Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #45698

# BOARD # 88: WIP: Detecting Academic Dishonesty in Online Exams Using Machine Learning Techniques

**Sumaya Binte Zilani Choya, George Mason University**
**Dr. Mihai Boicu, George Mason University**

Mihai Boicu, Ph.D., is Associate Professor in the Information Sciences and Technology Department at George Mason University. He is an expert in artificial intelligence, structured analytical methods, probabilistic reasoning, evidence-based reasoning, personalized education, active learning with technology, crowd-sourcing, and collective intelligence. He is the main software architect of the Disciple agent development platform and coordinates the software development of various analytical tools used in IC and education. He has over 150 publications, including 2 books and 3 textbooks. He has received the Innovative Application Award from the American Association for Artificial Intelligence, and several certificates of appreciation from the U.S. Army War College and the Air War College. He is a GMU Teacher of Distinction.

# WIP: Detecting Academic Dishonesty in Online Exams Using Machine Learning Techniques


## Abstract

The COVID-19 pandemic has increased the need for robust methods to uphold academic integrity in online examinations, where issues like impersonation and cheating are prevalent. Most machine learning techniques rely on image or video data, but few are looking at other indicators, such as post-score analysis. This study assesses how proficient machine learning models, such as Random Forest, Support Vector Machine, Logistic Regression, and Gradient Boosting, identify suspicious behaviors based on response accuracy and timing on exam datasets. The Gradient Boosting model achieved the best performance with an accuracy of 97.99% and an F1 score of 98.56%, highlighting the viability of post-score analysis for scalable and reliable academic integrity detection. These findings emphasize the potential of post-score analysis to safeguard the integrity of online education through effective and trustworthy detection techniques.


## Introduction

The COVID-19 pandemic has changed education around the world, accelerating the acceptance of remote exams and online learning environments. This unexpected shift also introduced additional educational challenges, such as upholding academic integrity during online exams [1]. In the absence of conventional in-person supervision, educators must deal with the severe problem of dishonest practices like plagiarism, impersonation, and unapproved collaboration which erode public confidence in online learning and diminish the value of genuine academic accomplishments [2]-[4].

Preserving academic integrity is crucial to maintaining educational institutions' legitimacy and ensuring student evaluations fairly represent learning objectives. Identifying and addressing instances of academic dishonesty not only reinforces fairness in assessments but also cultivates a culture of integrity and accountability among students, ensuring that educational outcomes accurately reflect individual effort and performance [5]-[6].

Machine learning techniques provide scalable and efficient methods for detecting cheating in online exams by analyzing patterns in student performance data [7]. Unlike traditional proctoring methods that rely on visual monitoring, machine learning can detect anomalies in response times, answer accuracy, and response patterns-offering a more privacy-conscious and resource-efficient approach to maintaining academic integrity [8]. This study explores the potential of machine learning models to detect cheating through post-score analysis, addressing a gap in the existing research.

By comparing the performance of algorithms such as Random Forest, Support Vector Machines (SVM), Logistic Regression, and Neural Networks, this research evaluates their effectiveness in identifying suspicious behaviors in online exam data. The findings aim to benefit educators and administrators by offering actionable insights to enhance the fairness and credibility of remote assessments. This study emphasizes how technology can reinforce educational integrity and guarantee that online learning will remain a reliable and fair educational practice.

## Literature review

**Prevalence of academic dishonesty in online learning:** Multiple studies report the increase in incidents of cheating during online examinations [17-19]. In a comprehensive evaluation covering 25 samples with 4,672 individuals, Newton et al. [17] discovered that self-reported cheating increased from 29.9% prior to the pandemic to 54.7% during COVID-19. Additionally, the study shows that individual cheating was more common than group cheating, and that, despite ethical concerns, remote proctoring decreased the risk of misbehavior. In support of this, Janke et al. [18] polled 1,608 university students in Germany and discovered that because online examinations are easier to cheat on and have less oversight, cheating rates are much higher. Patrzek et al. [19] found that 75% of students admitted to some form of academic misconduct, with 36% specifically confessing to cheating on exams. One of the main causes of academic dishonesty, according to their study, is procrastination. These results highlight the importance of the problem and challenge claims that online education guarantees fairness.
**Factors advancing academic dishonesty:** A wide range of factors have influenced academic dishonesty in online learning. Individual, institutional, medium-specific, and assessment-specific factors are the four primary categories into which Holden et al. [2] divide them. Their analysis highlights that detection and prevention procedures are the mainstays of integrity maintenance. In addition, Aristovnik et al. [1] conducted a thorough examination of 8,303 papers related to online education during the COVID-19 epidemic, emphasizing the difficulties in preserving the integrity of assessments, the quick changes in educational methods, and the growing dependence on technology. Their results support the necessity of creative approaches to academic integrity in online settings.

