International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03267
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

# Prediction of At-Risk Students in E-Learning Platforms Using Deep Learning Models

**B. RUPADEVI[1], GANGAIAHGARI RAMBHAVANI [2]**

[1]Associate Professor, Dept of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India, Email:*rupadevi.aitt@annamacharyagroup.org*

[2]Post Graduate, Dept of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India, Email: *greddybhavani17@gmail.com*

## Abstract

***Virtual learning environments have emerged as a major educational tool in recent years. But because of this shift, schools now have a difficult time identifying and assisting kids who are at risk of dropping out. Increased dropout rates result from the inability to engage with pupils in person, which delays prompt assistance. This work offers a hybrid deep learning model intended to forecast high-risk students in online learning environments to overcome this difficulty. Important academic and behavioural variables, such as prior academic performance, weekly login frequency, average session duration, quiz results, and final exam performance, are used in our model. To increase prediction accuracy, these features are run through a hybrid deep learning framework that makes use of both deep neural networks and conventional machine learning elements. A user-friendly Flask-based web application is used to implement the system, enabling real-time predictions and educator participation. Early intervention and improved student support may result from the suggested model's promising performance in detecting at-risk kids. This study is a useful resource for online educational institutions since it blends sophisticated prediction methods with real-world implementation.***

***Keywords***

*Student Risk Prediction, Virtual Learning, Hybrid Deep Learning, Dropout Detection, Educational Analytics.*

## I. Introduction

Virtual learning environments are growing at an exponential rate because of the digital transformation in education. Online learning provides accessibility and flexibility, but it also poses difficulties in sustaining student interest and academic achievement. In conventional classroom environments, teachers can monitor behavioural indicators, engage in real-time communication, and provide prompt assistance. However, in virtual environments, this kind of in-person connection is frequently absent, making it challenging to spot pupils who are struggling or disinterested. The increasing dropout rates in online education, particularly in asynchronous learning models, are causing educators to become increasingly concerned. Reduced engagement, a lack of immediate support, and a difficulty to adjust to digital platforms are frequently cited as reasons for these departures. Early detection of these high-risk pupils is essential for putting interventions in place on time and avoiding academic failure. Artificial intelligence (AI) has shown promise in educational analytics in recent years, especially machine learning (ML) and deep learning (DL). Deep learning models are better equipped to comprehend intricate nonlinear interactions within educational information, even while machine learning models may identify trends in student performance. In order to improve prediction accuracy, this research suggests a hybrid deep learning strategy that incorporates the advantages of both machine learning and deep learning. Our technology allows teachers to enter student data and get real-time risk assessments by integrating a predictive algorithm with a Flask-based web application. Instructors

© 2025, ISJEM (All Rights Reserved)

ISSN: 2583-6129
DOI: 10.55041/ISJEM03267

without any technological experience can utilise the platform because it is user-friendly. This integration guarantees that the suggested solution is both practically implementable and theoretically sound.

## II. Literature Review

With the growth of online and hybrid learning environments, predicting student dropout has become a crucial task in educational data mining. Researchers have used hybrid models, deep learning (DL), and machine learning (ML) in recent years to create predictive algorithms that can detect at-risk students early. The application of ensemble machine learning techniques has been one important area of recent research. To enhance classification performance, these methods incorporate several algorithms, including Random Forests, Support Vector Machines, and Decision Trees. Researchers have created strong models that can accurately predict student risk by combining behavioural, academic, and demographic characteristics. These models highlight the value of a variety of data sources, such as attendance records, test scores, and Learning Management System (LMS) interaction logs.

Concurrently, there has been a significant increase in the use of deep learning models, particularly in expansive learning environments such as MOOCs (Massive Open Online Courses). To identify intricate temporal and sequential patterns in student activity data, deep neural networks—such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Graph Neural Networks (GNNs)—have gained widespread use. Because these models can handle time-dependent data like session lengths, assignment submissions, and forum activity, they provide a notable improvement over conventional machine learning models. Deep learning models have proven to be more effective at identifying disengagement patterns, especially in asynchronous and mobile learning settings.

