

Received 31 October 2022, accepted 15 November 2022, date of publication 28 November 2022, date of current version 6 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3225404



Imbalanced Classification Methods for Student Grade Prediction: A Systematic Literature Review

SITI DIANAH ABDUL BUJANG^{1,2}, ALI SELAMAT^{ID1,3,4}, (Member, IEEE), ONDREJ KREJCAR^{1,4},

FARHAN MOHAMED^{ID3}, (Senior Member, IEEE), LIM KOK CHENG⁵, (Member, IEEE),

PO CHAN CHIU^{ID6}, AND HAMIDO FUJITA^{ID4,7,8,9}, (Life Senior Member, IEEE)

¹Malaysia Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia Kuala Lumpur, Kuala Lumpur 54100, Malaysia

²Faculty of Information and Communication Technology, Politeknik Sultan Idris Shah, Sungai Ayer Tawar, Selangor 45100, Malaysia

³School of Computing, Faculty of Engineering, Malaysia & Media and Games Center of Excellence (MagicX), Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

⁴Faculty of Informatics and Management, University of Hradec Kralove, 50003 Hradec Kralove, Czech Republic

⁵College of Computing and Informatics, Universiti Tenaga Nasional Malaysia, Kajang, Selangor 43000, Malaysia

⁶Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak 94300, Malaysia

⁷Faculty of Software and Information Science, Iwate Prefectural University, Takizawa, Iwate 020-0693, Japan

⁸Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18011 Granada, Spain

⁹i-SOMET Incorporated Association, Morioka 020-0104, Japan

Corresponding authors: Ali Selamat (aselamat@utm.my), Siti Dianah Abdul Bujang (sdianah84@gmail.com), and Ondrej Krejcar (Ondrej_krejcar@uhk.cz)

This work was supported in part by the Fundamental Research Grant Scheme (FRGS) through the Ministry of Education under Grant FRGS/1/2022/ICT08/UTM/01/1; and in part by the Faculty of Informatics and Management, University of Hradec Kralove, through the Specific Research Project (SPEV) under Grant 2102/2022.

ABSTRACT Student success is essential for improving the higher education system student outcome. One way to measure student success is by predicting students' performance based on their prior academic grades. Concerning the significance of this area, various predictive models are widely developed and applied to help the institution identify students at risk of failure. However, building a high-accuracy predictive model is challenging due to the dataset's imbalanced nature, which caused biased results. Therefore, this study aims to review the existing research article by providing a state-of-the-art approach for handling imbalanced classification in higher education, including the best practices of dataset characteristics, methods, and comparative analysis of the proposed algorithms, focusing on student grade prediction context problems. The study also presents the most common balancing methods published from 2015 to 2021 and highlights their impact on resolving imbalanced classification in three approaches: data-level, algorithm-level, and hybrid-level. The survey results reveal that the data-level approach using SMOTE oversampling is broadly applied in determining imbalanced problems for student grade prediction. However, the application of hybrid and feature selection methods supporting the generalization of the predictive model to boost student grade prediction performance is generally lacking. Other than that, some of the strengths and weaknesses of the proposed methods are discussed and summarized for the direction of future research. The outcomes of this review will guide the professionals, practitioners, and academic researchers in dealing with imbalanced classification, mainly in the higher education field.

INDEX TERMS Imbalanced classification, prediction model, machine learning, student grade prediction, education, systematic literature review.

I. INTRODUCTION

Student grade prediction is one of the essential areas that can determine and monitor student performance in higher

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

educational institutions (HEI). This area has gained significant attention in the education sector over the years as many studies have been interested and proven the reliability of student grade prediction with many help of the existing machine learning algorithms to enhance student success [1], [2], [3].

The aim is to facilitate the educational sector to evaluate the risk of academic failure and provide feedback to improve student outcomes for each semester. With early grade prediction, the development and progress of students can be assessed more effectively [4], [5]. Other than that, student grade prediction is also one of the common indicators to determine student performance success [6]. The high accuracy of students' grade performances is beneficial and helps the HEI identify the students at risk of failure early in academics. However, as the student dataset becomes more extensive and complex, the effects of imbalance distribution on the target class become higher, which results in poor performance on the predictive model [7]. Therefore, knowing the imbalanced classification methods is significant for building an effective predictive model to improve students' future teaching and learning performance.

There have been various published reviews regarding predicting student academic performances and their relevance, as presented in TABLE 1. Nevertheless, based on the results, few studies highlight the algorithm involved in boosting the predictive model accuracy in predicting student performances. Most of the existing surveys were focused on the machine learning methods and summarized the prominent findings with interesting future directions on student performances, but the review of the algorithm resolving the imbalanced dataset was not discussed comprehensively enough. In this paper, the useful methods to resolve the imbalanced classification to improve the efficiency of the predictive model will be discussed in more detail in three different approaches; data-level, algorithm-level, and hybrid-level.

Therefore, this paper thoroughly reviews and summarizes the most common methods for addressing imbalanced classification in the education domain, focusing on improving the performance of student grade prediction. The contributions of this comprehensive review are summarized and highlighted as follows:

1. This SLR analyzes and summarizes the imbalanced classification methods in detail from three different approaches, data-level, algorithm-level, and hybrid-level, to improve the accuracy of predictive models.

2. Provide a taxonomy of current imbalanced classification methods used for predicting student grades to highlight the most applied algorithms in the education field that will ease the professionals, practitioners, and academic researchers to understand the significance of this technique.

3. A comparative study of existing balancing methods with their classifiers in both aspects (binary and multi-class) and accuracy scores more comprehensively that can be used for future educational research.

4. Provide an overview of the existing evaluation performance metrics applied for an imbalanced classification problem to improve predictive model performances in student grade prediction.

The rest of the paper is organized as follows. Section II gives background information to this research for the reader's basic understanding. Section III describes the review method

and how the SLR was conducted to formulate the selected articles' research questions and search strategy. Section IV provides data extraction and synthesis of SLR results. Section V discusses the results of the overall findings. Section VI discusses the future direction of this research, and finally, the study is summarized and concluded in Section VII.

II. IMBALANCED CLASSIFICATION IN STUDENT GRADE PREDICTION

Data-driven in education is a new trend accelerated by global changes lately. The knowledge and insightful information gained from this area provide many advantages that can improve HEI decision-making. To achieve this, educational datasets are collected from various online databases and platforms such as Course Management and Learning Management Systems (LMS) or known as Moodle, Massive Open Online Courses (MOOC), Open Course Ware (OCW), Open Educational Resources (OER), and social media sites such as Twitter, Facebook, YouTube and Personal Learning Environments (PLE) [10].

Student grade prediction uses machine learning to predict the final score to improve student academic performance by the end of the semester [11]. The aim is to help educators determine the potential students at risk of low results and help them overcome their learning difficulties. Hence, identifying the relevant factors, including student background, academic information, environmental factors, test scores, and Grade Point Average (GPA) or Cumulative Grade Point Average (CGPA), are significant in predicting student performance [6]. However, when a tremendous amount of data is collected and analyzed without being classified in a balanced way, it becomes a significant problem for predicting students' grades.

During the training phases of student grade prediction, imbalanced classification appears when there is an unequal distribution of instances within the target class in the training dataset [12], [13]. Most datasets involve binary classification consisting of two target outputs: the "pass" class as the majority and the "fail" class as a minority. In contrast, some of it comprises more than two different classes, known as multi-class classification. When one class significantly outnumbers the class of the other, the training model usually spends more time processing on the majority classes than the minority ones, which could be less informative. Consequently, it usually leads classifiers to become biased and produce high erroneous. Due to this, many empirical studies are interested in exploring various methods to enhance student grade prediction performance [14], [15], [16], [17]. However, the methods and algorithms used in dealing with various class-imbalanced distributions to predict student grades are not being highlighted and are not comprehensive enough.

Several approaches have been proposed to handle class imbalance to improve the prediction model's performance. These approaches can be categorized into three levels of solutions:

TABLE 1. Comparative analysis of the related existing review article.

Features	Current Review	[8] - 2022	[2] - 2021	[3] - 2021	[9]-2021	[6] - 2020
Type of Review	Systematic Literature Review	Systematic Literature Review	Systematic Literature Review	Systematic Literature Review	Literature Review	Literature Review
Year Covered	2015 – 2021 (7 years)	2017-2021 (5 years)	2009 – 2021 (13 years)	2010-2020 (11 years)	2014-2019 (6 years)	2015-2019 (5 years)
No. of Databases Explored	5	12	6	3	N/A	7
Topic/ Focus Area	Student Grade Prediction, Imbalanced Classification, Machine Learning	Student Retention, Educational Data Mining, Predictive Analytics	Student Performance Prediction, Machine Learning	Machine Learning, Student Performance Prediction	Multi-class Imbalanced, Data Classification	Predicting Academic Success in Higher Education
No. Paper Reviewed	41	100	78	176	6	N/A
Aim	Investigate the imbalanced classification methods to improve the predictive model in the student grade prediction domain.	Provide existing techniques using Learning Analytics to further improve education and the learning environment	Investigate different techniques of data mining and machine learning in predicting student dropout and at risk.	Identify the prominent differences, trends, and problems in these studies and to guide future studies with possible remedies	Identify the existing methods at the level of multi-class imbalance data in the education field	Provide a step-by-step of guidelines for educators to apply data mining techniques to predict student success
Taxonomy	Imbalanced Classification Technique	Machine Learning and Deep Learning Approach	None	None	None	Potential factors impacting the prediction of student academic success
Keyword	Imbalanced Classification, Prediction Model, Machine Learning, Student Grade Prediction, Education, Systematic Literature Review	Educational data mining, Learning analytics, Machine learning, Predictive models, Student retention	Education data mining, Machine learning, MOOC, Student performance, Prediction, Classification	Student Performance Prediction, AI, Machine learning, Deep learning, Education, in-term, end-of-term, Future estimation, Systematic Literature Review	Educational Data, Imbalanced Data, Multi-class Classification	Higher education, Student success, Prediction, Data mining, Review guideline
Research Gap Issues/ Limitation	Refer to Section VII	Data imbalanced issues were reviewed based on previous papers but not discussed on the methods in detail.	It does not include a comparison of the machine learning method; few papers discussed the imbalanced data issues based on the reviewed article.	It does not include detailed methods for resolving imbalanced data; lacks a discussion on the generalization of the proposed model in deep learning.	A limited number of papers were reviewed for handling the multi-class imbalanced problem in education.	Only highlight the influential factors; do not include details of methods that lead to improving the accuracy of the predictive model.