**Online learning concerns:** Toprak et al. [3] highlight that enforcing academic integrity in online learning environments is more challenging due to ethical concerns investigate differences in how students and teachers view privacy and the application of rules. According to their research, 78% of students prefer moderate punishments for misbehavior, but 52% of teachers support harsher punishments. Despite these disagreements, both sides agreed that it is critical to set clear ethical guidelines to maintain a balance between preserving institutional values, promoting student diversity, and protecting privacy.

**Machine learning for cheating detection**: Automated techniques for detecting cheating have been established by the incorporation of machine learning (ML) into academic integrity research. Alsabhan et al. [4] outperformed conventional Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) by recognizing dishonest conduct with 90% accuracy using a Long Short-Term Memory (LSTM)-based model on the 7WiseUp dataset. By combining Kernel Density Estimation (KDE) with LSTMs, Kamalov et al. [8] expanded on this strategy and achieved a 0.05 False Positive Rate (FPR) and a 0.95 True Positive Rate (TPR), which greatly increased detection precision. Their approach effectively identified abnormalities in test results, proving that integrity measures beyond real-time proctoring can be improved by post-score analysis. Zawacki-Richter et al. [5] divided AI-driven solutions for academic integrity into adaptive systems, intelligent tutoring, evaluation, and profiling to highlight the ethical issues surrounding machine learning. Five major AI applications in education were noted by Crompton and Burke [6], who also observed a two to threefold growth in academic integrity research with an AI focus between 2021 and 2022. Their research is in line with that of Moya et al. [7], who looked at 14 papers on the ethical aspects of AI, such as inclusiveness, data privacy, and academic dishonesty.

**Advancements in AI-driven proctoring**: Proctoring systems with AI capabilities have shown excellent accuracy in detecting anomalous activity. By combining face recognition, gaze tracking, and deep learning models (CNN, YOLO v3), Ozdamli et al. [9] were able to achieve 96.95% gaze tracking accuracy and 99.38% identity verification accuracy. In their analysis of 41 deep learning-

based online proctoring studies, Abbas and Hameed [10] showed how well AI models can identify irregularities. In a similar vein, Kaddoura and Gumaei [12] improved test fairness using predictive analytics by utilizing Support Vector Regression Machines (SVRM). Together, these research results highlight AI's expanding use in automated cheating detection. Measures of integrity are further reinforced by recent developments in academic fraud detection using Natural Language Processing (NLP). The first anonymized dataset created specifically for online fraud detection, FraudNLP, was presented by Boulieris et al. [11]. While preserving user anonymity, their approach significantly improved fraud detection accuracy. In their investigation of NLP-based external plagiarism detection, Toprak et al. [13] used dependency parsing and language models to achieve a classification accuracy of 70.53%, particularly effective for identifying paraphrased content. Building on this line of research, Kundu et al. [14] and Masud et al. [15] developed keystroke dynamics-based AI models (modified TypeNet) that successfully identified AI-assisted writing with an accuracy of 85.72%.

**Research gaps**: Most of the previous studies focus on image-based fraud detection and real-time proctoring instead of post-score analysis [16]. Despite the efficacy of AI-powered proctoring, dependence on image-based identification raises privacy issues and may be biased. To close this gap, our study uses machine learning approaches to analyze post-exam performance and detect anomalies in test results that go beyond visual monitoring. Our method assesses statistical suspicious patterns in exam results, offering a scalable and privacy-conscious substitute for prior research that monitors gaze movements and behavior.