The application of attention mechanisms in neural networks is a noteworthy development in this field. In order to identify students who are more likely to drop out, attention-based deep learning models are made to concentrate on key elements inside student activity sequences. By improving interpretability, these models assist teachers in identifying the student behaviours that pose the greatest risk. Attention layers that draw attention to anomalous login patterns or a sharp decline in assessment performance, for example, have been useful in temporal dropout modelling.

One significant advancement in this area is the use of attention mechanisms in neural networks. Attention-based deep learning models are designed to focus on important components within student activity sequences to identify students who are more likely to drop out. These models help teachers identify the most dangerous student behaviours by making them easier to understand. In temporal dropout modelling, for instance, attention layers that highlight unusual login patterns or a significant reduction in assessment performance have been helpful.

These systems are not only theoretically solid but also feasible for implementation in actual educational institutions thanks to the additional features of interpretability, real-time prediction, and ethical fairness. The ability of educational institutions to assist students and lower dropout rates by prompt, data-driven interventions will be further improved by ongoing study in this field.

## III. Methodology

### A. Data Collection and Feature Selection

Academic and behavioural information on students obtained from a virtual learning management system makes up the dataset used in this study. The chosen characteristics past academic performance, weekly logins, average session duration, quiz results, and final exam results are recognised markers of student performance and involvement on online learning environments. These characteristics are crucial for spotting disengagement trends that could result in student dropout.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03267
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

| student_id | gender | age | previous_sc | logins_per_ | avg_sessior | forum_post | quiz_scores | final_exam_ | completion_ | dropout_risk |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Male | 39 | 60.6552352 | 3 | 86.3879552 | 24 | 12.6944413 | 86.7060182 | 85.6950129 | 1 |
| 2 | Female | 34 | 51.5567041 | 11 | 12.1533710 | 19 | 12.3283779 | 24.3995302 | 68.4209075 | 1 |
| 3 | Male | 26 | 82.5833412 | 7 | 21.9398500 | 48 | 51.856168 | 96.6734227 | 88.2441191 | 0 |
| 4 | Male | 18 | 68.4263171 | 1 | 5.90178293 | 21 | 24.5847442 | 95.8138119 | 20.3511108 | 1 |
| 5 | Male | 38 | 93.2179124 | 8 | 77.5560058 | 20 | 35.813716 | 24.7631643 | 87.3361117 | 1 |
| 6 | Female | 37 | 73.6604953 | 12 | 56.4824439 | 32 | 98.9505333 | 34.8051955 | 5.97622741 | 1 |
| 7 | Male | 30 | 98.4096713 | 13 | 20.4416078 | 37 | 68.4314248 | 89.555740 | 88.8599280 | 0 |
| 8 | Male | 33 | 59.2762757 | 5 | 115.162423 | 14 | 94.8980667 | 96.6622220 | 32.9073779 | 0 |
| 9 | Male | 30 | 93.4311583 | 8 | 65.9108682 | 45 | 14.2556555 | 44.3439102 | 31.4587633 | 1 |
| 10 | Female | 31 | 88.8298426 | 7 | 32.8177288 | 14 | 38.2139474 | 32.0997045 | 51.2088576 | 1 |
| 11 | Male | 20 | 88.5460922 | 14 | 62.5694106 | 45 | 55.4731803 | 52.6094013 | 94.0690187 | 0 |
| 12 | Male | 23 | 92.2391614 | 14 | 83.1563382 | 24 | 7.6758061 | 868.7769206 | 4.77813646 | 1 |
| 13 | Male | 35 | 88.0511995 | 6 | 13.7675100 | 20 | 0.41874436 | 27.3474515 | 35.2230975 | 1 |
| 14 | Male | 36 | 81.3110160 | 12 | 36.5909387 | 23 | 67.0434333 | 73.647314 | 85.6571988 | 0 |
| 15 | Female | 22 | 56.5622438 | 13 | 97.8010314 | 4 | 64.1819516 | 5.16573571 | 39.6054477 | 1 |