*N/A – denotes as not available information found in the articles based on author's understanding.

- Data-level or pre-processing approach:** This solution is more straightforward, feasible, and easy to implement. The aim is to rebalance the class distribution by either increasing the instances of a minority class or decreasing the instances of a majority class with a sampling method [18], [19]. The data-level approaches include three sampling methods: oversampling, undersampling, and hybrid.
- Algorithm-level approach:** These solutions try not to change the training dataset but to balance the class distribution by modifying the classifier learning [20].
- Hybrid-level approach:** These incorporate data and algorithm-based approaches based on ensemble learning to balance the dataset [21].

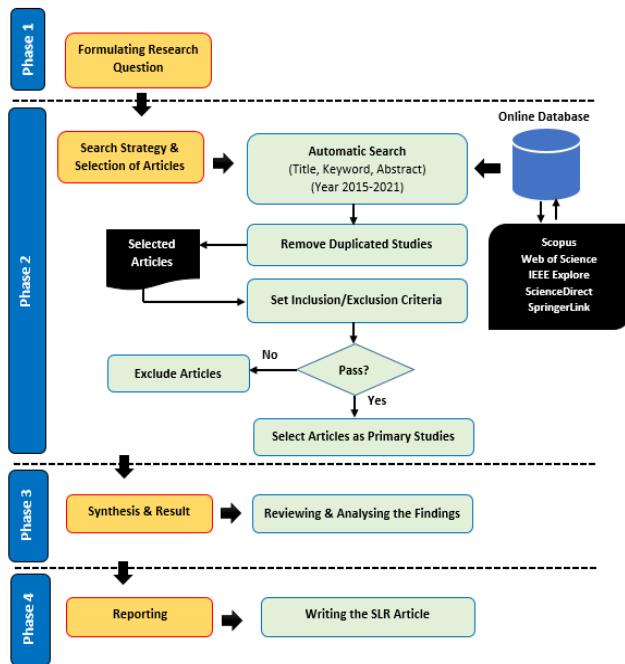
III. SYSTEMATIC REVIEW PROTOCOL

A systematic review is a review of evidence on a formulated question using a systematic method to summarize

research related to a comprehensive study plan [22]. This systematic review was conducted to review and identify current data pre-processing techniques for imbalanced classification problems and find the best solution for machine learning classifiers in student grade prediction. This study applied Kitchenham guidelines [23] to perform the systematic review. The review protocol consists of four phases: formulate research questions, search strategy and selection articles, synthesis and results, and report. Each phase in the review protocol is shown in FIGURE 1.

A. FORMULATING RESEARCH QUESTION (RQ)

There are three RQs have been formulated to explain the exact idea of this SLR, as indicated in TABLE 2. This subsection presents the RQ to address the relevant areas in this study. The aim is to provide an understanding of imbalanced classification methods and discuss how these methods can have a

**FIGURE 1.** Implementation of review protocol.

significant impact in the context of improving student grade prediction performance.

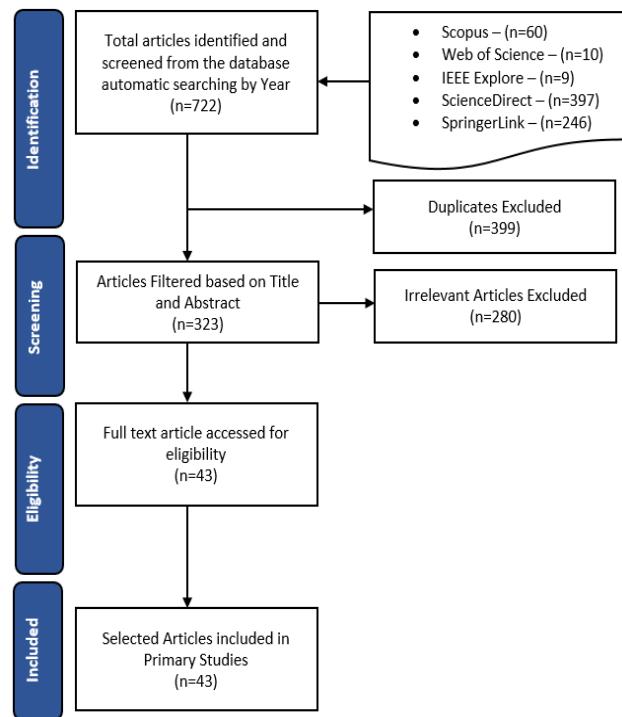
TABLE 2. List of research questions and motivation of related studies.

No.	Research Question	Motivation
RQ1	What are the existing imbalanced classification approaches and methods applied for student grade prediction problems?	Summarize the available state-of-the-art solutions based on data-level, algorithm-level and hybrid-level perspectives.
RQ2	How does the problem of class imbalance affect student grade prediction performance?	It shows the methodological approach used by previous researchers and the need to balance the imbalanced datasets with their impact on accuracy.
RQ3	What metrics are used to appraise the performance of imbalanced classification in student grade prediction?	Identify the metric used to measure the accuracy of the developed prediction model.

B. SEARCH STRATEGY AND SELECTION ARTICLES

This systematic literature review was performed based on the articles extracted from academic journals and conference proceedings from five online citation databases: Scopus, Web of Science, IEEE Explore, ScienceDirect, and SpringerLink. According to the RQs, we formulate the main keywords to search the literature consisting of three terms: (“student grade prediction”), (“student grade prediction”) AND (“imbalanced classification”), (“student grade prediction”) AND (“data pre-processing”). To narrow the search results, we limited our search to those published in peer-reviewed

journals and articles from 2015 to 2021 and selected the subject areas that focus on the computer science field based on the relevant title.

**FIGURE 2.** Selection process based on PRISMA flow diagram.

The search strategy of this review is based on PRISMA [24] flow diagram as shown in FIGURE 2. The process of article selection was conducted using the previously mentioned search keywords. Initially, the prospective articles through automated search generated 722 records based on the five selected databases. Next, all duplicated articles found during the automatic selection were removed using Mendeley software, ending up with 323 records. Then, the search was filtered by reading the article's title and abstract comprehensively through a detailed inclusion and exclusion criteria. After screening through this process, the articles were then classified based on QA criteria to remove all irrelevant articles, ending up with 43 remaining articles for in-depth review.

1) INCLUSION AND EXCLUSION CRITERIA

Specific criteria were set for inclusion and exclusion to determine the relevance of this study's selected journals and articles. This ensures that the selection criteria are reliable and correctly classified based on the defined RQs. The detailed criteria for this review are presented in TABLE 3.

2) QUALITY ASSESSMENT

Quality assessment (QA) is another criterion that we considered in this SLR to ensure that the selected articles are conducted based on specific quality measurements [23], [24]. These criteria were used to investigate the suitability of each

TABLE 3. Inclusion and exclusion criteria.

No.	Inclusion	Exclusion
1	Research papers published between 2015 until 2021	Research papers published less than 2015
2	Peer-reviewed studies from journals and conference papers	Non-peer-reviewed studies and book chapters
3	Publications are written in standard English language	Publications not written in standard English language
4	Research papers have free access	Research papers are not free to access
5	Studies that applied imbalanced methods in student grade prediction at all levels	Studies that have not applied imbalanced methods in student grade prediction

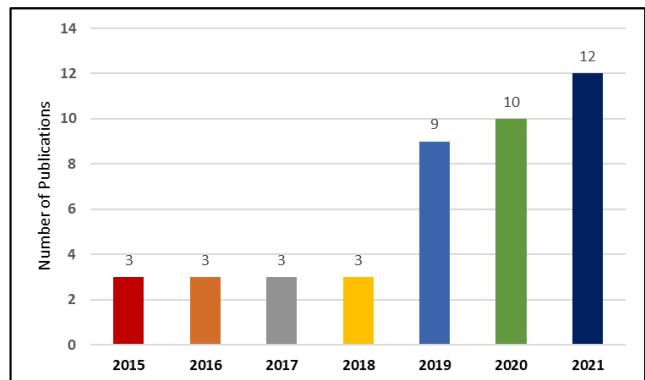
article selected for this SLR. For each QA, there were only three optional answers of whether the paper under review answered the QAs completely (“Yes”), partly (“Partly”) or not at all (“No”) abbreviated as “Y”, ‘N’ and “P” respectively. The five QA criteria carried out in this SLR is shown in TABLE 4.

TABLE 4. List of quality assessment criteria.

