Proceedings of the International Conference on Education, Vol. 11, Issue 1, 2025, pp.1-17

ISSN 2424-6700 online
DOI: https://doi.org/10.17501/24246700.2025.11101

COMPARATIVE STUDY IN PREDICTIVE ANALYTICS OF PERSONALIZED LEARNING AND TRADITIONAL CURRICULUM IN MEDICAL EDUCATION

Tangpanithandee S[1], Kitsiranuwat S[1], Chantra M[2], Puranitee P[2], and Hongwarittorrn N[1]*

[1]*Department of Computer Science, Faculty of Science and Technology, Thammasat University, Thailand*
[2]*Faculty of Medicine Ramathibodi Hospital, Mahidol University, Thailand.*

Abstract: Personalized education, known as Student-Selected Components (SSCs), empowers students to partially design their study plans, allowing for individualized learning experiences. In medical education, SSCs aim to provide self-directed learning, but evaluating student performance within such diverse curricula is challenging. Recently, an SSC-based curriculum was implemented at Ramathibodi Hospital, Mahidol University, Thailand. This study aims to assess the impact of the SSC curriculum on medical student performance by comparing failure rates on comprehensive exams with those of the traditional curriculum. Additionally, machine learning techniques were developed to identify key factors influencing learning outcomes and to predict the likelihood of fifth-year medical students passing or failing their exams. A dataset of 205 fifth-year medical students, encompassing 20 demographic and academic variables, was analyzed. The hypothesis testing and analysis were conducted by open-source tools including Python and Google Colab. The hypothesis testing result indicated that the SSC-based curriculum exhibited a slightly higher failure rate compared to traditional cohorts; however, this difference was not statistically significant. Support Vector Machine (SVM), Decision Tree, Random Forest, Adaptive Boosting, and Extreme Gradient Boosting were employed and tuned with various data-balancing techniques in order to get best results. Those models were evaluated based on F1-score. The SVM model outperformed with the highest F1-score of 84.8%. Key predictors included GPA of pre-clinical years, followed by grades in Neurology, Ophthalmology-Otolaryngology, Pediatrics, and Mother-Baby care. These findings suggest potential challenges in adapting to the increased self-directed nature of the SSC curriculum. Although the predictive models could effectively identify students at risk of failing, enabling targeted early interventions, it could be improved with additional data on student performance metrics. While the SSC curriculum offers greater flexibility and personalization, further refinements are necessary to optimize learning outcomes and ensure consistent quality in medical education.

Keywords: curriculum assessment, machine learning, medical education, personalized learning, predictive modeling

Introduction

Student-Selected Components (SSCs), sometimes called special study modules or selective study units, provide undergraduate medical students with dedicated periods to explore medical fields of personal interest (Clark et al., 2016). First introduced in the UK during the 1990s (Riley, 2009), SSCs were part of a broader initiative by the General Medical Council (GMC) to reform medical education through its 'Tomorrow's Doctors' framework. This reform proposed that medical curricula should consist of two main parts: a core curriculum accounting for two-thirds of the program and an optional, flexible component for the remaining third (Riley, 2009).

The success of SSCs in the UK has led to their adoption in various countries, including Ireland, Brazil, and Malaysia (Falk, Robb, Khan, & Hill, 2009; Sobral, 2008; Thomas, Dhanoa, & Palanisamy, 2012). Evidence from these regions indicates positive educational outcomes, such as increased Likert scale

*Corresponding Author's Email: *nuttanon@tu.ac.th

scores for skills development and greater student confidence in beginning clinical work (Falk et al., 2009). Additionally, SSCs have been reported to inspire students to pursue specialized careers, such as surgery or neurosurgery (Clark et al., 2016).

In 2019, the Faculty of Medicine at Ramathibodi Hospital, Mahidol University, Thailand, introduced major SSCs, referred to as "Selectives." These were considered in-house electives that provided advanced specialized knowledge, increasing the proportion of personalized study options, including research opportunities, while ensuring that students met the necessary requirements to become doctors. However, assessing the performance of students in such varied curricula poses a significant challenge (Riley, 2009). To address this, the failure rates of the comprehensive examination will be used as the key outcome measure, providing a common benchmark for evaluating student performance.

Comprehensive examinations are vital in medical education as they determine whether students are ready to transition from the academic environment to professional clinical practice. These exams are essential in assessing students' readiness to assume the responsibilities of medical professionals, making it also crucial to identify the factors that most influence exam outcomes. Numerous studies have explored these factors, which are key to improving educational strategies and enhancing medical student performance (Abdu, 2024; Adam et al., 2015; Al Shawwa et al., 2015; Dickman, Sarnacki, Schimpfhauser, & Katz, 1980; Dunleavy, Kroopnick, Dowd, Searcy, & Zhao, 2013; Kleshinski, Khuder, Shapiro, & Gold, 2009; Kreiter & Kreiter, 2007; Puddey & Mercer, 2014).

