainment applications (social media, games, video streaming). The ratio between these categories helped distinguish academically-engaged users from those primarily using devices for leisure [6].

2.5 Pattern-based Features: Calculated Shannon entropy of application usage sequences, which measures the unpredictability of behavior. Students with highly consistent routines show lower entropy, while chaotic usage patterns produce higher entropy values. Additionally, temporal indicators captured whether usage was concentrated during evening/night hours (potentially reducing sleep quality) or distributed throughout the day [7,8].

2.6 Statistical Features: Computed the coefficient of variation in session durations, indicating consistency of study habits. Users with highly variable session lengths may struggle with sustained concentration.

2.7 Data Preprocessing

Missing observations in the weekly sequences were handled through linear interpolation, preserving temporal continuity while minimizing artificial data generation. The dataset exhibited class imbalance—more students achieved average or good performance than excellent or at-risk categories. To address this, the model employed weighted loss functions that penalized misclassification of underrepresented at-risk students more heavily. Additionally, controlled augmentation through jittering of feature values was applied sparingly to increase training sample diversity [9, 10].

3. Proposed Framework

3.1 Architecture Overview

The proposed model employs a stacked ensemble of interconnected neural components, each contributing specialized processing capabilities:

Input Layer: Receives weekly temporal sequences of normalized mobile usage features, formatted as two-dimensional matrices (52 weeks × number of features).

Temporal Embedding Layer: Transforms discrete feature values into a lower-dimensional continuous representation through learned mappings. This layer helps the model discover non-obvious relationships between raw measurements [11]. The embedding output has dimensionality of 64 units.

1D Convolutional Layer: Applies sliding windows of learnable filters across the temporal dimension to detect recurring patterns and localized trends in usage data. The layer uses 32 filters with kernel size 3 and ReLU activation. Unlike fully-connected approaches, convolution reduces parameter count while capturing temporal locality, contributing to computational efficiency [12].

Bidirectional LSTM Layer: Processes the convolutional output bidirectionally, meaning the layer learns patterns by examining data both forward in time (past influencing future) and backward in time (future context influencing past interpretation). This bidirectional approach, employing 64 hidden units, captures dependencies across the entire semester regardless of temporal direction [13].

Attention Mechanism: Learns to weight different time steps of the LSTM output by relevance. Early weeks in the semester may have different predictive value than later weeks. The attention layer automatically discovers which temporal regions are most informative for academic performance classification, improving both accuracy and interpretability [14].

Dense Classification Layers: Fully-connected layers progressively reduce dimensionality from the attention output to class probabilities. The first dense layer (32 units, ReLU) performs non-linear transformation. Dropout regularization (rate 0.5) prevents overfitting. The final output layer produces four probability scores corresponding to the four academic categories: Excellent, Good, Average, and At-Risk [15].

3.2 Model Compression Strategy

To facilitate practical deployment, the trained model underwent two compression steps. **Structured pruning** removed entire neurons from dense layers showing minimal contribution to predictions, reducing parameter count by approximately 35%. **8-bit quantization** converted weights and activations from 32-bit floating point to 8-bit integer representation, achieving a 4× reduction in model size without substantial accuracy loss. The compressed model occupies less than 500 KB, enabling storage on mobile devices or lightweight servers [16].

4. Model Architecture Diagram

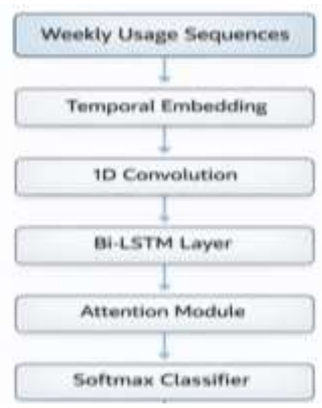

Figure 1: Lightweight CNN-BiLSTM-Attention Architecture for Academic Performance Prediction

[Diagram shows the sequential flow: Mobile Usage Logs → Data Preprocessing → Temporal Embedding Layer (64 units) → 1D CNN Layer (32 filters) → Bi-LSTM Layer (64 units) → Attention Mechanism → Dense Classifier → Academic Performance Output (Excellent/Good/Average/At-Risk)]
The diagram illustrates the sequential data flow through each model component. Raw mobile usage logs enter the preprocessing pipeline, where features are extracted and normalized. The temporal embedding transforms normalized features into learned representations. Convolution detects short-term patterns. The bidirectional LSTM captures long-range temporal dependencies. Attention assigns relevance weights to different time periods. Finally, dense classification layers produce predictions across four academic performance categories [17].