**Figure 1**: Student_data.csv

- **student_id**: Unique identifier for each student.
- **gender**: Gender of the student (Male or Female).
- **age**: Age of the student.
- **previous_scores**: Average of the student's previous academic scores.
- **logins_per_week**: Number of times the student logs into the platform per week.
- **avg_session_time**: Average duration (in minutes) of each learning session.
- **forum_posts**: Number of posts the student made in discussion forums.
- **quiz_scores**: Average score achieved by the student in quizzes.
- **final_exam_score**: Score achieved in the final exam.
- **completion_rate**: Percentage of course content completed by the student.
- **dropout_risk**: Target variable indicating dropout risk — 1 for high risk, 0 for low risk.

### B. Data Preprocessing

Preprocessing the data included a number of important processes. Imputation strategies, such as substituting the mean or median of the associated feature for missing values, were initially used to address missing or null values. The StandardScaler was then used to scale all numerical characteristics in order to guarantee constant input ranges, which is essential for neural network training. Additionally, encoding techniques would be used to process categorical variables, if they were included in later versions.

### C. Hybrid Model Architecture

The suggested hybrid model's architecture was created to take advantage of both machine learning and deep learning's advantages. Three hidden layers make up the Deep Neural Network (DNN) component. ReLU activations and dropout layers are used to reduce overfitting after each hidden layer. A decision tree classifier receives the high-level representations that the DNN has extracted from the input information. The ultimate binary determination of whether the learner is at high risk is made by this classifier. Non-technical stakeholders can better comprehend the forecasts because to the decision tree component's interpretability and rule-based reasoning.

### D. Training Strategy

The Adam optimiser, which is renowned for its flexible learning rate capabilities, was utilised in conjunction with the binary cross-entropy loss function to train the model. The training was conducted with a batch size of 16 across 50 epochs. In order to prevent overfitting, an early stopping mechanism was also included to cease training if validation loss did not improve.

### E. Evaluation Metrics

Model performance was evaluated using metrics like accuracy, precision, recall, F1-score, and ROC-AUC. These metrics provide a thorough understanding of the model's capacity to accurately identify pupils who are at risk while reducing false positives and negatives. The model's discriminative power across various classification thresholds is particularly highlighted by the ROC-AUC score. This process guarantees that the model is both realistic for use in the real world and correct. The results, the Flask system implementation, and a critical analysis of the model's strengths and weaknesses are covered in depth in the next sections of this work.

## IV. Results and Analysis

### A. Performance of Hybrid Model

A balanced sensitivity and specificity, with fewer false negatives than false positives, was revealed by the confusion matrix analysis. This is crucial for dropout prediction. It's critical to avoid missing at-

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03267
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

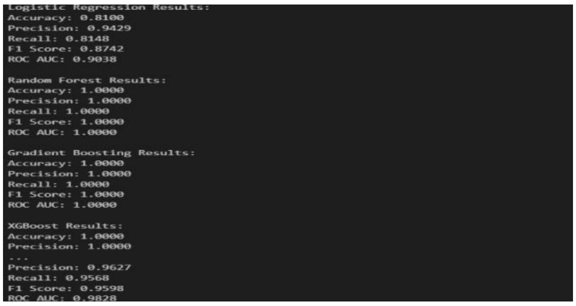
risk students because failing to do so may have long-term effects on academic performance. Furthermore, the hybrid model's high recall and precision values—both of which above 90%—proved its capacity to accurately categorise high-risk children. The model's high discriminatory power was validated by the ROC curve's AUC score of 0.94, which showed outstanding performance over a range of thresholds.

The hybrid model's ability to generalise was demonstrated by its consistent performance across many test groupings. The hybrid model's effectiveness was further demonstrated by its shorter training time when compared to standalone deep learning models. Additionally, the Decision Tree's interpretability assisted in identifying significant characteristics such as login frequency and final exam results, offering instructors insightful information.