No.	Criteria	Y	N	P
QA1	Are the objectives of the research clearly defined?	43	0	0
QA2	Is the proposed technique well-defined to perform the related issues?	42	1	0
QA3	Are the proposed methods for addressing imbalanced classification described in detail?	35	0	8
QA4	Are the strength and weaknesses of the proposed methods mentioned to perform the related issues?	42	1	0
QA5	Are the evaluation metrics used in the research well-defined?	43	0	0

IV. DATA EXTRACTION AND SYNTHESIS

As mentioned in the previous section, this systematic review selected the publication between 2015 and 2021 from five different databases. A total of 43 articles fulfilled all the inclusion and QA used in this SLR. These selected articles were conducted and filtered according to the RQs to be the most relevant that focused on the solution of imbalanced classification and high dimensionality in student grade prediction. In contrast, some other articles only applied conventional data pre-processing for their study findings. FIGURE 3 indicates the trend of selected primary studies published over seven years. Out of 43 articles, the highest number of surveyed studies with 12 articles (27.9%) were published in 2021, followed by 10 articles (23.3%) were released in the year

**FIGURE 3.** Publishing trends for imbalanced classification in higher education.

2020, 9 articles (20.9%) in the year 2019, and 3 articles (7.0%) were in the year between 2015 to 2018 respectively.

The number of systematic reviews shows that an increasing trend in student grade prediction has gained prominence since 2019 among researchers. Moreover, based on our observation, the application of methods related to imbalanced classification problem has gained interest among researchers to improve the predictive model performance in the education field. As listed in TABLE 5, we exploited five research databases to find the relevant articles. Among this, 25 articles (58.1%) were published in journals and 18 articles (41.9%) in conferences proceedings.

TABLE 5. Source of digital library.

Digital library	Link
Scopus	https://www.scopus.com
Web of Science	https://www.webofscience.com
IEEE Explore	https://ieeexplore.ieee.org
Science Direct	https://www.sciencedirect.com
Springer Link	https://link.springer.com

The number of articles from different available databases, including IEEE Explore (30.2%), SpringerLink and Science Direct (23.3%), Scopus (18.6%), and Web of Science (4.6%) is shown in FIGURE 4. The details of each selected article are referred to in TABLE 6 based on year, publication title, number of citations, publication type, publication name, and rankings based on impact factor (IF). Meanwhile, FIGURE 5 visualizes the bibliometric analysis of the citation and publication using VOSviewer. The node size indicates the frequency of occurrence, whereas the curve between the nodes represents their co-occurrence in the same publication.

V. RESULTS AND DISCUSSION

This section presents and discusses the findings by answering the RQs formulated into three subsections; (1) a summary of state-of-the-art approaches and methods in handling imbalanced classification; (2) discusses the proposed methods, algorithms, and datasets which significantly affect the

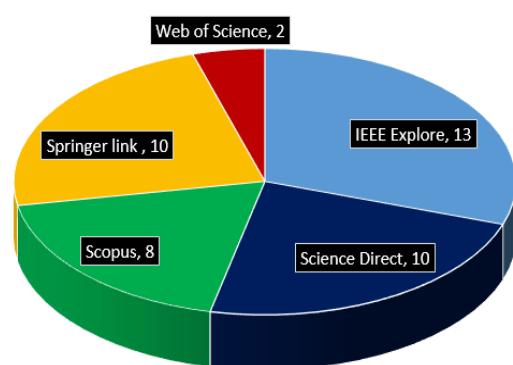
TABLE 6. List of selected studies for the SLR cont.

Year	Pub. Title	Cited	Pub. Type	Pub. Name	IF	Ref
2021	Multi-class Prediction Model for Student Grade Prediction Using Machine Learning	12	Journal	IEEE Access	3.367	[25]
2021	Redefining profit metrics for boosting student retention in higher education	10	Journal	Decision Support Systems	5.795	[26]
2021	Balancing sequential data to predict students at-risk using adversarial networks	10	Journal	Computers and Electrical Engineering	3.818	[27]
2021	Fuzzy-based active learning for predicting student academic performance using autoML: A step-wise approach	9	Journal	Journal of Computing in Higher Education	4.970	[28]
2021	Deep analytic model for student dropout prediction in massive open online courses	18	Journal	Computers & Electrical Engineering	5.795	[29]
2021	Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks	9	Journal	IEEE Access	3.367	[30]
2021	HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students	1	Journal	Education and Information Technologies	2.917	[31]
2021	Predicting Course Grade through Comprehensive Modelling of Students' Learning Behavioural Pattern	5	Journal	Complexity	2.833	[32]
2021	Predicting students academic performance using a hybrid of machine learning algorithms	NA	Conference Paper	IEEE AFRICON Conference	0.500	[33]
2021	Back to Basics: An Interpretable Multi-class Grade Prediction Framework	NA	Journal	Arabian Journal for Science and Engineering	2.334	[34]
2021	Oversampling technique in student performance classification from engineering course	4	Journal	International Journal of Electrical and Computer Engineering	1.616	[35]
2021	Enhancing Prediction of Student Success: Automated Machine Learning Approach	48	Journal	Computers & Electrical Engineering	5.795	[36]
2020	Multi-split optimized bagging ensemble model selection for multi-class educational data mining	33	Journal	Applied Intelligence	5.086	[37]
2020	Graphs Regularized Robust Matrix Factorization and Its Application on Student Grade Prediction	16	Journal	Applied Sciences (Switzerland)	2.679	[11]
2020	An Intelligent Prediction System for Educational Data Mining Based on Ensemble and Filtering approaches	62	Conference Paper	International Conference on Computational Intelligence and Data Science	0.350	[38]
2020	Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining	13	Conference Paper	Journal of Physics: Conference Series	0.547	[39]
2020	Comparing Different Resampling Methods in Predicting Students' Performances Using Machine Learning Techniques	72	Journal	IEEE Access	3.367	[40]
2020	Improving Learning Experience of Students by Early Prediction of Student Performance using Machine Learning	8	Conference Paper	IEEE International Conference for Innovation in Technology	0.920	[41]
2020	Implementation of Data Mining for Drop-Out Prediction using Random Forest Method	6	Conference Paper	8th International Conference on Information and Communication Technology	0.910	[42]
2020	A Decision Support System for Undergraduate Students Admissions using Educational Data Mining	4	Journal	International Conference on Information Technology, Computer, and Electrical Engineering	0.650	[43]
2020	Forward Selection Technique to Choose the Best Features in Prediction of Student Academic Performance Based on Naïve Bayes	4	Conference Paper	Journal of Physics: Conference Series	0.547	[44]
2020	Features Exploration for Grades Prediction using Machine Learning	2	Conference Paper	ACM International Conference Proceeding Series	0.550	[45]
2019	Using machine learning to predict student difficulties from learning session data	146	Journal	Artificial Intelligence Review	8.139	[46]
2019	Prediction of Student's performance by modelling small dataset size	110	Journal	International Journal of Educational Technology in Higher Education	4.944	[47]
2019	Tracking student performance in introductory programming by means of machine learning	39	Conference Paper	MEC International Conference on Big Data and Smart City	1.770	[48]
2019	S PRAR: A novel relational association rule mining classification model applied for academic performance prediction	25	Conference Paper	Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference	NA	[49]
2019	Improving the evaluation process of students' performance utilizing a decision support software	18	Journal	Neural Computing and Applications	5.606	[50]
2019	Dimensionality reduction for predicting student performance in unbalanced data sets	4	Journal	International Journal of Advances in Soft Computing and its Applications	1.780	[51]
2019	Imbalanced Learning Techniques for Improving the Performance of Statistical Models in Automated Essay Scoring	2	Conference Paper	Procedia Computer Science	2.270	[52]
2019	Early Student Grade Prediction: An Empirical Study	11	Conference Paper	International Conference on Advancements in Computational Sciences	0.580	[4]

TABLE 6. (Continued.) List of selected studies for the SLR cont.

2019	An approach to educational data mining model accuracy improvement using histogram discretization and combining classifiers into an ensemble	2	Conference Paper	Smart Innovation, Systems and Technologies	0.700	[53]
2018	Student performance predictor using multi-class support vector classification algorithm	16	Conference Paper	IEEE International Conference on Signal Processing and Communication	1.440	[16]
2018	Are we meeting a deadline? classification goal achievement in time in the presence of imbalanced data	8	Journal	Knowledge-Based Systems Journal	8.038	[54]
2018	Grade prediction of student academic performance with multiple classification models	10	Conference Paper	International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery	0.460	[55]
2017	Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses	339	Journal	Computers in Human Behaviour	6.829	[56]
2017	Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbor	102	Conference Paper	Canadian Conference on Electrical and Computer Engineering	0.470	[57]
2017	Classification and Prediction based Data Mining Algorithms to Predict Students' Introductory programming Performance	37	Conference Paper	International Conference on Inventive Computing and Informatics,	1.110	[58]
2016	Models for early prediction of at-risk students in a course using standards-based grading	285	Journal	Computers and Education	8.538	[59]
2016	Grade prediction with models specific to students and courses	68	Journal	International Journal of Data Science and Analytics	2.520	[60]
2016	Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department	68	Conference Paper	Procedia Computer Science	2.270	[61]
2015	Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector	304	Conference Paper	Procedia Computer Science	2.270	[62]
2015	Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique	152	Journal	Decision Analytics	1.560	[63]
2015	Modeling students' academic performance based on their interactions in an online learning environment	22	Journal	Elementary Education Online	0.364	[64]

*N/A – denotes as not available information found in the articles based on author's understanding.

**FIGURE 4.** Number of articles based on publication database.

performance of student grade prediction and; (3) explore the various evaluation metrics that are used to measure the performance of student grade prediction.