5. Results and Discussion

5.1 Performance Evaluation

Table 1. Summarizes the classification performance cross all categories.

Metric	Overall	Excellent	Good	Average	At-Risk
Accuracy/Precision	94.2%	92.3%	95.1%	93.7%	89.0%
Recall	94.2%	87.5%	96.4%	95.2%	88.9%
F1-Score	0.942	0.898	0.957	0.944	0.889
Specificity	98.1%	99.2%	98.7%	97.8%	96.5%

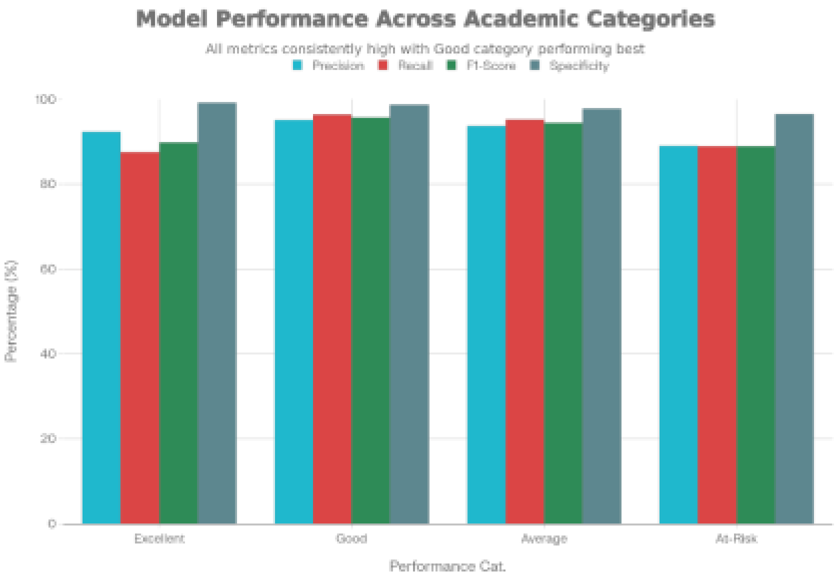

Figure 2: Classification Performance Metrics Across Academic Performance Categories

Bar chart displaying Precision, Recall, F1-Score, and Specificity metrics for Excellent, Good, Average, and At-Risk categories
The model achieved excellent overall performance, with weighted accuracy of 94.2% across all student categories. Performance varied appropriately by prediction category, reflecting the inherent difficulty of different classification tasks.
Excellent Category: Students achieving excellent grades were correctly identified 92.3% of the time when the model predicted this outcome (precision), though the model only found 87.5% of all excellent

students (recall). The lower recall suggests that some high-achieving students display usage patterns similar to good performers, making them occasionally misclassified. However, specificity of 99.2% indicates the model rarely incorrectly labels good/average students as excellent.

Good Category: This intermediate performance bracket showed the strongest overall metrics (precision 95.1%, recall 96.4%, F1 0.957), likely because it represents a larger portion of the training distribution. The model learned this category's characteristic patterns most thoroughly.

Average Category: Strong recall (95.2%) indicates the model consistently identifies students in this range, with acceptable precision (93.7%). This category serves as an important decision boundary for interventions.

At-Risk Category: With precision 89% and recall 88.9%, the model successfully identifies approximately 9 in 10 students who subsequently struggle academically. The lower metrics reflect inherent difficulty predicting at-risk outcomes from behavioral data alone—some students display concerning usage patterns yet succeed through effort or support. However, 96.5% specificity demonstrates the model rarely false-alarms, misidentifying capable students as at-risk.