The role of technology in education, especially Artificial Intelligence (AI) and Machine Learning (ML), is transformative. AI and ML technologies provide powerful tools to analyze large datasets and generate new insights beyond conventional methods (Chassignol, Khoroshavin, Klimova, & Bilyatdinova, 2018; Hoti, Zenuni, Hamiti, & Ajdari, 2023; Khudhur & Ramaha, 2023; Perez, Domínguez, Omatu, Herrera-Viedma, & Corchado Rodríguez, 2021). In the educational field, ML has been widely applied to improve learning experiences, predict student performance, and optimize outcomes (Anuradha & Thambusamy, 2015; Asif, Merceron, Abbas, & Haider, 2017; Asif, Merceron, & Pathan, 2014; Chassignol et al., 2018; Garg, 2018; Mesaric & Šebalj, 2016; Mohamed & Waguih, 2017; Mueen, Zafar, & Manzoor, 2016; Putpuek, Rojanaprasert, Atchariyachanvanich, & Thamrongthanyawong, 2018; Singh & Kaur, 2016; Sivasakthi, 2017). These technologies allow for the analysis of extensive datasets, enabling more tailored and effective interventions that can help students succeed (Birihanu & Akmel, 2017; Hussain et al., 2019; Pallathadka et al., 2021). Predictive analytics, particularly through ML, can identify students who may need additional support, allowing for timely intervention and improving overall performance (Hussain et al., 2019; Ofori, Maina, & Gitonga, 2020).

In medical education, ML is emerging as a critical tool for analyzing student performance data, predicting outcomes on key assessments such as licensure exams, and providing early intervention opportunities (Mastour, Dehghani, Moradi, & Eslami, 2023). By accurately predicting which students may struggle, educators can take proactive measures to provide the necessary support before students are at risk of failing. This ensures better preparation and increases the likelihood of success, helping medical students become more competent professionals.

Systematic reviews showed that neural networks are the preferred method for analyzing student data, providing superior accuracy in predicting academic results (Alalawi, Athauda, & Chiong, 2023). The predictive capacity of machine learning extends to evaluating student engagement and performance, offering valuable insights for academic support and planning. In the study by Ramaswami et al., machine learning was employed to forecast students' final academic performance, utilizing data from their interactions with the learning management system (LMS). The accuracy of these predictions was around 75%, with slight improvements as the course progressed (Ramaswami, Susnjak, Mathrani, & Umer, 2020). Another comprehensive study conducted at Eindhoven University of Technology analyzed 17 different LMS indicators to distinguish between strong and weak predictors of student performance (Conijn, Snijders, Kleingeld, & Matzat, 2017). Additionally, Chango and colleagues enhanced prediction methods by integrating various data types, including classroom recordings, online forum participation, and quiz results, to assess student outcomes more effectively (Chango, Cerezo, & Romero, 2021) These studies underscore machine learning's potential in utilizing diverse educational data to predict academic success, thereby providing critical insights for optimizing student engagement, performance, and educational planning.

As shown in Table 1, Historical academic performance has been identified as a key predictor of future success, highlighting the link between past and present academic achievements (Al-Barrak & Al-Razgan, 2016; Almarabeh, 2017; Anuradha & Thambusamy, 2015; Asif et al., 2017; Asif et al., 2014; Garg, 2018; Mesaric & Šebalj, 2016; Mohamed & Waguih, 2017; Mueen et al., 2016; Singh & Kaur, 2016; Sivasakthi, 2017). Demographic factors, including gender, age, and socioeconomic status, also play a crucial role in influencing educational outcomes, suggesting the importance of a comprehensive approach to student support (Ahmad, Ismail, & Aziz, 2015; Almarabeh, 2017; Anuradha & Thambusamy, 2015; Garg, 2018; Mohamed & Waguih, 2017; Sivasakthi, 2017).

Table 1. Examples of literatures using machine learning to predict students' outcome

Author	**Year**	**Predictive Factors**	**Predictive Model**
Ahmed, E. (Abdu, 2024)	2024	gender, region, entrance_result, num_of_prev_attempts, studied_credits, disability	Decision trees, Naïve Bayes, K-nearest neighbors, Support vector machine
Mastour, H., et al. (Mastour et al., 2023)	2023	gender, GPA, residency status, entrance semester, number of attempts, scores in intended course, age at entrance, age when taking the CMBSE	Logistic regression, Support vector machine, K-nearest neighbors, Random forest,

			Adaptive boosting, Extreme gradient boosting
Ramaswami, G.S., et al. (Ramaswami et al., 2020)	2020	assignment score, weekly quizzes, assignment viewed, quiz viewed, age, folder viewed, book viewed	Naïve Bayes, Random forest, Logistic regression, K-nearest neighbors
Mesaric, J., et al. (Mesaric & Šebalj, 2016)	2016	highschool, highschool program, state exam, state exam total points, foreign language, math, enrollment status	Decision tree, Random tree, Random forest
AI-Barrak, M.A., et al. (Al-Barrak & Al-Razgan, 2016)	2016	final GPA, semester of graduation, major, nationality, campus, course taken, course' grade	Decision tree
Anuradha, C., et al. (Anuradha & Thambusamy, 2015)	2015	gender, high school grade, living location, living in hostel, family size, class test grade, seminar performance, general proficiency, attendance	Decision tree, Bayesian classifiers, K-nearest neighbor, Rule learners
Admad, F., et al. (Ahmad et al., 2015)	2015	first year GPA, gender, family income, university entry mode, english, mathematic	Decision tree, Naïve Bayes, Rule based

Additionally, environmental factors which are program types and class structures significantly impact student success (Mesaric & Šebalj, 2016; Mohamed & Waguih, 2017; Mueen et al., 2016; Sivasakthi, 2017). Recent studies question the use of traditional pre-matriculation scores, such as MCAT results and undergraduate GPAs, as predictors of academic success in medical education (Adam et al., 2015; Al Shawwa et al., 2015; Dickman et al., 1980; Puddey & Mercer, 2014). While there are some hypotheses about the predictive value of such pre-admission metrics, a consensus is emerging on the necessity to incorporate a wider array of factors. These factors include not only entrance scores and GPAs but also internal exam results, study habits, and even social networking behaviors (Adam et al., 2015; Al Shawwa et al., 2015; Collins, White, & Kennedy, 1995; Dickman et al., 1980; Dunleavy et al., 2013; Kleshinski et al., 2009; Kreiter & Kreiter, 2007; Puddey & Mercer, 2014; Siu & Reiter, 2009). Those studies drive a move towards a more holistic approach in evaluating predictors of student success, recognizing the complexity and multi-dimensional nature of academic achievement in the medical field. Furthermore, the importance of this research is underscored by the limited existing evidence on the application of machine learning techniques to evaluate and enhance personalized curricula, particularly within the complex domain of medical education.