A. SUMMARY STATE-OF-THE-ART FOR IMBALANCED CLASSIFICATION METHOD

Most prevailing methods to address imbalanced classification problems in education, mainly in the student grade prediction domain, are broadly discussed in three approaches: data, algorithm, and hybrid based. In the student grade prediction domain, we found that 17 articles from 43 primary studies selected data-level based on sampling methods more often

to reduce the imbalanced classification, whereas 14 articles considered feature selection to solve the high dimensionality, which helps address imbalanced classification. Meanwhile, the remaining papers use other methods, such as algorithms and hybrid approaches to enhance student grade prediction performance. In answering **RQ1**, the taxonomy of several methods and algorithms proposed by previous researchers for improving student grade prediction performance is summarized in FIGURE 6.

1) DATA-LEVEL OR DATA PRE-PROCESSING APPROACH

Data-level, known as data pre-processing, is one of the most commonly used approaches to address imbalanced problems at the training set level. Three methods utilized under these approaches are oversampling, undersampling and hybrid sampling. Though much research has been devoted to investigating the strength and weaknesses of different selection approaches in the education domain, choosing the most appropriate methods for a given task is often difficult. From the literature, oversampling [30], [54], [55], [56], [63], [65], [66] methods have been the most preferable so far due to their effectiveness in dealing with a high proportion of class imbalance compared to undersampling [38], [39].

a: OVERSAMPLING

A method that increases the size of minority class instances by generating new instances to obtain balanced classes. SMOTE is one of the most popular and classic

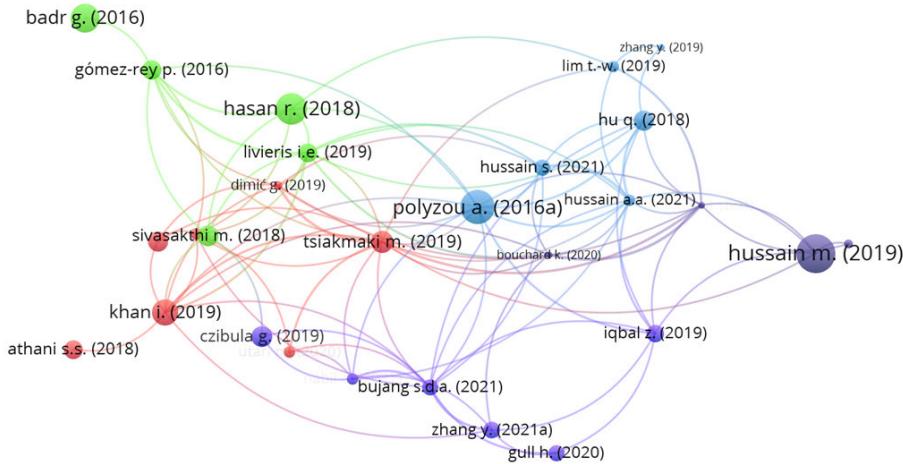


FIGURE 5. Bibliometric analysis of the citation and publications between 2015 to 2021.

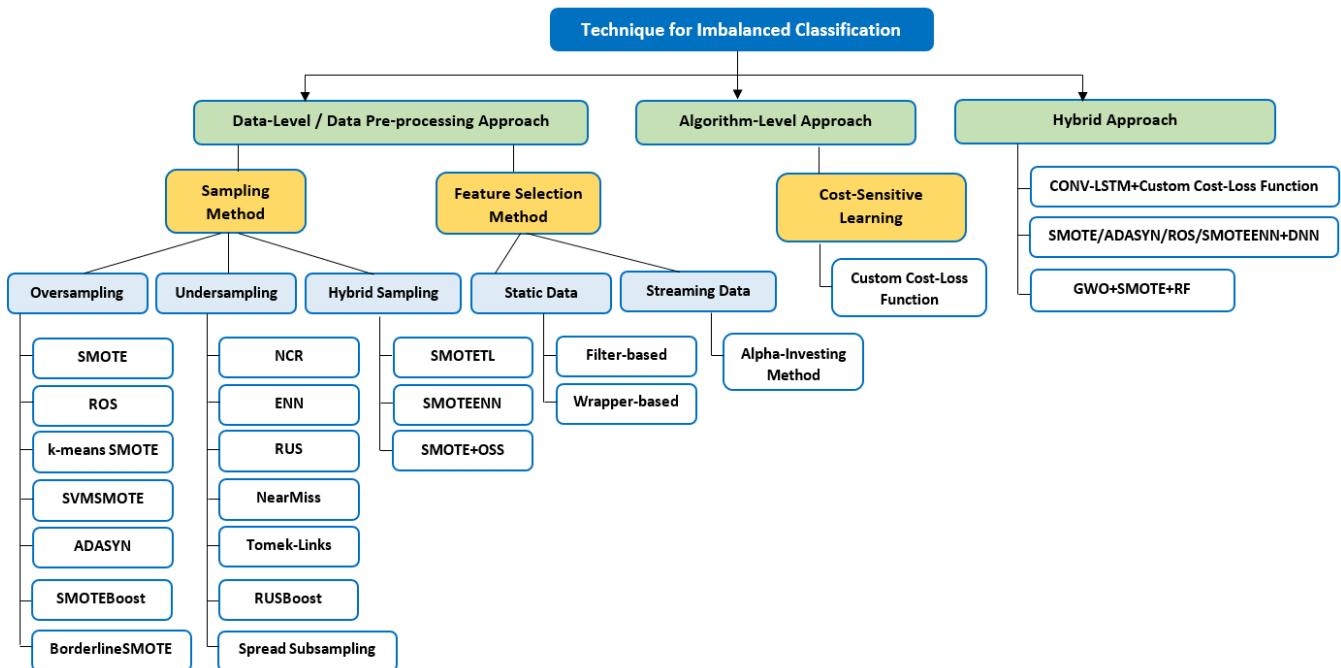


FIGURE 6. Taxonomy of methods proposed in the literature to address imbalanced classification in student grade prediction domain.

oversampling algorithms. Chawla et al. [67] introduced an oversampling SMOTE algorithm that uses interpolation methods to increase the size of the minority class. It oversamples the minority class by generating “synthetic” instances rather than oversampling by replacement. Based on the similar principle of SMOTE, [35] compared many oversampling in 463,956 student data records using SMOTE, SVMMSMOTE and ADASYN to solve imbalance class for predicting students’ performance. Utari et al. [42] developed a high-performance model to predict student dropout using 2492 datasets with balancing methods using SMOTE and RF algorithms. Meanwhile, Bouchard et al. [45] used SMOTE

on a high dimensional imbalance dataset with feature selection to improve student grade prediction. Tanha et al. [21] presented another oversampling algorithm SMOTEBoost to analyze the performance of binary and multi-class imbalanced problems. This algorithm introduces SMOTE in each iteration, boosting learning more minority class instances. However, it takes the highest computational time to process large datasets.

b: UNDERSAMPLING

A method that reduces the size of the majority class instances by eliminating the cases from the training dataset. RUS [39]

is an example of a naïve version and the most straightforward undersampling algorithm. It simply removes a random portion of samples in the majority class to make all classes equally represented. However, randomly discarding the samples may delete some potentially useful information from the majority class instances, reducing the prediction performance [68]. Therefore, to overcome this problem, some researchers have proposed heuristic undersampling algorithms such as ENN [69] and Tomek-Links [70] that effectively avoid the blindness of RUS. Some advanced methods that combine both oversampling and undersampling, known as hybrid sampling, are proposed to achieve reliable results.

c: HYBRID SAMPLING

This method aims to increase the number of minority samples and reduce the number of majority samples to mitigate the sample imbalance. A study [17] used SMOTETL combined SMOTE and TL algorithms to overcome the imbalance problem in predicting student performance. Meanwhile, Pristyanto et al [71] proposed a data hybrid solution using a combination of SMOTE and OSS to handle the distribution of imbalanced classes in predicting student success. The algorithm used SMOTE to reduce the risk of data duplication, whereas OSS overcame the loss of information and misclassification problem in the majority classes. Hassan et al. [39] proposed an algorithm combining SMOTE and undersampling ENN known as SMOTEENN, which can overcome the multi-class imbalanced problem. The experimental results show that SMOTEENN consistently produce high results with ensembles classifiers to improve students' performance.

2) ALGORITHM-LEVEL APPROACH

Unlike the data-level approach, algorithm-level is a dedicated algorithm that directly learns the imbalanced distribution from different datasets based on cost-sensitive methods and ensemble learning [20]. Instead of creating balanced data distributions using different sampling strategies, cost-sensitive learning overcomes the imbalanced learning problem by using different cost matrices that describe the costs for misclassifying any particular data class [72]. It generally learns the imbalance distribution from the classes in the datasets by modifying the decision threshold when assigning the minority class. The improved algorithm has a significant effect on maximizing the classification accuracy of imbalanced datasets [29].

3) HYBRID APPROACH

Other works use hybrid approaches combining both data-level and algorithm-level approaches to optimize the prediction models. Mubarak et al. [29] proposed a custom loss function that introduces different weights for different classes into a new hyper-model CONV-LSTM, so that it selects significant features. It captures temporal dependencies automatically to optimize the prediction results. The proposed

architecture combined the CNN and LSTM deep neural network, which creates complex non-linear features from the interaction of thousands of attributes, and were trained by updating the weights through forward and back propagation to achieve optimal parameters. Then, applying the cost-sensitive to the loss function minimizes the misclassification cost of the imbalanced class, thus increasing the performance of dropout prediction. In another study, Deepika et al. [73] proposed a combination of Optimal feature selection methods using GWO, SMOTE and RF Classifier. They adopted the GWO optimization method to select the optimal parameter for the RF classifier. In contrast, SMOTE solved the imbalanced data in the datasets chosen to enhance the efficiency of student performance prediction.