**Data preprocessing**

The dataset was downloaded from Kaggle [16] and consists of 1,600 instances with 12 attributes used for classifying exams as either genuine (1,523 cases) or cheating (77 cases). The dataset captures critical aspects of student behavior and performance during exams and includes the following attributes: Student Name (anonymized identifier for each participant), Exam Mode (type of exam environment, e.g., remote or center-based), Total Questions (the total number of questions presented in the exam), Total Questions Attempted (the number of questions a student attempted to answer), Questions Attempted Within Ideal Time (the number of questions answered within a predefined ideal response time), Ideal Time Correct Questions (the number of correctly answered questions within the ideal response time), Total Correct Questions (the overall number of questions answered correctly), Questions to Pass (the minimum number of questions required to pass the exam), Result (the outcome of the exam, e.g., pass or fail), Total Exam Time (the total time taken by the student to complete the exam in minutes), Exam Finish Time in Seconds (the exact time in seconds at which the student completed the exam), and Prediction (the label for the exam as "Genuine" or "Cheating").Label encoding: Categorical variables were converted into numerical values using the LabelEncoder from the sklearn.preprocessing module. This step ensured that the machine learning models could interpret and utilize the categorical data effectively.
Feature scaling: Features have various values, and they are on different scales (some with big values, others with small values). Machine learning models may get confused by the difference between large values in one scale and small values on another scale. For this reason, the feature values are scaled to a small interval (e.g., -1 to 1) with a mean of 0 and standard deviation of 1. This way all features will have the same type of values when used in a machine learning model.
**Feature selection with random forest classifier**

Figure 1 presents the feature importance analysis derived from the Random Forest classifier, which identified Ideal Time Correct Questions, Questions Attempted Within Ideal Time, Exam Finish Time

in Minutes, Result, Total Correct Questions, Total Questions Attempted, Total Exam Time, Total Questions, and Questions to Pass as the most influential predictors. These features collectively capture key aspects of timing, accuracy, and exam-related behaviors, which are essential for predicting outcomes in the model. The selection of these nine features ensures a balance between model simplicity and predictive power by excluding less informative features such as Exam Mode.