The current study aims to advance this field by applying machine learning to predict the performance of medical students, utilizing a wide range of predictive factors to enhance the accuracy of interventions and educational results.

The primary objective of this research is to assess the impact of the newly introduced SSC-based curriculum at Ramathibodi Hospital, Mahidol University, by comparing it with the previous curriculum in terms of failure rates on comprehensive exams. As a secondary objective, this study will use machine learning techniques to develop a predictive model to determine whether fifth-year medical students will pass or fail their comprehensive exams and explores for potential predictive factors. By incorporating various factors, including demographic and academic data, the model aims to improve prediction accuracy and provide valuable insights to enhance student performance.

Materials and Methods

This study evaluates the impact of a newly implemented SSC-based curriculum by comparing comprehensive examination outcomes with those from previous traditional curricula and by developing machine learning (ML) models to predict student performance. The methodology integrates data collection, preprocessing, handling of imbalanced data, model development, and performance evaluation. The conceptual framework of this study is shown in Figure 1. The study protocol was approved by Ramathibodi Hospital Institutional Review Board. (COA.No.MURA2025/427).

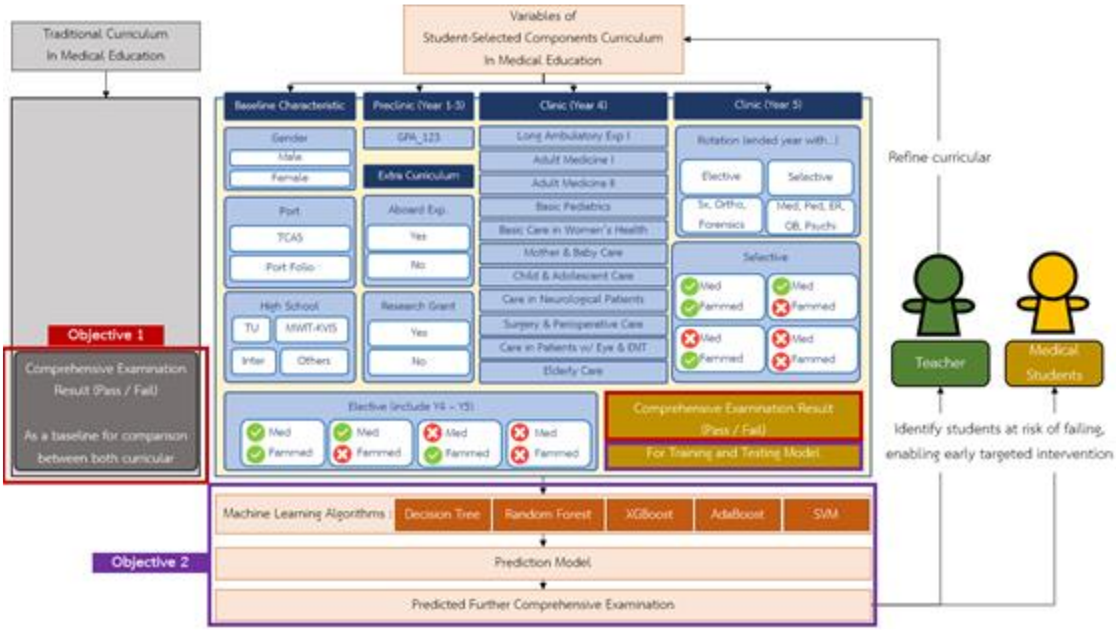

Figure 1. The conceptual framework in this study

Datasets and Study Population

Data were collected from the Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Thailand. The primary dataset comprises records from 209 fifth-year medical students enrolled in the newly SSC-based curriculum in year 2019. In addition, data were obtained from five groups of students following the five traditional curriculum cohorts spanning year 2014–2018.

Demographic information and background data of the SSC-based students are also shown in Figure 1. This dataset includes 20 features—such as gender, admission type, high school background, cumulative preclinical GPA, clinical subject grades, elective and selective choices, and extracurricular participation—with the outcome defined as a pass or fail in the comprehensive examination. The ML process is demonstrated in Figure 2.