Few previous works explore the DNN model to improve the class imbalance problem. Nabil et al. [30] compared the efficiency of DNN with SMOTE, ADASYN, ROS and SMOTEENN using 4266 students' previous academic achievement records. When the DNN is implemented to class imbalance dataset, it helps generalization. It enables the network to correctly discover hidden patterns and extract insightful knowledge, thus achieving better and more reliable results. Similarly, Aslam et al. [74] proposed SMOTE into DNN based on a dense model using eight hidden layers to overcome the imbalance problem. TABLE 7 summarizes all approaches, balancing methods and algorithms regarding their strengths and weaknesses in handling imbalanced classification, especially for the student grade prediction domain.

B. IMPACT OF BALANCING METHOD ON STUDENT GRADE PREDICTION PERFORMANCE

The characteristic of an imbalanced dataset has caused the problem of low accuracy for the minority class in classification, which can produce biased results. Due to this, many methods and algorithms have been proposed to solve imbalanced classification among many researchers. To answer **RQ2**, several datasets are commonly used, considering the actual and online data, which give the most significant impact for each study experiment. TABLE 8 summarizes the detailed dataset characteristics with a comparative analysis of the balancing method and algorithm to improve student grade prediction performance.

Utari et al. [42] proposed SMOTE to predict student dropout using 2406 imbalanced datasets. The proposed technique outperformed the accuracy performance of the RF classifier up to 93.4% from 92.3%. Similarly, Gull et al. [41] have applied SMOTE using 250 imbalanced datasets for the multi-class imbalanced problem to predict final student grades early. The proposed techniques show high performance for the LDA with 90.7% accuracy by generating a balanced number of minority class samples. However, SMOTE always leads to overfitting risk when generating and comparing six different algorithms, such as ROS, ADASYN, and duplicating new synthetic instances, which become less effective.

TABLE 7. Summary of strengths and weaknesses on different approaches for handling imbalanced classification.

Approach	Balancing Method	Algorithm	Strengths	Weaknesses	Ref.
Data-Level	Oversampling	SMOTE	<ul style="list-style-type: none"> Rebalance the sample on an imbalanced dataset to reduce the effects of asymmetrical (skewed) class distribution in the learning process. Easy and most popular among researchers. Efficient for low dimensional imbalanced datasets. 	<ul style="list-style-type: none"> The need to apply other classification algorithms in order to optimize the functionality of SMOTE. This leads to overfitting while duplicating and creating new synthetic instances. Sensitive to the data complexity. 	[42], [74], [39], [45], [40]
Data-Level	Oversampling	ADASYN	<ul style="list-style-type: none"> Generate minority class samples that are difficult to learn based on k-nearest neighbors 	<ul style="list-style-type: none"> Less generate minority class samples that are easy to learn. 	[27], [30]
Data-Level	Oversampling	ROS	<ul style="list-style-type: none"> Increase the performance of GB algorithm 	<ul style="list-style-type: none"> It can cause overfitting as it randomly duplicates the minority class 	[30]
Data-Level	Oversampling	BorderlineSMOTE	<ul style="list-style-type: none"> Computes k-nearest neighbors in the minority and the majority class 	<ul style="list-style-type: none"> Fail to discover all minority instances within the decision boundary 	[27]
Data-Level	Oversampling	k-means SMOTE	<ul style="list-style-type: none"> Reduces the noise Minority samples are sparsely distributed based on cluster sparsity using Euclidian. Enhance class regions by creating samples within naturally occurring clusters of the minority class 	<ul style="list-style-type: none"> Ignore small disjuncts in the minority class 	[27]
Data-Level	Oversampling	SVMSMOTE	<ul style="list-style-type: none"> Create new minority samples within the decision boundary based on the support vector obtained from the standard SVM classifier. 	<ul style="list-style-type: none"> The output does not affect by class imbalanced 	[27], [75]
Data-Level	Undersampling	RUS, ENN, NearMiss and Tomek-Links	<ul style="list-style-type: none"> Suitable for vast data 	<ul style="list-style-type: none"> Cause information loss due to data reduction 	[39]
Data-Level	Undersampling	Spread Subsampling	<ul style="list-style-type: none"> Perform effective results with NB classifier by increasing the prediction accuracy 	<ul style="list-style-type: none"> Not considerable for an ensemble approach 	[38]
Data-Level	Hybrid Sampling	SMOTEENN	<ul style="list-style-type: none"> Increase the accuracy performance of the LR and DT model Suitable for education data multi-class classification problem 	<ul style="list-style-type: none"> Not effective for deep artificial neural network 	[39]
Data-Level	Feature Selection (Wrapper-based)	Backward search direction	<ul style="list-style-type: none"> Reduces the dimension of the full datasets by finding the relevant feature subset 	<ul style="list-style-type: none"> Computationally expensive 	[51]
Data-Level	Feature Selection (Streaming-based)	Alpha-investing method	<ul style="list-style-type: none"> Reduces the number of predictor variables by selecting the features sequentially 	<ul style="list-style-type: none"> Consume high computational memory 	[46]
Algorithm-Level	Cost-sensitive Learning	Custom cost loss function	<ul style="list-style-type: none"> Design to reduce different types of misclassification errors by improving the standard loss function on the training dataset 	<ul style="list-style-type: none"> Setting cost-sensitive categories for different classifications are needed by considering class imbalance. 	[29]
Data-Level	Undersampling	RUSBoost	<ul style="list-style-type: none"> Simplicity, more speed, and high performance 	<ul style="list-style-type: none"> Cause to data loss 	[21], [76]
Data-Level	Oversampling	SMOTEBoost	<ul style="list-style-type: none"> Construct an ensemble of diverse classifiers by assigning weights to instances to reduce bias 	<ul style="list-style-type: none"> More complex and time-consuming 	[21], [76]
Hybrid-level	Hybrid	SMOTE+DNN	<ul style="list-style-type: none"> Improved accuracy with a combination of sampling and AI learning algorithm 	<ul style="list-style-type: none"> Need extra work on designing the evaluation to complement each other 	[30]

Some researchers applied RUS, NearMiss, SMOTEENN, and SMOTETL [39] to overcome this problem to improve

the imbalanced classification. A study in [52] proposed four algorithms based on oversampling and undersampling

TABLE 8. Summary of dataset characteristics, balancing methods, algorithm and limitation in related work.

No	Paper	Dataset	Size Sample	Class Label	No. Attribute	Balancing Method	Classifier
1	Intayoad et al. [75]	General education course at Mae Fah Luang University	D1: 1353	{Pass, Fail}	N/A	SMOTE, Borderline-SMOTE1, Borderline-SMOTE2, SVM-SMOTE	NB, DT, kNN
2	Lim et al. [51]	Kaggle	D1: 649 D2: 395	{Pass, Fail} {Pass, Fail}	32	Features approach – wrapper-based	DT, NB, NBT, LibSVM
3	Nabil et al. [30]	Public university	4266	{Pass, Fail}	12	SMOTE, ADASYN, ROS SMOTEENN	DNN, DT, RF, Gradient Boosting, LR, SVM, kNN
4	Saifudin et al. [44]	UCI Math	395	{Pass, Fail}	33	Features approach – forward selection algorithm	NB
5	Utari et al. [42]	ABC Faculty at XYZ university	2406	{Graduated, Dropout}	32	SMOTE	RF
6	Mubarak et al. [29]	https://datastage.stanford.edu http://kddcup2015.com	D1: 78,623 D2: 120,542	{Dropout, Non-Dropout}	N/A	Custom Cost-Loss Function, CONV-LSTM	CNN
7	Waheed et al. [27]	OULAD	32,593	{Distinction-Pass, Distinction-Fail, Pass-Fail, Withdrawn-Pass}	N/A	SMOTE, Borderline-SMOTE, ADASYN, k-Means SMOTE, SVM-SMOTE, ROS	SC-GAN, CGAN
8	Gull et al. [41]	Historical data from undergraduate courses	250	{A,B,C,D,E,F}	5	SMOTE	LR, LDA, KNN, CART, Gaussian NB, SVM
9	Jishan et al. [63]	Numerical Analysis course at North South University, Dhaka, Bangladesh	181	{A,B,C,D,F}	6	SMOTE	DT, NB, NN
10	Bouchard et al. [45]	Quebec's primary schools and high schools	63040	{A,B,C,D,E}	270	SMOTE	1-NN, 5-NN, DT, CART, NB, BayesNet, SVM, LR, MLP, Adaboost, Bagging, RF
11	Ayienda et al. [33]	Kaggle	649	{Fair, Good, Excellent}	33	Quantile Based Discretization, Filter-Based, Feature Selection,	SVM, MLP, LR, kNN, NB, Weighted Voting Classifier
12	Zhang et al. [55]	University of electronic science and technology students	538	{Excellent, Good, Medium, Average, Poor}	N/A	SMOTE, RF Feature Selection	NB, SVM, DT, MLP
13	Pristyanto et al. [71]	Results from Data Structure course at University of Indonesia	105	{successful, not successful}	10	SMOTE + OSS	kNN, SVM, NB
14	Deepika et al. [73]	UCI Math and Portuguese dataset	382	{Pass, Fail}	33	SMOTE, GWO Feature Selection, RF	kNN, ANN
15	Khan et al. [48]	Buraimi University College in Oman	50	{High, Medium, Low}	50	SMOTE	DT (J48), NB, SMO, MLP, IBK, Dec.Table, Jrip, OneR, ZeroR, RF, RTree
16	Ghorbani R. and Ghousi R. [40]	Iran University of Science and Technology Portugal Secondary Education	D1: 650 D2: 394	{Poor, Medium, Good, Excellent}	19	Borderline-SMOTE, ROS, SMOTE, SMOTEENN, SVM-SMOTE, SMOTE-TL	RF, kNN, ANN, XGBoost, SVM, DT, LR, NB
17	Ashraf et al. [38]	University of Kashmir	N/A	{Division I, Division II, Division III}	23	SMOTE. Spread subsampling	RF, J48, kNN, NB, Boosting