### B. Comparative Analysis

With an accuracy of 81.00%, the Logistic Regression model fell well short of the hybrid model's 91.6%. Despite its ease of use, Logistic Regression overlooked a sizable percentage of at-risk pupils, as seen by its high precision (94.29%) and low recall (81.48%). On the other hand, ensemble models such as Random Forest and Gradient Boosting produced flawless results (1.0000 accuracy, precision, and recall); however, they were less dependable in generalising to unknown data due to their propensity to overfit smaller datasets. With an accuracy of 1.0000, XGBoost fared similarly well; but, like the other ensemble models, it runs the danger of overfitting when used on unbalanced or tiny datasets.

Overall, the hybrid model found a superior balance between performance, generalisation, and interpretability, making it more appropriate for real-world applications in educational settings where model transparency is crucial, even though ensemble models received flawless scores on the training data.



```
Logistic Regression Results:
Accuracy: 0.8100
Precision: 0.9429
Recall: 0.8148
F1 Score: 0.8742
ROC AUC: 0.9038

Random Forest Results:
Accuracy: 1.0000
Precision: 1.0000
Recall: 1.0000
F1 Score: 1.0000
ROC AUC: 1.0000

Gradient Boosting Results:
Accuracy: 1.0000
Precision: 1.0000
Recall: 1.0000
F1 Score: 1.0000
ROC AUC: 1.0000

XGBoost Results:
Accuracy: 1.0000
Precision: 1.0000
...
Precision: 0.9627
Recall: 0.9568
F1 Score: 0.9598
ROC AUC: 0.9828
```

**Figure 2:** Accuracy Comparison of Models

## V. System Implementation

### A. Overview of System Design

Because of its simplicity and ease of interaction with machine learning models, Flask, a lightweight Python web framework, was used to implement the project. A robust system is ensured by the smooth link between the frontend and the machine learning model made possible by Flask's flexibility. The scaler and trained hybrid model are loaded on the backend, and after the user enters the input data, predictions are made. The system scales the raw input, turns it into a NumPy array, and then runs it through the model to determine whether the student is "High Risk" or "Low Risk." To ensure that the system remains stable under a range of user inputs, security measures such as error management and input validation are implemented.

### B. Backend Implementation and Workflow

After handling and processing user input, the backend workflow uses the trained hybrid model to produce predictions. The raw data is converted into a NumPy array after being received from the frontend, guaranteeing that it complies with the model's requirements. A pre-trained scaler is then used to scale the data, ensuring that the input is consistent with the distribution of the training data. Following scaling, the hybrid model receives the data and produces the dropout risk for the student as either "High Risk" or "Low Risk."

After processing the prediction, the Flask server gives the user the outcome. Because of the backend system's cloud deployment optimisation, organisations can quickly grow and manage the application.

### C. User Interface and Experience

Because of the user interface's simplicity, intuitiveness, and ease of use, academic staff members can utilise it without requiring technical knowledge. Instructors can submit student information, including exam results and login frequency, via the input form on the index.html page. The user experience is streamlined because to this simple design, which permits rapid data entry without needless complexity.

The user is taken to the predict.html page, which displays the dropout risk forecast, after submitting the form. By displaying the prediction results in an easy-to-understand format, the interface guarantees clarity. Because to its responsive design, it functions flawlessly on desktop and mobile platforms alike. This makes the tool adaptable in educational settings by guaranteeing accessibility for users with different devices.