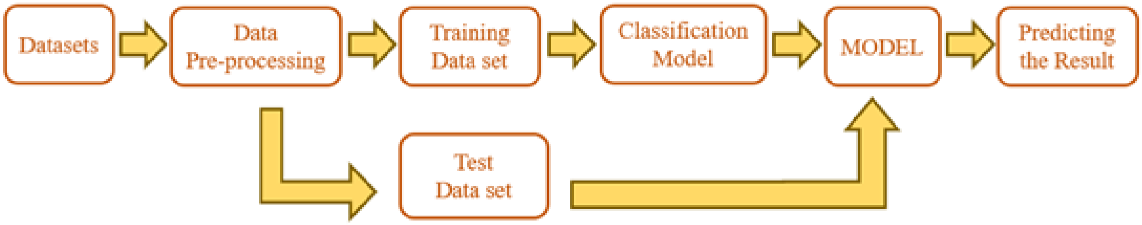

Figure 2. The machine learning workflow in this research

Data Preprocessing and Transformation

The integrity of the input data significantly influences the effectiveness of the predictive model, underscoring the critical role of data preparation. Initially, the dataset included records from 209 fifth-year medical students. However, to ensure data consistency and relevance, entries from three students enrolled in a previous curriculum and one student who did not complete the test were excluded, resulting in a refined dataset comprising 205 students.

To facilitate analysis, it was essential to convert student background information into a format suitable for machine learning models. This transformation involved converting qualitative string data into nominal values. For instance, the gender attribute was coded numerically, with 'male' represented as '1' and 'female' as '0'. Similarly, academic grades for each subject were standardized to a numerical scale to maintain uniformity across the dataset. Specifically, grades were quantified as follows: 'A' was converted to 4.0, 'B+' to 3.5, 'B' to 3.0, 'C+' to 2.5, 'C' to 2.0, 'D+' to 1.5, 'D' to 1.0, and 'F' to 0. Moreover, the cumulative Grade Point Average (GPA) from the first to the third year was categorized into four distinct levels based on the score range: 2.00-2.49, 2.50-2.99, 3.00-3.49, and 3.50-4.00. This categorization not only simplified the analysis but also allowed for a clearer segmentation of student performance over time, which is vital for understanding trends and patterns that might influence their outcomes in the comprehensive examination.

For the Selective component, students were offered six subjects to choose from: Internal Medicine (also known as Medicine), Family Medicine, Obstetrics & Gynecology, Orthopedics & Rehabilitation, Pediatrics, and Surgery. Each student was required to select three out of these six subjects. Thus, it will be 20 possible groups as the combination output for each student. In this study, we categorized all 20 possible combinations into 4 groups by using only Medicine and Family Medicine. This is due to the fact that these two subjects are the major part of the comprehensive examination. As detailed in Table

2, these groups of selective choices included 1. Those with both Medicine and Family Medicine, 2. Those with only Medicine (without Family Medicine), 3. Those with only Family Medicine (without Medicine), and 4. Those that did not include either of these two subjects. This classification aimed to assess the potential impact of these key subjects on students' comprehensive exam performance.

Table 2. Group details showing categorization members of selective combinations

Group	Sub group	Selective 1	Selective 2	Selective 3
1. Have both Medicine and Family Medicine	1	Medicine	Family Medicine	OB-GYN
	2	Medicine	Family Medicine	Pediatrics
	3	Medicine	Family Medicine	Orthopedics
	4	Medicine	Family Medicine	Surgery
2. Have Medicine (without Family Medicine)	5	Medicine	OB-GYN	Pediatrics
	6	Medicine	OB-GYN	Orthopedics
	7	Medicine	OB-GYN	Surgery
	8	Medicine	Pediatrics	Orthopedics
	9	Medicine	Pediatrics	Surgery
	10	Medicine	Orthopedics	Surgery
3. Have Family Medicine (without Medicine)	11	Family Medicine	OB-GYN	Pediatrics
	12	Family Medicine	OB-GYN	Orthopedics
	13	Family Medicine	OB-GYN	Surgery
	14	Family Medicine	Pediatrics	Orthopedics
	15	Family Medicine	Pediatrics	Surgery
	16	Family Medicine	Orthopedics	Surgery
4. Have none of both Medicine and Family Medicine	17	OB-GYN	Pediatrics	Orthopedics
	18	OB-GYN	Pediatrics	Surgery
	19	OB-GYN	Orthopedics	Surgery
	20	Pediatrics	Orthopedics	Surgery

Train-Test Dataset Splitting with Synthetic Minority Oversampling Technique (SMOTE)

The data were split into training and testing sets using an 80:20 ratio. This split was stratified, ensuring that this 80:20 proportion was maintained for instances within both the 'pass' and 'fail' groups respectively, thus preserving the original class distribution in both sets. To address the overall class imbalance in the training data (where 'pass' instances significantly outnumbered 'fail' instances) and enhance the predictive model's robustness, three distinct techniques were applied using the Synthetic Minority Oversampling Technique (SMOTE), a method specifically designed for handling such imbalanced data. The first method involved applying SMOTE solely to the imbalanced training set. The second method extended the use of SMOTE to both the imbalanced training and testing sets. The third method began with an undersampling approach to balance the testing set first. This undersampling process was repeated five times to ensure thoroughness. Subsequently, SMOTE was applied to the remaining data in the training set. Each of these approaches aimed to create a more balanced dataset that would enable more accurate modeling and analysis. Illustrations of these three techniques can be found in Figure 3.