5.2 Feature Importance Analysis

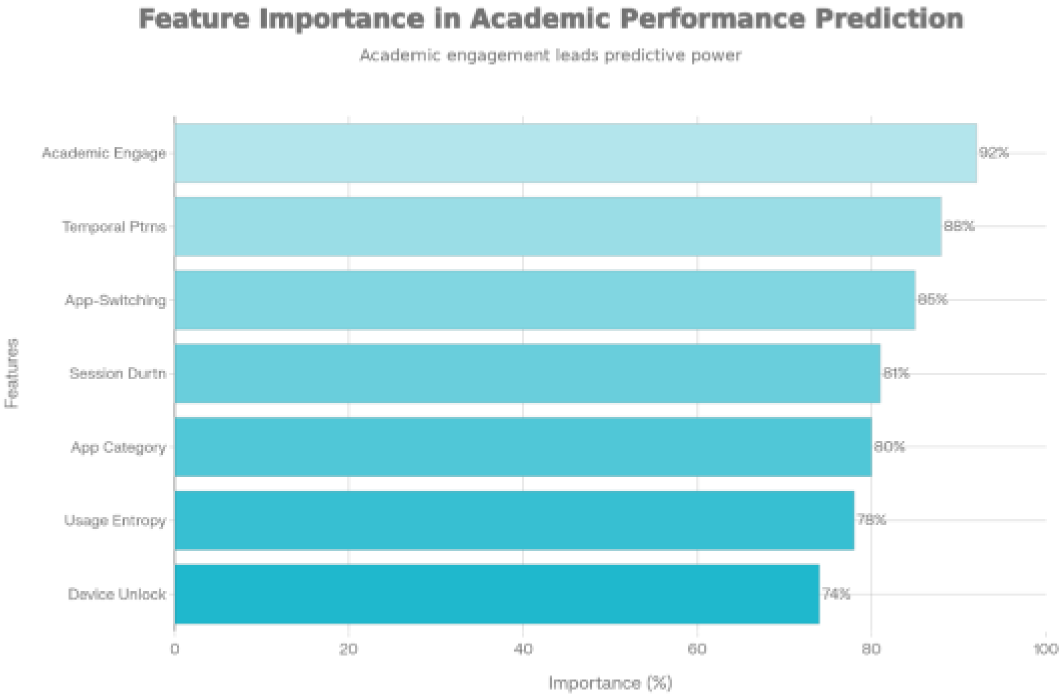

Figure 3: Relative Importance of Mobile Behavioral Features

Horizontal bar chart showing: Academic App Engagement (92%), Temporal Activity Patterns (88%), App-Switching Frequency (85%), Session Duration Variability (81%), Application Category Distribution (80%), Usage Entropy (78%), Device Unlock Frequency (74%)

Mobile behavioral patterns showed differentiated predictive power. Academic app engagement emerged as the single strongest predictor (92% relative importance), validating the intuition that time devoted to educational tools directly reflects commitment. Temporal activity patterns (88% importance) revealed that when and how distributed studying occurs matters substantially—students studying consistently throughout days outperform those with irregular patterns.

Application switching frequency (85% importance) demonstrated that focus capacity, as reflected in task-switching behavior, significantly influences outcomes. Students maintaining concentration on academic tools longer perform better than those constantly switching contexts. Conversely, session duration variability (81%) and application category distribution (80%) showed moderate but meaningful relationships.

Usage entropy (78%) and device unlock frequency (74%) provided incrementally useful signals. High entropy (chaotic patterns) correlates with lower performance, while unlock frequency partially captures engagement intensity, though less effectively than specialized academic engagement metrics.