*N/A – denotes as not available information found in the articles based on author's understanding

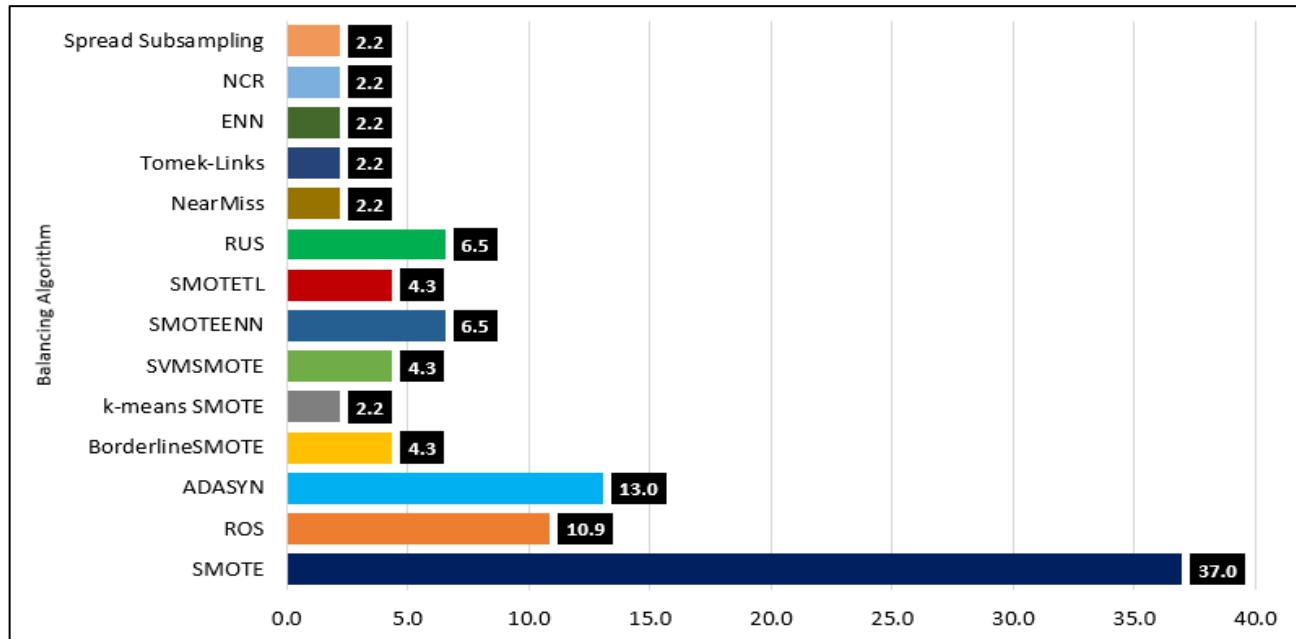


FIGURE 7. Frequency of balancing algorithms based on sampling applied in student grade prediction.

techniques to minimize the effects of imbalanced problems in predicting essay grading. The distribution of each class is based on a stratified sampling approach so that the training dataset is maintained in the same ratios. It was observed that SMOTE and ADASYN were not effectively performed as ROS and RUS due to unusual patterns found in the spatial distribution of feature vectors extracted from textual data. On the other hand, [30] analyzed sampling methods using SMOTE, ROS, and ADASYN and hybrid sampling SMOTEENN algorithms to handle the highly binary imbalanced dataset. The studies used 4266 records of anonymized students with 12 features to predict students' grades based on previous course results. The imbalanced class distributions show that 91.16% and 8.84% were labeled as pass and fail, respectively.

Generally, after the sampling techniques, model classifiers' performance improves and achieves better results with a balanced dataset. However, ROS produced consistent results for all classifiers, including DNN, DT, LR, SVC, KNN, RF and GB. Intayoad et al. [75] employed three combinations of SMOTE, including Borderline SMOTE1, Borderline SMOTE2, and SVMSMOTE, to enhance the accuracy of the minority class samples using two educational datasets. Their results revealed that the proposed algorithms improve the precision, recall, and F1-score of minority class for KNN and NB compared to DT classifiers. However, SMOTE is not influenced by the imbalance ratio to affect classification performance.

On the other hand, we also considered methods to address the issues of noise in data, including outliers and bias for the imbalanced dataset. Jishan et al. [63] applied the

discretization technique using optimal equal-width binning and SMOTE to predict students' final grades with 181 instances. Results demonstrate high accuracy for NN and NB classifiers with 75% compared to 61% after the proposed algorithms were performed.

FIGURE 7 shows the percentages for various balancing algorithms frequently used in the student grade prediction domain. In total, there are fourteen (14) sampling methods have been extracted from this SLR. It is observed that most practices have used SMOTE oversampling (37.0%), then followed by ADASYN (13.0%) and ROS (10.9%). Meanwhile, RUS undersampling and hybrid sampling, SMOTEEN shared an almost equivalent distribution with 6.5%, followed by SMOTETL, SVMSMOTE and BorderlineSMOTE contributed with the same percentage of 4.3%. However, from the findings, we found very few researchers explored methods from different undersampling algorithms (NearMiss, Tomek-Links, ENN, NCR, and Spread Subsampling) and K-means SMOTE oversampling with only 2.2%. These findings indicate a valuable research gap that other researchers can further explore for the important contribution to the performance of predictive models in the future educational domain.

Another important issue that may adversely affect an imbalanced class distribution is the high data dimensionality. Lim et al. [51] investigated the effectiveness of feature selection using the Wrapper approach to reduce the number of 32 features for the imbalanced datasets. The results were validated using four popular classification algorithms: DT, NB, NBT, and LibSVM. The results found that using relevant features to assist C4.5 produced better prediction accuracy

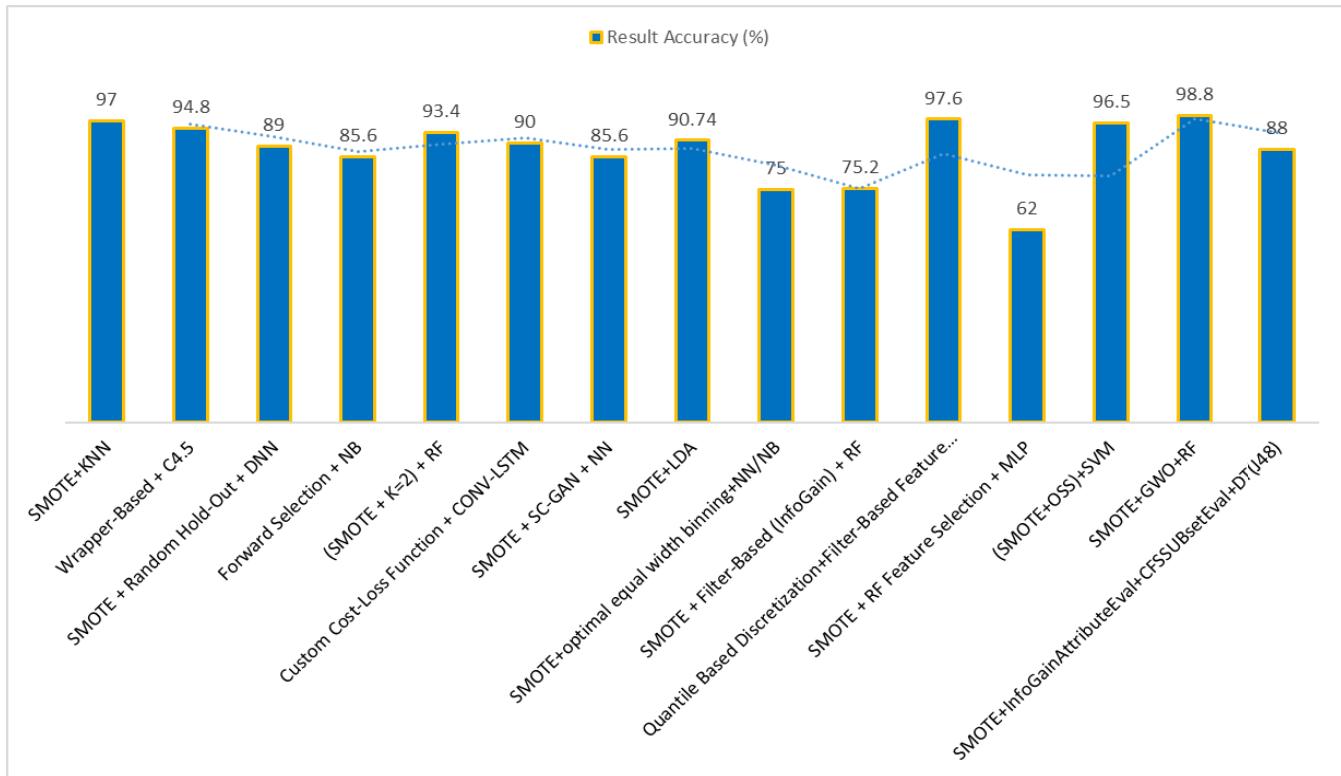


FIGURE 8. Comparison of best accuracy using various algorithms for handling imbalanced classification based on selected studies.

than others. Besides, selecting the relevant features helps to upgrade the generalization abilities of the classifiers and reduce computational time and resources. Hussain et al. [46] proposed a feature selection algorithm based on a streaming model using the Alpha-investing method to predict students' difficulties in the learning session. Other than that, Saifudin et al. [44] identified that using a forward selection algorithm for feature selection reduces the computational complexity of the imbalanced dataset and improves the NB model's accuracy.