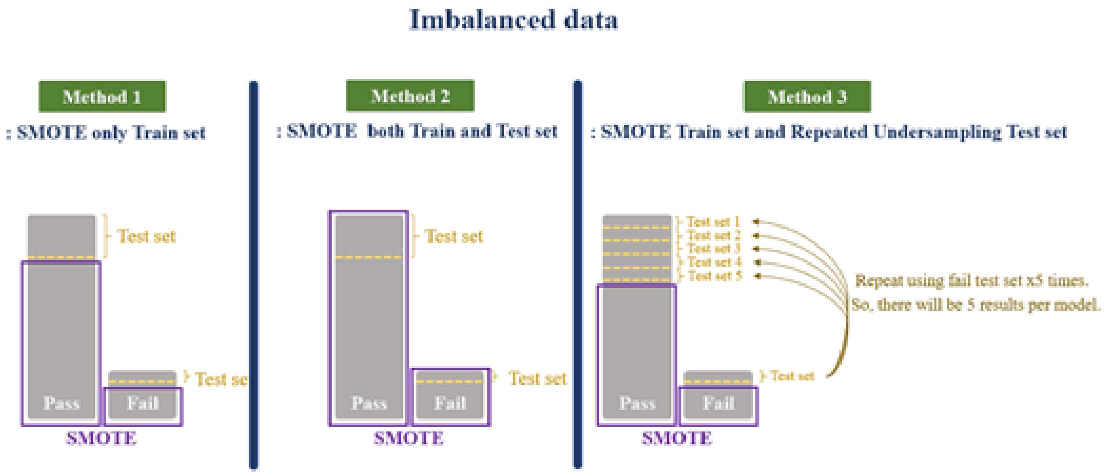

Figure 3. Three methods employed to deal with the imbalanced data.

Prediction Models

Five machine learning prediction/classification algorithms are utilized in this study comprising Decision Tree, Random Forest, Extreme Gradient Boosting (XGBoost), Adaptive Boosting (Adaboost), and Support Vector Machines (SVM). These algorithms are employed due to their excellent rule-based modeling abilities for classification-type prediction issue.

Performance Measure

In this research, the efficacy of the classification strategy was evaluated using a confusion matrix. This tool effectively illustrates the strengths and weaknesses of the classification model by providing detailed insights into its performance. Key performance metrics used include accuracy, precision, recall, and F1-score. Accuracy represents the ratio of correctly predicted observations to the total number of observations. High accuracy often suggests that the model performs well (Baldi, Brunak, Chauvin, Andersen, & Nielsen, 2000). While these models are intuitive and relatively easy to visualize, decision trees are prone to overfitting when dealing with extensive feature sets or complex datasets (Farhood, Joudah, Beheshti, & Müller, 2024). Hence, additional metrics are essential for a more comprehensive performance evaluation (Baldi et al., 2000).

Precision is defined as the proportion of true positive observations out of all positive predictions made by the model. Recall measures the proportion of true positive observations out of all actual positives in the dataset (Powers, 2008). Specifically focused in this study, the F1-score, a balanced measure, is the harmonic mean of precision and recall, serving as a critical metric for the evaluation of binary classification systems. The performance of the model in this study is quantified using these parameters, detailed in equations (1)-(4)

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\textbf{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\textbf{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\textbf{F1-score} = \frac{2*Precision*Recall}{Precision+Recall} \quad (4)$$

where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

Results and Discussion

This part presents the main findings of the study by comparing examination outcomes between the SSC-based curriculum and the traditional curriculum, evaluating the performance of various machine learning (ML) models under different data balancing techniques, and exploring potential predictors of student success.

Assessment of SSC-based Curriculum vs. Traditional Curriculum

Table 3 summarizes failure rates across six academic years—recent five traditional curriculum cohorts (year 2014–2018) and one newly SSC-based curriculum cohort (year 2019). The SSC-based curriculum showed a failure rate of 17.07% (35/205 students). In contrast, the traditional curriculum, after balanced sampling, had an average failure rate of 13.17 ± 2.36%. Statistical analysis ($p = 0.270$) indicated that the difference was not statistically significant. It is important to note that some traditional-cohort failure rates were greatly affected by external circumstances. For example, students whose application year is in 2017—during COVID-19 in their clinical year—failure rates reached 30.22%, likely due to the shift from hands-on clinical learning to remote methods and the added stress of a global health crisis. In comparison, the SSC-based group's failure rate (17.07%) sits below these peak pandemic-era rates, although it is slightly higher than the overall average for the traditional curriculum (13.17%). This outcome may indicate an adjustment period as students' transition to a more self-directed learning environment.

Table 3. Failure rates compared among traditional curriculum (before COVID pandemic), traditional curriculum (during COVID pandemic), and new curriculum

Application Year	**Traditional Curriculum**		**New Curriculum**		**p-value**
	Failure ratio*	Failure rates (%)	Failure ratio*	Failure rates (%)	
2014	4/172	2.33	-	-	

2015	7/158	4.43	-	-	
2016 (COVID)	11/169	6.51	-	-	
2017 (COVID)	55/182	30.22	-	-	
2018 (COVID)	45/196	22.96	-	-	
2019	-	-	35/205	17.07	
Total Failure	122/877	13.91	35/205	17.07	
Balanced Sampling	**27/205**	**13.17±2.36**	**35/205**	**17.07±2.63**	**0.270**

* *The failure ratio was defined as the proportion of medical students who did not pass among those who took the comprehensive examination.*

Interestingly, while the failure rate in the SSC-based curriculum is higher than that observed in the old curriculum's pre-pandemic years, it is lower than the pandemic-affected cohorts. This suggests that while the SSC-based curriculum offers more personalized learning opportunities, it may also introduce variability in academic performance, possibly due to differences in self-directed learning efficacy among students. Further investigation is needed to determine whether this higher failure rate is an artifact of the early stages of curriculum implementation or if it reflects a need for curricular refinement to optimize learning outcomes.