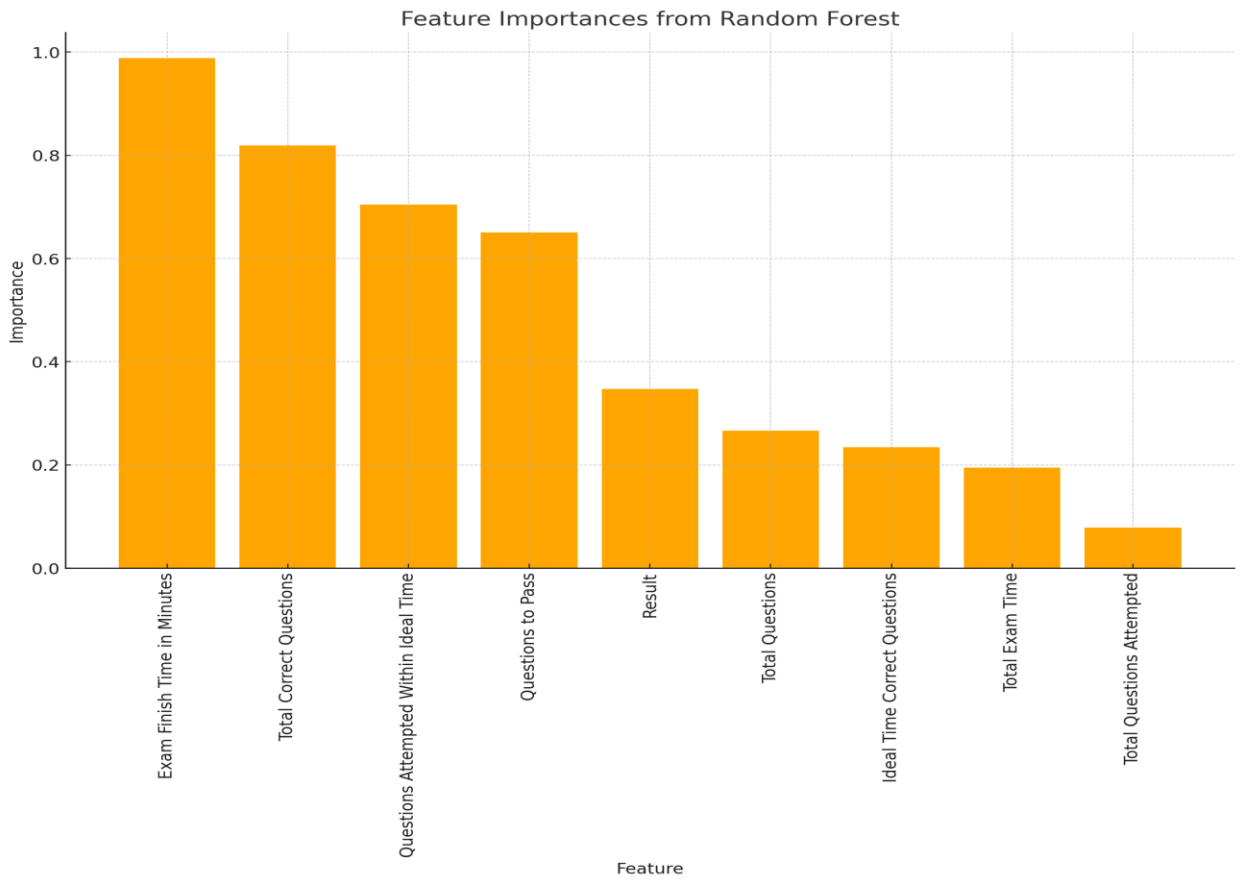


Figure 1. Key features based on their significance in predicting the target variable

We used a process of preprocessing, feature selection, and supervised learning models to assess trends in student exam behavior and forecast results. The dataset was normalized and encoded for numerical stability. It included 10,000 samples and nine key features. Multiple classifiers were trained using the most influential features found through Random Forest feature selection.

**Machine Learning Models**

We applied several machine learning algorithms to build predictive models. We evaluated six machine learning models, including Random Forest, SVM, Gradient Boosting, Logistic Regression, a Neural Network with two hidden layers and Naive Bayes, Cross-validation ensured reliable accuracy metrics, and hyperparameter tuning via GridSearchCV optimized model performance. Each model was chosen for its unique methodological approach and capability to handle diverse data

structures. The models and their configurations are as follows:

**Random Forest:** To achieve reliable classification, Random Forest (RF), an ensemble learning technique, builds several decision trees and combines their results. Random Forest was used in our study to describe non-linear relationships in the data. For RF, the optimal configuration included 100 estimators (n_estimators=100) and an unrestricted tree depth (max_depth=None), enabling the model to capture complex patterns without overfitting. To choose the most important predictors, the model was also used to assess the significance of the features. It became one of the best-performing models with a mean accuracy of 97.19%, handling mixed feature types with ease and avoiding overfitting by ensemble averaging. However, because the Neural Network was able to identify more complex and profound patterns in the dataset, Neural Network fared marginally better than  did.

**Support Vector Machine:** SVM identifies an optimal hyperplane to separate classes, employing kernel functions to handle both linear and non-linear decision boundaries. To successfully handle class independence, SVM was applied in our work using a linear kernel and an optimal penalty value C=1.0. Because it outperformed non-linear kernels in cross-validation experiments, the linear kernel was selected based on the data structure. With a mean accuracy of 96.88%, SVM showed excellent performance in class distinction. In contrast to flexible models like Neural Networks and ensemble approaches like Random Forest, its accuracy was constrained by its sensitivity to outliers and dependence on a fixed hyperplane. Nevertheless, SVM was a useful model in our investigation because of its robustness and simplicity.

**Logistic Regression:** Logistic Regression is a statistical model used for binary classification, predicting the probability of a target variable belonging to a specific class. In our study, Logistic Regression was implemented without hyperparameter tuning, relying on its default configuration to evaluate its baseline performance. The model achieved a mean accuracy of 96.56%, indicating solid performance in linear separability scenarios. However, it struggled with capturing non-linear relationships in the data, which limited its overall effectiveness compared to more advanced models like Random Forest and Neural Networks. Despite its lower accuracy, Logistic Regression's simplicity, interpretability, and efficiency made it an important benchmark for evaluating other classifiers in our analysis.

**K-Nearest Neighbors (KNN):** K-Nearest Neighbors (KNN) is a non-parametric algorithm that classifies samples based on the majority vote of k-nearest neighbors, using distance metrics like Euclidean distance. In this study, k = 5 was used, and the data was standardized to ensure features contributed equally to the distance calculations. KNN achieved an accuracy of 96.56%, indicating moderate performance. Its accuracy was limited by the high dimensionality of the data and overlapping class distributions which can affect the reliability of distance metrics Despite these constraints, KNN remains valuable for identifying local patterns in the dataset.