In contrast, Khan et al. [48] proposed a correlation-based filter approach to reduce the number of overlapping features and overfit the training dataset to improve the accuracy performance in student grade prediction. Besides the sampling method, Pristyanto et al. [71] proposed hybrid sampling known as SMOTE + OSS that combined SMOTE and OSS methods using 105 instances to improve the prediction performance on the binary imbalanced problem. In their study, the instances of minority classes were generated randomly based on k-nearest neighbors using SMOTE to reduce the risk of duplication. In contrast, the majority of class instances were selected using OSS to remove noise and borderline samples. The results concluded that SMOTE + OSS is more effective with a high average g-mean increase of up to 96.5% using SVM compared to KNN (89.4%) and NB 85.4%) classifiers. Other methods in [73] proposed a hybrid model based on the GWO algorithm, SMOTE oversampling and RF classifier to improve student academic performance. Their experiment

showed that the proposed model could solve the imbalance problem in the selected datasets with high accuracy. Meanwhile, Mubarak et al. [29] noticed that imbalanced classes could cause poor performance results on dropout prediction models. Therefore, the author proposed a novel cost loss function algorithm to address the imbalanced problem with CONV-LSTM using deep learning to optimize the student dropout prediction. The proposed model also applied automatic feature extraction to extract the important features of raw clickstream data obtained from MOOC, showing better results than those who applied manual feature engineering. FIGURE 8 shows the best accuracy percentage using various algorithms proposed from previous studies for handling imbalanced classification to improve student grade prediction. The related articles are summarized in TABLE 9.

C. EVALUATION PERFORMANCE METRICS USED

In order to ensure that the model proposed is suitable for handling imbalanced classification problems, this subsection identifies the most often used performance metrics which respond to **RQ3**. As depicted in FIGURE 9, fourteen (14) frequently used metrics for evaluating the performance of classifiers based on accuracy, f1-measure, precision, recall, AUC, ROC, RMSE, False Positive Rate (FPR), True Positive Rate (TPR), Specificity, Kappa, MAE, incorrectly classified and correctly classified were proposed.

From the findings, 72.2% of the selected articles most considered accuracy, f1-measure, precision, and recall with

TABLE 9. Comparison of accuracy with methods and algorithms proposed for addressing imbalanced classification.

No	Ref	Year	Class Type	Balancing Method	Best Algorithm	Accuracy (%)	Limitation
1	[75]	2018	binary	Oversampling	SMOTE+ kNN	97.0	The imbalance ratio does not affect the performance of SMOTE.
2	[51]	2019	binary	Feature selection	Wrapper-Based + DT	94.8	Limited use of attributes.
3	[30]	2021	binary	Hybrid Approach	SMOTE + Random Hold-Out + DNN	89.0	Feature selection is not considered.
4	[44]	2019	binary	Feature Selection	Forward Selection + NB	85.6	Small dataset and limited to one classifier. Not generalizable.
5	[42]	2020	binary	Oversampling	(SMOTE + K=2) + RF	93.4	Limited use of sampling methods.
6	[29]	2021	binary	Hybrid Approach	Cost-Loss Function + CONV-LSTM	90.0	Lacking generalization of the model.
7	[27]	2021	multi-class	Hybrid Approach	SMOTE + SC-GAN + NN	85.6	Difficult to identify the true dependency.
8	[41]	2020	multi-class	Oversampling	SMOTE +LDA	90.7	The limited size of the dataset. Limited use of attributes. Not generalizable.
9	[63]	2015	multi-class	Oversampling	SMOTE + optimal equal width binning+ NN/NB	75.0	Small dataset and limited attributes.
10	[45]	2020	multi-class	Oversampling + Feature Selection	SMOTE + Filter-Based (InfoGain) + RF	75.2	SMOTE not efficient on high dimensional imbalanced dataset. Use manual feature selection.
11	[33]	2021	multi-class	Hybrid Approach	Quantile Based Discretization + Filter-Based Feature Selection + Hybrid WVC (10 CV) + (SVM, NB, MLP, KNN, LR)	97.6	Not use an imbalanced method to address imbalanced data problems.
12	[55]	2018	multi-class	Oversampling + Feature Selection	SMOTE + RF Feature Selection + MLP	62.0	Limited sample size.
13	[71]	2017	binary	Hybrid sampling	(SMOTE + OSS) + SVM	96.5	Small dataset and only considered for the binary class.
14	[73]	2020	binary	Hybrid Approach	SMOTE + GWO + RF	98.8	Not resolving data sparsity issues.
15	[48]	2019	multi-class	Oversampling + Feature Selection	SMOTE + InfoGainAttributeEval + CFSSUBsetEval + DT(J48)	88.0	Small dataset.
16	[40]	2020	multi-class	Oversampling	SVM-SMOTE + RF + 5 fold CV	81.3	Feature selection is not considered.
17	[38]	2020	multi-class	Oversampling + Ensemble	SMOTE + NB with Boosting	96.1	Feature selection is not considered.

the rates of 23% (29/126), 16.7% (21/126), 16.7% (21/126) and 15.9% (20/126) respectively to evaluate the imbalanced classification performance in student grade prediction. Accuracy is the most intuitive performance measure defined as the ratio of the total number of correctly calculated predictions. In contrast, precision is the ratio of the correctly predicted positive to the total number of correctly predicted positive cases. Recall, also known as sensitivity, is the ratio of correctly classified cases to the total number of all actual class cases. Meanwhile, the f-measure or f1-score is the weighted

average of precision and recall, which are considered a good indicator of the relationship between them. Usually, most previous researchers found the f-measure score is more valuable than accuracy for imbalanced classifications because it considers both false positives and false negatives rates. For example, studies in [30], [33], [48], and [63] have compared the score of accuracy, f1-measure, precision, and recall to evaluate the effectiveness of the predicted models. On the other hand, only 27.8% of articles reflect their performance evaluation based on the remain parameters. This includes

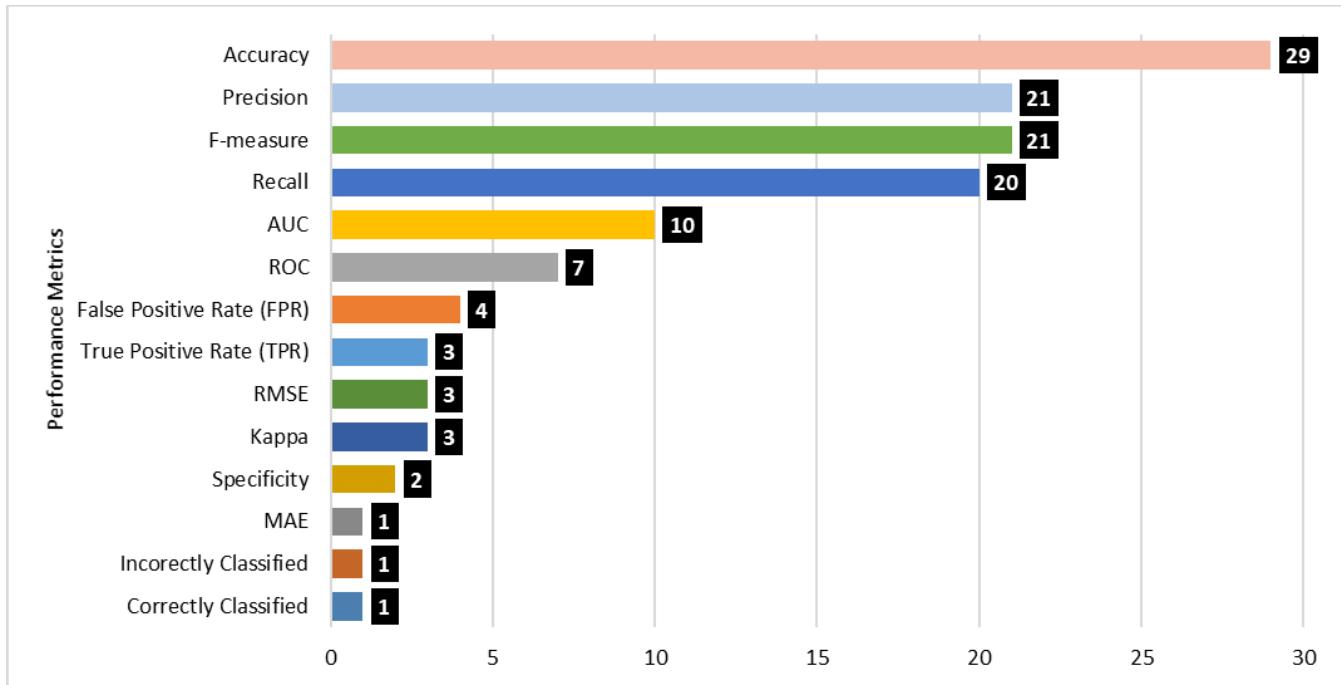


FIGURE 9. Parameter of the performance metric.

AUC at 7.9% (10/126), ROC at 5.6% (7/126), and FPR at 3.2% (4/126), whereas TPR, Kappa, and RMSE contributed the same distribution at 2.4% (3/126). Apart from them, the correctly and incorrectly classified MAE was one of the least used metrics in the student grade prediction, with a score of 0.8% (1/126). The RMSE and MAE are metrics used to measure the regression model error.