Comparison of Data Splitting Techniques and Machine Learning Algorithms

SSC-based curriculum was explored deeper by machine learning. A detailed comparative study was carried out to assess the predictive power of several ML algorithms, including Decision Trees, Random Forest, XGBoost, AdaBoost, and Support Vector Machines (SVM), in forecasting student performance outcomes. The evaluation of these algorithms was based on metrics comprising accuracy, precision, recall, and F1-score. The results in Table 4 were categorized according to the three data splitting techniques described earlier. The findings indicated that Method 2 yielded the most favorable overall results, however, it could potentially be influenced by the synthesized data in the test set through SMOTE. Subsequently, SVM demonstrated the highest prediction score in Method 3, achieving an F1-score of 84.8%, with Random Forest following at 77.0% F1-score.

Table 4. Model Performances separated into three SMOTE methods for imbalanced data

Prediction method	**Accuracy**	**Precision**	**Recall**	**F1-score**

	(%)	(%)	(%)	(%)
Method 1: SMOTE only train set				
- Decision Tree	63.0	10.0	14.0	12.0
- Random Forest	80.0	33.0	14.0	20.0
- XGBoost	73.0	17.0	14.0	15.0
- Adaboost	71.0	22.0	29.0	25.0
- Support Vector Machines	73.0	33.0	59.0	42.0
Method 2: SMOTE both train and test set				
- Decision Tree	78.0	85.0	68.0	75.0
- Random Forest	88.0	86.0	91.0	89.0
- XGBoost	84.0	79.0	91.0	85.0
- Adaboost	84.0	81.0	88.0	85.0
- Support Vector Machines	78.0	77.0	79.0	78.0
Method 3: 5x-undersampling test set and SMOTE train set				
- Decision Tree	68.4	76.8	57.0	65.2
- Random Forest	78.6	85.0	71.0	77.0
- XGBoost	75.8	78.2	71.0	74.6
- Adaboost	52.8	58.2	29.0	37.8
- Support Vector Machines	84.6	83.8	86.0	84.8

This analysis confirms the robustness of machine learning models like SVM, Decision Tree, and others in predicting educational outcomes, aligning with previous studies. For instance, research by Esmael Ahmed highlighted the SVM's superior performance post-parameter optimization, achieving a 96% accuracy, outperforming Decision Tree, Naïve Bayes, and K-nearest neighbor algorithms (Abdu, 2024).

Identifying Potential Predicting Factors by Feature Importance

Following the results presented in section 3.2, further analysis was undertaken to explore the impact of feature importance of this dataset to find essential predictors. The detail is shown in Figure 4

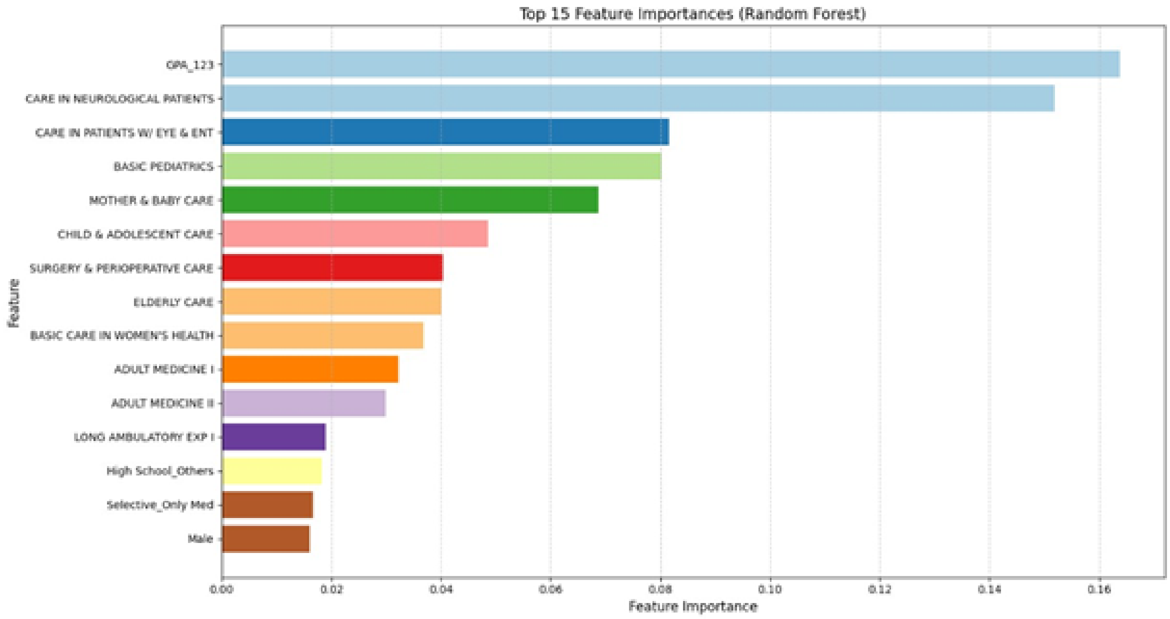

Figure 4. Potential predictors, identified by Feature Importances (Random Forest) method

To identify which variables had the most influence on the pass/fail outcome, a feature importance analysis was performed with Random Forest. Among the 20 features used, the following stood out: 1. Preclinical GPA (Years 1–3): A higher GPA during the first three years strongly correlated with success in the comprehensive exam. 2. Grades in Specific Subjects: Students' performance in Neurology, Ophthalmology-Otolaryngology (Eye & ENT), Pediatrics, and Mother-Baby Care carried significant weight in predicting the final outcome. 3. Selective Choices in the Clinical Year: Selective subjects (especially Internal Medicine) had some effect, though less than GPA or certain clinical courses.