**Gradient Boosting:** Gradient Boosting emerged as the best-performing model in this study, achieving an impressive accuracy of 97.99% and an F1 score of 98.56%. This superior performance can be attributed to the model's iterative approach, where each tree focuses on correcting the errors of the previous ones, thus capturing complex non-linear relationships in the data. By combining multiple weak learners into a strong predictive model, Gradient Boosting effectively minimizes residual errors and classifies even challenging samples. Additionally, the fine-tuned hyperparameters, including 50 estimators, a learning rate of 0.1, and a maximum depth of 5, optimized its ability to balance model complexity and overfitting. These factors collectively enabled Gradient Boosting to outperform other models, making it an excellent choice for academic integrity detection, where precision and reliability are critical.**Naive Bayes (NB):** Naive Bayes is a

probabilistic classifier that relies on Bayes' theorem and assumes independence between features. This simple yet effective algorithm was implemented without hyperparameter tuning in this study. While it demonstrated moderate accuracy compared to more complex models, its reliance on the independence assumption limited its ability to capture intricate feature relationships. Despite this, Naive Bayes offered computational efficiency and served as a baseline for comparison. It achieved an accuracy of 90.00%, reflecting its utility in straightforward classification tasks with well-separated classes.

**Neural Network (NN):** The Neural Network was designed with two hidden layers, each containing 32 neurons, and employed the ReLU activation function along with dropout layers to mitigate overfitting. Using the Adam optimizer and binary cross-entropy as the loss function, the model was trained over 100 epochs with a batch size of 32. Cross-validation and test set evaluation confirmed its robustness, making it the best-performing model in this study with an accuracy of 97.81%. The NN excelled due to its ability to capture complex non-linear relationships, learning patterns from the data that were inaccessible to other models. Its performance underscores the effectiveness of deep learning in applications requiring detailed feature interaction.

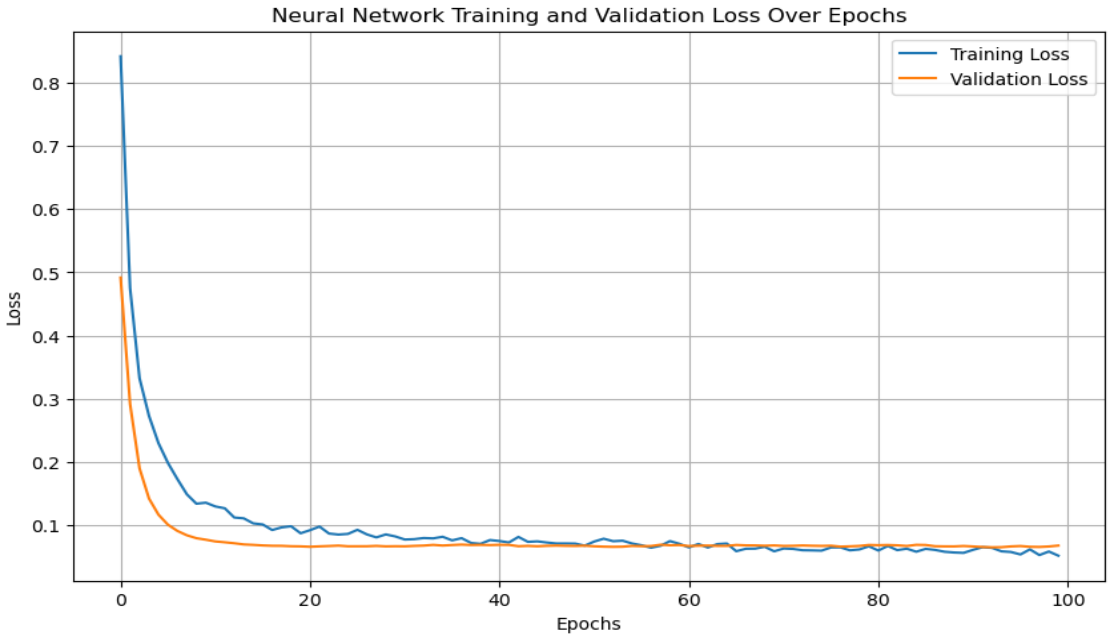


Figure 2. The graph illustrates the decrease in loss over epochs, indicating the neural network's learning progress and optimization.