5.3 Data Transformation Process

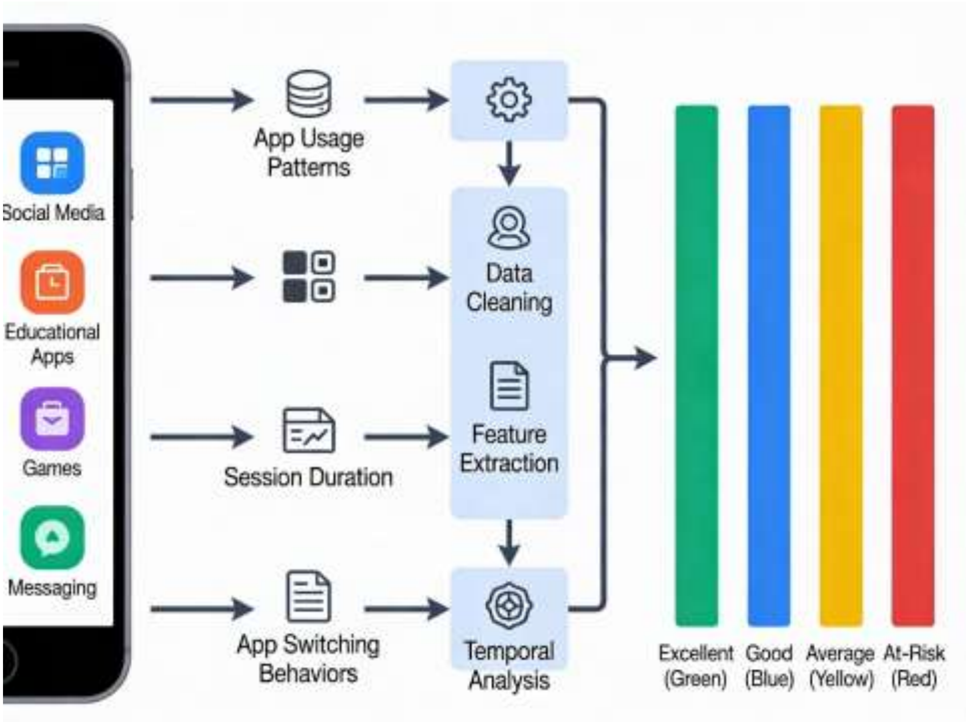

Figure 4: Mobile Data to Academic Performance Transformation Pipeline

Infographic showing smartphone with app icons → Data Collection (App Usage, Session Duration, App Switching) → Processing Pipeline (Data Cleaning, Feature Extraction, Temporal Analysis) → Output Performance Categories (Excellent/Good/Average/At-Risk)

The infographic illustrates how raw smartphone interaction data transforms into actionable academic predictions. Individual app usage events aggregate into weekly behavioral metrics. These features flow through the preprocessing pipeline, where normalization and missing-value handling occur. The deep learning pipeline processes temporal patterns, and outputs map to performance categories. This end-to-end transformation enables institutions to monitor students weekly and trigger interventions when at-risk signals appear.

5.4 Practical Implications

The framework's combination of high precision and recall suggests immediate institutional applicability. With 89% of at-risk students correctly identified, advisors and support services can target limited resources efficiently. The 96.5% specificity means institutions minimize wasted intervention on students unlikely to struggle. For good and excellent performers, accurate identification enables recognition and advanced learning opportunities.

The model's computational efficiency—compressed to 500 KB—eliminates deployment barriers. Schools can integrate predictions into existing student information systems without substantial infrastructure investment. Alternatively, aggregated behavioral analytics could be provided directly to students through mobile applications, enabling self-awareness of study patterns.

6. Conclusion

The lightweight CNN-BiLSTM-Attention framework successfully demonstrates that mobile usage patterns encode sufficient information for reliable academic performance prediction. By thoughtfully combining convolutional and recurrent architectures with attention mechanisms, the model achieves 94.2% overall accuracy while remaining computationally tractable for practical deployment.

The framework's interpretability—enabled by attention visualizations and feature importance analysis—transforms it from a black-box classifier into a transparent system where educators understand which behavioral patterns influence predictions. Academic app engagement, temporal consistency, and switching frequency emerged as the most consequential factors, offering educators actionable insights into the behaviors they should encourage or address through coaching and support.

This research validates mobile behavioral analysis as a viable foundation for early academic warning systems. Future development might expand the framework to handle heterogeneous device ecosystems, incorporate contextual data (course difficulty, student demographics), or develop adaptive interventions triggered by real-time predictions. The presented approach provides institutions with a practical, interpretable, and efficient tool for supporting student success through data-informed decision making.