These results suggest that building a strong academic foundation in the early years of medical school is crucial for success in high-stakes final exams. Excelling in certain specialized subjects may further strengthen a student's chance of passing. Program administrators might use this information to offer targeted support in areas that appear most critical.

Discussion of Predictive Model Performance

The findings from this study indicate that machine learning algorithms, particularly SVM, can effectively predict SSC-based curriculum student performance in medical education. The strong performance of SVM in Method 3 (F1-score of 84.8%) is likely due to its robustness in handling imbalanced data, managing non-linear feature interactions, and minimizing overfitting.

Moreover, the variation in failure rates between the old and SSC-based curricula highlights the need for ongoing assessment of the SSC-based curriculum's effectiveness. The higher failure rate, even being not statistically significant, may point to an adjustment period for students adapting to a more self-directed learning environment. It is crucial for future studies to incorporate additional student feedback and academic support mechanisms to ensure that personalized learning approaches do not inadvertently disadvantage certain student groups.

Conclusions

This study first aimed to compare a newly introduced SSC-based curriculum with a traditional curriculum. Secondly, it employed machine learning methods to identify key predictors of medical students' comprehensive exam outcomes. Although the SSC-based group showed a slightly higher failure rate than the traditional cohorts in the curriculum comparison, this difference was not statistically significant, suggesting that students may need additional support when transitioning to a more self-directed learning format.

Among the machine learning techniques tested, Support Vector Machines (SVM) demonstrated the strongest predictive capability, reaching an F1-score of 84.8% and highlighting essential factors such as preclinical GPA and specific subject grades. While the 4 groups of 20 different selective-subject combinations did not emerge as top predictors in this dataset, some of these selective choices may become significant with larger sample sizes or additional cohorts. As the SSC-based curriculum continues to expand, future investigations should focus on gathering more extensive, multi-institutional data to confirm whether these selective pathways consistently influence exam performance.

References

Abdu, E. (2024). Student Performance Prediction Using Machine Learning Algorithms. Applied Computational Intelligence and Soft Computing, 2024. doi:10.1155/2024/4067721

Adam, J., Bore, M., Childs, R., Dunn, J., McKendree, J., Munro, D., & Powis, D. (2015). Predictors of professional behaviour and academic outcomes in a UK medical school: A longitudinal cohort study. Med Teach, 37(9), 868-880. doi:10.3109/0142159X.2015.1009023

Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. 9, 6415-6426. doi:10.12988/ams.2015.53289

Al-Barrak, M., & Al-Razgan, M. (2016). Predicting Students Final GPA Using Decision Trees: A Case Study. International Journal of Information and Education Technology, 6, 528-533. doi:10.7763/IJIET.2016.V6.745

Al Shawwa, L., Abulaban, A. A., Abulaban, A. A., Merdad, A., Baghlaf, S., Algethami, A., . . . Balkhoyor, A. (2015). Factors potentially influencing academic performance among medical students. Adv Med Educ Pract, 6, 65-75. doi:10.2147/AMEP.S69304

Alalawi, K., Athauda, R., & Chiong, R. (2023). 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. doi:https://doi.org/10.1002/eng2.12699

Almarabeh, H. (2017). Analysis of Students' Performance by Using Different Data Mining Classifiers. International Journal of Modern Education and Computer Science, 9, 9-15. doi:10.5815/ijmecs.2017.08.02

Anuradha, C., & Thambusamy, V. (2015). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance. Indian Journal of Science and technology, 8, 974-6846. doi:10.17485/ijst/2015/v8i15/74555

Asif, R., Merceron, A., Abbas, D.-S., & Haider, N. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113. doi:10.1016/j.compedu.2017.05.007

Asif, R., Merceron, A., & Pathan, M. (2014). Predicting Student Academic Performance at Degree Level: A Case Study. International Journal of Intelligent Systems and Applications, 7, 49-61. doi:10.5815/ijisa.2015.01.05

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C., & Nielsen, H. (2000). REVIEW Assessing the Accuracy of Prediction Algorithms for Classification: An Overview. Bioinformatics, 16. doi:10.1093/bioinformatics/16.5.412

Birihanu, E., & Akmel, F. (2017). Student Performance Prediction Model using Machine Learning Approach: The Case of Wolkite University. International Journal of Advanced Research in Computer Science and Software Engineering, 7, 2277. doi:10.23956/ijarcsse/V7I2/01219

Chango, G., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. Computers & Electrical Engineering, 89, 106908. doi:10.1016/j.compeleceng.2020.106908

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. Procedia Computer Science, 136, 16-24. doi:10.1016/j.procs.2018.08.233

Clark, D. J., Kolias, A. G., Garnett, M. R., Trivedi, R. A., Price, S. J., & Hutchinson, P. J. (2016). Student-selected components in neurosurgery. Br J Neurosurg, 30(1), 4-6. doi:10.3109/02688697.2015.1114590

Collins, J. P., White, G. R., & Kennedy, J. A. (1995). Entry to medical school: an audit of traditional selection requirements. Med Educ, 29(1), 22-28. doi:10.1111/j.1365-2923.1995.tb02795.x

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. IEEE Transactions on Learning Technologies, 10(1), 17-29. doi:10.1109/TLT.2016.2616312

Dickman, R. L., Sarnacki, R. E., Schimpfhauser, F. T., & Katz, L. A. (1980). Medical students from natural science and nonscience undergraduate backgrounds. Similar academic performance and residency selection. JAMA, 243(24), 2506-2509.