**Algorithm performance**

This study evaluated the performance of various machine learning algorithms in detecting potential cheating during online examinations. To ensure a comprehensive and balanced assessment, we incorporated feature selection and extended our evaluation beyond accuracy by including additional metrics such as F-Score, Matthews Correlation Coefficient (MCC), false positive rate (FPR), and specificity.as suggested by the reviewers, to go beyond accuracy and include metrics such as F-Score, Matthews Correlation Coefficient (MCC), false positive rate (FPR), and specificity. In Table 3, ensemble methods, particularly Gradient Boosting and Random Forest, achieved the highest accuracies among all classifiers. Gradient Boosting emerged as the best-performing algorithm with a

mean accuracy of 97.99%, closely followed by Neural Network and Random Forest, with mean accuracies of 97.81% and 97.19%, respectively. Logistic Regression also performed competitively, achieving a mean accuracy of 96.56%.

The K-Nearest Neighbors (KNN) classifier demonstrated impressive performance with a mean accuracy of 96.56%, though it was slightly outperformed by ensemble methods and SVM. On the other hand, the Naive Bayes classifier exhibited significantly lower accuracy at 90.00%, due to its assumption of feature independence, which may not hold in the context of this dataset.

Interestingly, the Neural Network, despite its complexity to capture intricate patterns, achieved a mean accuracy of 97.81%, indicating robust performance. However, it was slightly outperformed by Gradient Boosting in this specific scenario.

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | MCC (%) | Specificity (%) | False Positive Rate (%) | False Negative Rate (%) | True Positive Rate (%) |
|---|---|---|---|---|---|---|---|---|---|
| Random Forest | 97.19% | 97.78% | 99.35% | 98.56% | 41.18% | 30.0% | 70.0% | 0.65% | 99.35% |
| SVM | 96.88% | 96.88% | 100.0% | 98.41% | 0.0% | 0.0% | 40.0% | 0.0% | 95.0% |
| Logistic Regression | 96.56% | 97.76% | 98.71% | 98.23% | 34.15% | 30.0% | 70.0% | 1.29% | 98.71% |
| KNN | 96.56% | 96.87% | 99.68% | 98.25% | -1.01% | 0.0% | 42.0% | 0.32% | 99.68% |
| Gradient Boosting | 97.99% | 97.78% | 99.35% | 98.56% | 41.18% | 30.0% | 70.0% | 0.65% | 99.35% |
| Naive Bayes | 90.00% | 98.60% | 90.97% | 94.63% | 28.78% | 60.0% | 40.0% | 9.03% | 90.97% |
| Neural Network | 97.81% | 97.79% | 100.0% | 98.88% | 54.16% | 30.0% | 70.0% | 0.0% | 100.0% |

Table 1: Performance Metrics of Machine Learning Models for Academic Integrity Detection

The Random Forest classifier, benefiting from its ensemble nature, which reduces variance and improves generalization, performed well with an accuracy of 97.19%. Similarly, with accuracies of 96.88% and 96.56%, respectively, Support Vector Machine (SVM) and Logistic Regression achieved competitive results. Despite its simplicity, the K-Nearest Neighbors (KNN) model fared reasonably well (96.56% accuracy); however, the ensemble techniques marginally outperformed it.

However, Naive Bayes had the lowest accuracy (90.00%), most likely due to its high independence assumption, which contradicts the feature correlations in the dataset. Despite its high precision (98.60%), its relatively lower recall (90.97%) suggests difficulty in identifying all instances of cheating accurately.

Fairness in online examinations is maintained by the low false positive rates (FPR), especially for SVM and KNN (0.0%), which indicate no misclassification of honest students as cheaters. Accurate identification of non-cheating instances is ensured by high specificity (SVM: 0.0%, KNN: 0.0%), minimizing unnecessary actions. With high accuracy (≥97%) and great recall (≥99%), ensemble methods—in particular, Random Forest and Gradient Boosting—performed

exceptionally well, successfully balancing false positives and false negatives. MCC values above 41% for Random Forest and Gradient Boosting demonstrate their resilience in handling class imbalances, while F1-scores above 98% confirm their dependability.

The highest overall performance was shown by Gradient Boosting (GB), which achieved a Matthews Correlation Coefficient (MCC) of 41.18%, 97.99% accuracy, and a 98.56% F1-score. However, with an MCC of -1.01%, K-Nearest Neighbors (KNN) produced the poorest performance balance, even though it had a high recall of 99.68%. These models are effective instruments for enforcing academic integrity, as they reliably and accurately detect instances of cheating. This research offers a robust and efficient framework for online examination monitoring by employing advanced machine learning techniques and thorough evaluation.