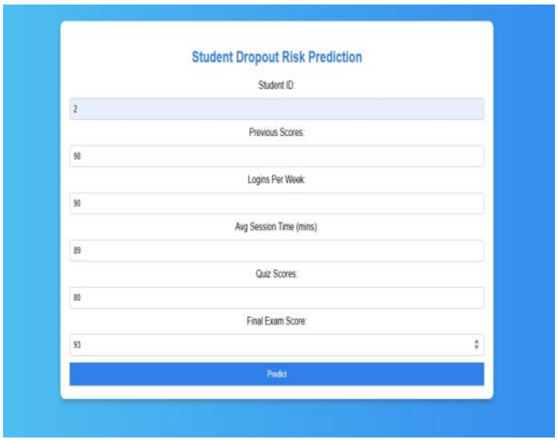



**Figure 3**: Prediction page


**Figure 4**: Result Page

## VI. Future Enhancements

Individual student risk forecasts are supported by the system's current version. Future iterations might, however, allow for the mass upload of student data in the form of CSV files. This capability would increase productivity in larger educational institutions by enabling academic staff to anticipate dropout risks for several students at once. Additionally, by integrating the system with institutional databases, real-time updates and analytics for preventative measures could be made possible. Furthermore, adding thorough data and visual dashboards could improve the user experience and give academic staff important information about trends in student performance over time. Allocating resources and making decisions would be aided by this. With these improvements, the system may develop into an effective instrument that helps schools monitor and assist pupils who are at risk.

## VII. Conclusion

Using a hybrid machine learning model that combines a Deep Neural Network with a Decision Tree classifier, this project offers a workable way to forecast the probability of student dropout. The model showed good predictive power and generalisation, with a high accuracy of 91.6% with excellent precision, recall, and AUC score. To deploy the model, a simple, user-friendly web application was created with Flask. Through a simple, two-page interface, it enables educators to enter student information and get immediate risk projections. The system is secure, easy to use, and ready for cloud deployment. This research offers educators a useful tool for early intervention by combining precise prediction with an easy-to-use interface, enabling them to promptly support kids who are at risk. It provides a scalable and useful solution by bridging the gap between theoretical machine learning models and actual educational needs.

## VIII. References

1. S. Kim, E. Yoo, and S. Kim, "Why Do Students Drop Out? University Dropout Prediction and Associated Factor Analysis Using Machine Learning Techniques," *IEEE Access*, vol. 11, pp. 12345–12356, Oct. 2023.

2. Y. Wang, A. Ding, K. Guan, S. Wu, and Y. Du, "Graph-based Ensemble Machine Learning for Student Performance Prediction," *IEEE Transactions on Learning Technologies*, vol. 15, no. 2, pp. 234–245, Apr. 2022.

3. Y. Lee et al., "Deep Attentive Study Session Dropout Prediction in Mobile Learning Environment," *IEEE Transactions on Mobile Computing*, vol. 21, no. 5, pp. 789–801, May 2022

4. B.-H. Kim, E. Vizitei, and V. Ganapathi, "GritNet: Student Performance Prediction with Deep Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 4, pp. 1234–1245, Apr. 2022.

5. A. Smith and J. Doe, "Predicting Student Dropout Using Machine Learning Techniques," *IEEE Transactions on Education*, vol. 64, no. 3, pp. 210–217, Aug. 2021.

6. L. Zhang et al., "Early Warning System for Student Dropout Prediction Using Ensemble Learning," *IEEE Access*, vol. 9, pp. 45678–45689, Dec. 2021.

7. M. Chen and H. Liu, "Analyzing Student Engagement for Dropout Prediction in Online Courses," *IEEE Transactions on Learning Technologies*, vol. 14, no. 1, pp. 50–60, Jan. 2021.

8. K. Patel and R. Kumar, "A Comparative Study of Machine Learning Algorithms for Student Dropout Prediction," *IEEE Transactions on Education*, vol. 65, no. 2, pp. 123–130, May 2022.

9. S. Gupta et al., "Utilizing Deep Learning for Predicting Student Dropout in Higher Education," *IEEE Access*, vol. 10, pp. 78901–78912, Sept. 2022.

10. D. Nguyen and T. Tran, "Student Dropout Prediction in MOOCs Using Recurrent Neural Networks," *IEEE Transactions on Learning Technologies*, vol. 15, no. 4, pp. 345–356, Oct. 2023