Dunleavy, D. M., Kroopnick, M. H., Dowd, K. W., Searcy, C. A., & Zhao, X. (2013). The predictive validity of the MCAT exam in relation to academic performance through medical school: a national cohort study of 2001-2004 matriculants. Acad Med, 88(5), 666-671. doi:10.1097/ACM.0b013e3182864299

Falk, G. A., Robb, W. B., Khan, W. H., & Hill, A. D. (2009). Student-selected components in surgery: providing practical experience and increasing student confidence. Ir J Med Sci, 178(3), 267-272. doi:10.1007/s11845-009-0306-8

Farhood, H., Joudah, I., Beheshti, A., & Müller, S. (2024). Evaluating and Enhancing Artificial Intelligence Models for Predicting Student Learning Outcomes. Informatics, 11, 46. doi:10.3390/informatics11030046

Garg, R. (2018). PREDICTING STUDENT PERFORMANCE OF DIFFERENT REGIONS OF PUNJAB USING CLASSIFICATION TECHNIQUES. International Journal of Advanced Research in Computer Science, 9, 236-240. doi:10.26483/ijarcs.v9i1.5234

Hoti, A. H., Zenuni, X., Hamiti, M., & Ajdari, J. (2023, 6-10 June 2023). Student Performance Prediction Using AI and ML: State of the Art. Paper presented at the 2023 12th Mediterranean Conference on Embedded Computing (MECO).

Hussain, S., Muhsen, Z., Salal, Y., Theodorou, P., Kurtoğlu, F., & Hazarika, G. (2019). Prediction Model on Student Performance based on Internal Assessment using Deep Learning. International Journal of Emerging Technologies in Learning (iJET), 14, 4. doi:10.3991/ijet.v14i08.10001

Khudhur, A., & Ramaha, N. T. A. (2023, 8-10 June 2023). Students' Performance Prediction Using Machine Learning Based on Generative Adversarial Network. Paper presented at the 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA).

Kleshinski, J., Khuder, S. A., Shapiro, J. I., & Gold, J. P. (2009). Impact of preadmission variables on USMLE step 1 and step 2 performance. Adv Health Sci Educ Theory Pract, 14(1), 69-78. doi:10.1007/s10459-007-9087-x

Kreiter, C. D., & Kreiter, Y. (2007). A validity generalization perspective on the ability of undergraduate GPA and the medical college admission test to predict important outcomes. Teach Learn Med, 19(2), 95-100. doi:10.1080/10401330701332094

Mastour, H., Dehghani, T., Moradi, E., & Eslami, S. (2023). Early prediction of medical students' performance in high-stakes examinations using machine learning approaches. Heliyon, 9(7), e18248. doi:10.1016/j.heliyon.2023.e18248

Mesaric, J., & Šebalj, D. (2016). Decision trees for predicting the academic success of students. Croatian Operational Research Review, 7, 367-388. doi:10.17535/crorr.2016.0025

Mohamed, M. H., & Waguih, H. (2017). Early Prediction of Student Success Using a Data Mining Classification Technique.

Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. International Journal of Modern Education and Computer Science, 11, 36-42. doi:10.5815/ijmecs.2016.11.05

Ofori, F., Maina, E., & Gitonga, R. (2020). Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review. 2616-3573.

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asis-Lopez, M., Flores-Albornoz, J., & Phasinam, K. (2021). Classification and prediction of student performance data using various machine learning algorithms. Materials Today: Proceedings, 80. doi:10.1016/j.matpr.2021.07.382

Perez, E., Domínguez, J., Omatu, S., Herrera-Viedma, E., & Corchado Rodríguez, J. (2021). Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. Journal of Artificial Intelligence and Soft Computing Research, 12, 79-100. doi:10.2478/jaiscr-2022-0006

Powers, D. (2008). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. Mach. Learn. Technol., 2.

Puddey, I. B., & Mercer, A. (2014). Predicting academic outcomes in an Australian graduate entry medical programme. BMC Med Educ, 14, 31. doi:10.1186/1472-6920-14-31

Putpuek, N., Rojanaprasert, N., Atchariyachanvanich, K., & Thamrongthanyawong, T. (2018). Comparative Study of Prediction Models for Final GPA Score: A Case Study of Rajabhat Rajanagarindra University.

Ramaswami, G. S., Susnjak, T., Mathrani, A., & Umer, R. (2020, 16-18 Dec. 2020). Predicting Students Final Academic Performance using Feature Selection Approaches. Paper presented at the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE).

Riley, S. C. (2009). Student Selected Components (SSCs): AMEE Guide No 46. Med Teach, 31(10), 885-894. doi:10.3109/01421590903261096

Singh, W., & Kaur, P. J. I. J. o. A. R. i. C. S. (2016). Comparative Analysis of Classification Techniques for Predicting Computer Engineering Students' Academic Performance. 7.

Siu, E., & Reiter, H. I. (2009). Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. Adv Health Sci Educ Theory Pract, 14(5), 759-775. doi:10.1007/s10459-009-9160-8

Sivasakthi, M. (2017). Classification and prediction based data mining algorithms to predict students' introductory programming performance.

Sobral, D. T. (2008). Student-selected courses in a medical school: scope and relationships. Med Teach, 30(2), 199-205. doi:10.1080/01421590701754169

Thomas, S., Dhanoa, A., & Palanisamy, U. (2012). Shaping the Future of Medicine: The Effect of 'Selective' Choices on Tomorrow's Doctors. Institutions and Economies.