**Result and discussion**

According to the findings, Gradient Boosting (GB) is the best classifier, achieving the highest accuracy (97.99%), F1-score (98.56%), and Matthews Correlation Coefficient (MCC) (41.18%), while maintaining a relatively low false positive rate (70.0%). Although Random Forest (RF) and Support Vector Machine (SVM) also exhibited high accuracy (97.19% for RF and 96.88% for SVM), their specificity (30.0% for RF and 0.0% for SVM) and false positive rates (70.0% for RF and 40.0% for SVM) were lower. Despite its superior specificity (60.0%), which reduces false positives, Naïve Bayes (NB) had the lowest overall accuracy (90.00%). The K-Nearest Neighbors (KNN) classifier showed a balanced performance, with an accuracy of 96.56% and an MCC of -1.01%. Neural Network (NN) performed well, achieving an accuracy of 97.81% and an F1-score of 98.88%, with the highest MCC (54.16%), making it a strong alternative for robust classification. This study demonstrates how academic dishonesty may be identified by post-score analysis utilizing machine learning, namely GB, without the need for visual proof. By comparing multiple classifiers across various evaluation metrics, this research provides practical insights into trade-offs between accuracy, specificity, and false positive rates. These findings offer a valuable foundation for implementing scalable, data-driven solutions to support academic integrity in online education.

**Real world implementation**

In order to implement these methods in a real educational environment, the tests must be done through a testing system that allows the capture of response time durations and to have a procedure to compute the remaining features. Several learning management systems have the possibility to download the students' response for each question and to have the total time the student spent on each question (e.g., Blackboard Learning Management System). To compute the ideal time per each question, a suggested method is to average the response from all students and to take a proportional larger interval as default ideal time (e.g., mean*0.5 to mean*1.5 but will vary depending on the specific question difficulties). A too long response time may indicate that the student tried to obtain the result from other sources, while a too fast response may indicate that the student used an AI system to provide the answer. Based on these data, one may aggregate the features used in these models: Ideal Time Correct Questions, Questions Attempted Within Ideal Time, Exam Finish Time in Minutes, Result, Total Correct Questions, Total Questions Attempted, Total Exam Time, Total Questions, and Questions to Pass.

Once the data is computed, the model can be applied, and the model will identify potential cheating cases. Because the model has a high false positive rate, it is essential that these cases are further

analyzed by the instructor. An ideal validation will be a direct discussion with the student to check that the student has the knowledge to solve the questions marked as outside the ideal time. Both cheating and non-cheating cases can now be labeled and added to the training data. This way the model will improve over time and there will be fewer misclassified instances.

## Conclusion and future work

Overall, the results show that patterns of academic dishonesty in online assessments may be successfully identified using machine learning approaches combined with appropriate feature selection and assessment measures. Through examining post-score behavioral data, including response time, accuracy patterns, and question completion patterns, the algorithms identified cases of cheating reliably and precisely. These results demonstrate how data-driven strategies may improve online examination integrity while lowering dependency on conventional proctoring techniques.

Future research will extend this approach to a variety of academic courses by adding subject-specific elements like response patterns and question difficulty, guaranteeing cross- disciplinary flexibility. Learning management system integration with real-time analysis will improve scalability and facilitate wider use in online learning.

## References

[1] A. Aristovnik, K. Karampelas, L. Umek, and D. Ravšelj, "Impact of the COVID-19 pandemic on online learning in higher education: a bibliometric analysis," Front. Educ., vol. 8, Article 1225834, Aug. 2023, doi: 10.3389/feduc.2023.1225834.

[2] O. L. Holden, M. E. Norris, and V. A. Kuhlmeier, "Academic Integrity in Online Assessment: A Research Review," Front. Educ., vol. 6, Art. no. 639814, Jul. 2021, doi: 10.3389/feduc.2021.639814.

[3] E. Toprak, B. Özkanal, S. Aydin, and S. Kaya, "Ethics in E-Learning," Turkish Online Journal of Educational Technology, vol. 9, no. 2, pp. 78, Apr. 2010.

[4] W. Alsabhan, "Student Cheating Detection in Higher Education by Implementing Machine Learning and LSTM Techniques," Sensors, vol. 23, no. 8, Art. no. 4149, Apr. 2023, doi: 10.3390/s23084149.