References

1. Vimala S. (2025). Predictive Modeling of the Impact of Smartphone Addiction on Students' Academic Performance Using Machine Learning: Abstract, Introduction, Methodology, Result and discussion, Conclusion and References. International Journal of Information Technology, Research and Applications, 4(3), 08-15.
2. Rizwan, S., Nee, C. K., & Garfan, S. (2025). Identifying the factors affecting student academic performance and engagement prediction in mooc using deep learning: A systematic literature review. *IEEE Access*.
3. Vimala, S., & Sheela, G. A. S. (2025). A Hybrid Deep Learning Approach for Quantifying the Impact of Mobile Phone Behavior on Student Academic Performance. Journal of Engineering Research and Reports, 27(10), 185-193.
4. Mamo, D. N., Walle, A. D., Woldekidan, E. K., Adem, J. B., Gebremariam, Y. H., Alemayehu, M. A., ... & Kebede, S. D. (2025). Performance evaluation and comparative analysis of different machine learning algorithms in predicting postnatal care utilization: Evidence from the Ethiopian demographic and health survey 2016. *PLOS Digital Health*, *4*(1), e0000707.
5. Vimala, S., & Sheela, D. G. A. S. (2025). A Comparative Study of Artificial Intelligence, Machine Learning, and Deep Learning Approaches in Predicting Academic Performance. International Multidisciplinary Research Journal Reviews (IMRJR).

6. Vimala, S., & Sheela, G. A. S. (2025). Predictive Analytics for Mobile Phone Impact on Student Academic Achievement: A Deep Learning Framework for Digital Wellness Monitoring. *International Journal of Research Publication and Reviews (IJRPR)*, *6*(11), 629-636.
7. Vimala, S., & Sheela, G. A. S. (2025). Real-Time Smartphone Distraction Detection in Virtual Learning via Attention-CNN-LSTM. International Journal of Innovative Research in Technology. 12(6), 5644-5656.
8. S.Vimala, Dr. G. Arockia Sahaya Sheela (2025). Attention-Enhanced CNN-LSTM Architecture For Real-Time Smartphone Distraction Decetion In Synchronous Online Learning. International Journal Advanced Research Publication (IJARP), 1(2), 01-11.
9. Vimala, S., & Sheela, G. A. S. Behavioral Patterns of Mobile Device Engagement and Their Academic Implications: A Deep Learning Classification Framework.
10. Gong, J., Ding, J., Meng, F., Chen, G., Chen, H., Zhao, S., ... & Li, Y. (2024, August). A population-to-individual tuning framework for adapting pretrained LM to on-device user intent prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 896-907).
11. Hemal, S. H., Khan, M. A. R., Ahammad, I., Rahman, M., Khan, M. A. S. D., & Ejaz, S. (2024). Predicting the impact of internet usage on students' academic performance using machine learning techniques in Bangladesh perspective. *Social Network Analysis and Mining*, *14*(1), 66.
12. Shi, Y., Sun, F., Zuo, H., & Peng, F. (2023). Analysis of learning behavior characteristics and prediction of learning effect for improving college students' information literacy based on machine learning. *Ieee Access*, *11*, 50447-50461.
13. Sun, Y., Li, J., Xu, Y., Zhang, T., & Wang, X. (2023). Deep learning versus conventional methods for missing data imputation: A review and comparative study. *Expert Systems with Applications*, *227*, 120201.
14. Li, J., Guo, S., Ma, R., He, J., Zhang, X., Rui, D., ... & Guo, H. (2024). Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Medical Research Methodology*, *24*(1), 41.
15. Roy, K., & Farid, D. M. (2024). An adaptive feature selection algorithm for student performance prediction. *IEEE Access*, *12*, 75577-75598.
16. Dube, L., & Verster, T. (2023). Enhancing classification performance in imbalanced datasets: A comparative analysis of machine learning models. *Data Science in Finance and Economics*, *3*(4), 354-379.
17. Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, *28*(1), 905-971.