[5] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?" Int. J. Educ. Technol. High. Educ., vol. 16, Art. no. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[6] H. Crompton and D. Burke, "Artificial intelligence in higher education: the state of the field," Int. J. Educ. Technol. High. Educ., vol. 20, Art. no. 22, 2023, doi: 10.1186/s41239-023-00392-8.

[7] B. A. Moya, S. E. Eaton, H. Pethrick, K. A. Hayden, R. Brennan, J. Wiens, and B. McDermott, "Academic integrity and artificial intelligence in higher education contexts: A rapid scoping review," Can. Perspect. Acad. Integr., vol. 7, no. 3, pp. 1-19, 2023, doi: 10.55016/ojs/cpai.v7i3/78123.

[8] F. Kamalov, H. Sulieman, and D. S. Calonge, "Machine learning based approach to exam cheating detection," PLoS ONE, vol. 16, no. 8, Art. no. e0254340, Aug. 2021, doi: 10.1371/journal.pone.0254340.

[9] F. Ozdamli, A. Aljarrah, D. Karagozlu, and M. Ababneh, "Facial Recognition System to Detect Student Emotions and Cheating in Distance Learning," Sustainability, vol. 14, no. 20, Art. no. 13230, 2022, doi: 10.3390/su142013230.

[10] M. A. E. Abbas and S. Hameed, "A Systematic Review of Deep Learning Based Online Exam Proctoring Systems for Abnormal Student Behaviour Detection," Int. J. Sci. Res. Sci. Eng. Technol., vol. 9, no. 4, pp. 192-209, Jul.-Aug. 2022, doi: 10.32628/IJSRSET229428.

[11] P. Boulieris, J. Pavlopoulos, A. Xenos, and V. Vassalos, "Fraud detection with natural language processing," Mach. Learn., vol. 113, pp. 5087–5108, 2024, doi: 10.1007/s10994-023-06354-5.

[12] S. Kaddoura and A. Gumaei, "Towards effective and efficient online exam systems using deep learning-based cheating detection approach," Intell. Syst. Appl., vol. 16, Art. no. 200153, Nov. 2022, doi: 10.1016/j.iswa.2022.200153.

[13] M. Chong, L. Specia, and R. Mitkov, "Using Natural Language Processing for Automatic Detection of Plagiarism," Research Group in Computational Linguistics, University of Wolverhampton, UK. [Online]. Available: https://wlv.ac.uk/. [Accessed: Jul. 26, 2023].

[14] D. Kundu, A. Mehta, R. Kumar, N. Lal, A. Anand, A. Singh, and R. R. Shah, "Keystroke Dynamics Against Academic Dishonesty in the Age of LLMs," presented at MIDAS Lab, IIIT Delhi, India, and Bucknell University, USA. [Online]. Available: https://iiitd.ac.in/midas. [Accessed: Jul. 26, 2023].

[15] M. M. Masud, K. Hayawi, S. S. Mathew, T. Michael, and M. ElBarachi, "Smart Online Exam Proctoring Assist for Cheating Detection," presented at the College of Information Technology, United Arab Emirates University, UAE; College of Technological Innovations, Zayed University, UAE; and Faculty of Engineering & Info. Sciences, The Univ. of Wollongong in Dubai, UAE. [Online]. Available: https://uaeu.ac.ae/. [Accessed: Jul. 26, 2023].

[16] [Online]. Available: https://www.kaggle.com/datasets/sansjain/online-exam-data.

[17] P. M. Newton and K. Essex, “How common is cheating in online exams and did it increase during the COVID-19 pandemic? A systematic review,” J. Acad. Ethics, vol. 22, pp. 323–343, 2024. Available:https://doi.org/10.1007/s10805-023-09485-5

[18] S. Janke, S. C. Rudert, A. Petersen, T. M. Fritz, and M. Daumiller, “Cheating in the wake COVID-19: How dangerous is ad-hoc online testing for academic integrity?,” *Computers and Education Open*, vol. 2, p. 100055, 2021. Available:https://www.sciencedirect.com/science/article/pii/S2666557321000264

[19] J. Patrzek, S. Sattler, F. van Veen, C. Grunschel, and S. Fries, “Investigating the effect of academic procrastination on the frequency and variety of academic misconduct: A panel study,” Stud. High. Educ., vol. 39, no. 4, pp. 1–17, 2014. Available:

https://www.tandfonline.com/doi/full/10.1080/03075079.2013.854765