Studies from [4] and [11] used the RMSE and MAE to evaluate the matrix factorization techniques for predicting students' grades. Meanwhile, [45] applied accuracy, f1-measure, Kappa, and ROC to evaluate different classifiers using 270 features with imbalanced classes to predict students' final grades. Kappa is another metric that compares the observed and expected accuracy, which is a good measure for handling multi-class imbalanced problems. A higher Kappa score closer to score 1 indicates a more accurate prediction model. Hence, the diversity of different performance metrics shows a significant impact in evaluating the model's capabilities from different perspectives when handling binary and multi-class imbalanced problems. TABLE 10 lists the specific articles that have utilized the performance metrics for this studies.

VI. FUTURE DIRECTIONS OF STUDENT GRADE PREDICTION

This study presents an overview of imbalanced classification methods used in student grade prediction focusing on two aspects; (a) appropriate solutions to build high-accuracy prediction models; (b) appropriate algorithms and metrics to appraise the performance of imbalanced datasets. Based

on this review, there are some limitations, and underexplored methods have been shown in resolving the imbalanced datasets to build high-accuracy prediction models. Some aspects that are found to be potential and require further discussion for the future direction in this field are as follows:

A. SAMPLING BASED WITH FEATURE SELECTION METHOD FOR HIGH DIMENSIONAL IMBALANCED DATASET

Dealing with imbalanced classification on a high-dimensional dataset is very difficult to train and will negatively impact the performance of the predictive results. It is of utmost importance to overcome the high dimensionality of the data because it can produce the probability of data becoming sparse and highly imbalanced leading to inaccurate results. In this study, SMOTE is less effective for high-dimensional imbalanced dataset. Adopting feature selection can reduce the biased of minority classes to improve the predictive result. Therefore, building comprehensive data pre-processing is required by considering a hybrid feature selection method that can be integrated with different sampling methods to solve the problem of imbalanced classification effectively.

B. HYBRID-LEVEL APPROACH FOR IMBALANCED MULTI-CLASS CLASSIFICATION

Selecting appropriate hybrid methods and basic classifiers can increase the effectiveness of these algorithms, which can contribute to a better quality of educational data. Referring to FIGURE 8, we noticed that the performance of

TABLE 10. Articles related to performance metrics parameter used.

Performance Metrics	Reference
Accuracy	[11], [16], [25], [27], [28], [30], [31], [33], [42], [43], [45], [46], [47], [48], [51]–[53], [56], [58], [59], [63]
AUC	[26]–[29], [33], [44], [49], [54], [63], [64]
Correctly Classified	[38]
f-measure	[25], [27], [29]–[33], [37], [38], [41], [39], [42], [45], [46], [48], [49], [55], [56], [62], [63]
False Positive Rate (FPR)	[37], [38], [49], [53]
Incorrectly Classified	[38]
Kappa	[41], [45], [47]
MAE	[77]
Precision	[25], [27], [29], [31]–[33], [37], [38], [39], [41]–[43], [46], [48], [49], [53], [56], [58], [62], [63]
Recall	[25], [27], [29], [33], [38], [42], [46], [48], [56], [63], [47]
RMSE	[11], [60], [77]
ROC	[26], [38], [45], [54], [63]
Specificity	[47], [49]
True Positive Rate (TPR)	[38], [49], [78]

the hybrid approach achieved the highest accuracy in handling the imbalanced classification problem for student grade prediction. Furthermore, some methods discussed and proposed dealing with the binary imbalanced dataset cases, but the research on the hybrid approach of multi-class classification is still lacking. In future work, it is recommended to highlight the practice of using another aspect of hybrid approaches to resolving the multi-class problem in this domain to boost the accuracy of the predictive model.

C. IMPROVED ACCURACY OF IMBALANCED DATASET USING ENSEMBLE METHOD

The ensemble method is designed to increase the accuracy by not changing the base classifier to make a decision output of a single class. Based on the observation, we revealed that most of the studies used common traditional classifiers such as DT, SVM, kNN, NB, RF, etc., and only a few have investigated the potential of ensemble algorithms using ensemble algorithms such as Bagging, Stacking, and Boosting method for improving the performance of student grade prediction. It is more appropriate to highlight the performance of ensemble algorithms when dealing with imbalanced classes than the mere use of a primary classifier in predicting the student's grade prediction.

VII. CONCLUSION

This paper presents the survey of approaches and methods used to address the imbalanced class problem in the student

grade prediction domain, including the state-of-the-art methods, solutions, impacts of the methods, and future directions for solving imbalanced classification problems. Overall, this review achieved its objectives of enhancing student grade prediction performance by highlighting the impact of imbalanced classification as the key solution to improve student grade prediction. Various approaches and methods proposed to overcome the difficulties found in imbalanced student grade prediction can be grouped into sampling methods, feature selection, cost-sensitive learning, and hybrid approaches. Nevertheless, this approach has its limitations. Still, it has many aspects that need to be improved regarding the diversity of methods used, although some have shown exemplary performance in handling imbalanced classes. Hence, it can provide new advances in developing a more efficient model focus for improving data quality, especially in the education domain. We will expand our investigation for future plans by considering the new techniques and methods in other related data-driven issues.

APPENDIX

Acronym	Full Form
ADASYN	Adaptive Synthetic Sampling.
AUC	Area Under the Receiver Operating.
CGAN	Characteristics Curve.
CNN	Conditional Generative Adversarial Networks.
CONV-LSTM	Convolutional Neural Networks and Long Short-Term Memory.

ANN	Artificial Neural Network.
DT	Decision Tree.
EDM	Educational Data Mining.
ENN	Extended Nearest Neighbor.
FMARR	Fuzzy Mean Aggregation Reasoning Rule.
FPC GA/LGA	Fuzzy Pattern Classifier Genetic Algorithm.
FPC OWA	Fuzzy Pattern Classifier Ordered-Weighted Averaging.
GB	Gradient Boosting.
GWO	Grey Wolf Optimization.
HELA	Hybrid Ensemble Learning Algorithm.
KNN	K- Nearest Neighbor.
LA	Learning Analytics.
LDA	Linear Discrimination Analysis.
LibSVM	Library for Support Vector Machine.
LR	Logistic Regression.
LSTM	Long Short-Term Memory.
MFPC AIWA/OWA	Modified Fuzzy Pattern Classifier Andnessdirected Importance Weighting Averaging.
NB	Naïve Bayes.
NBT	Naïve Bayes Tree.
NN	Neural Network.
NCR	Neighbor Cleaning Rule.
OLUAD	Open University Learning Analytics Dataset.
OSS	One Sided Selection.
PT	Pattern Trees.
PTTD	Top-Down Pattern Trees.
RF	Random Forest.
ROC	Receiver Operating Characteristics Curve.
ROS	Random Oversampling.
RUS	Random Undersampling.
SFS	Streaming Feature Selection.
SC-GAN	Sequential-Conditional Generative Adversarial Networks.
SMOTE	Synthetic Minority Oversampling Technique.
SMOTEENN	Synthetic Minority Oversampling Technique + Extended Nearest Neighbor.
SMOTETL	Synthetic Minority Oversampling Technique + Tomek-Links.
SVC	Support Vector Classifier.
SVMSMOTE	Support Vector Machine + Synthetic Minority. Oversampling Technique.
TL	University of California.
UCI	
WVC	Weighted Voting Classifier

ACKNOWLEDGMENT

The authors are also grateful for the support of student Michal Dobrovolny in consultations regarding application aspects.

REFERENCES

- [1] N. Mduma, K. Kalegele, and D. Machuve, "A survey of machine learning approaches and techniques for student dropout prediction," *Data Sci. J.*, vol. 18, pp. 1–10, Apr. 2019, doi: [10.5334/dsj-2019-014](https://doi.org/10.5334/dsj-2019-014).
- [2] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, p. 552, 2021, doi: [10.3390/educsci11090552](https://doi.org/10.3390/educsci11090552).
- [3] B. Sekeroglu, R. Abiyev, A. Ilhan, M. Arslan, and J. B. Idoko, "Systematic literature review on machine learning and student performance prediction: Critical gaps and possible remedies," *Appl. Sci.*, vol. 11, no. 22, p. 10907, Nov. 2021.
- [4] Z. Iqbal, A. Qayyum, S. Latif, and J. Qadir, "Early student grade prediction: An empirical study," in *Proc. 2nd Int. Conf. Advancements Comput. Sci. (ICACS)*, Feb. 2019, pp. 1–7, doi: [10.23919/ICACS.2019.8689136](https://doi.org/10.23919/ICACS.2019.8689136).
- [5] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran, "Machine learning based student grade prediction: A case study," 2017, *arXiv:1708.08744*.
- [6] E. Alyahyan and D. Düstegör, "Predicting academic success in higher education: Literature review and best practices," *Int. J. Educ. Technol. Higher Educ.*, vol. 17, no. 1, pp. 1–21, Dec. 2020, doi: [10.1186/s41239-020-0177-7](https://doi.org/10.1186/s41239-020-0177-7).
- [7] S. H. Ebenuwa, M. S. Sharif, M. Alazab, and A. Al-Nemrat, "Variance ranking attributes selection techniques for binary classification problem in imbalance data," *IEEE Access*, vol. 7, pp. 24649–24666, 2019, doi: [10.1109/ACCESS.2019.2899578](https://doi.org/10.1109/ACCESS.2019.2899578).
- [8] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student retention using educational data mining and predictive analytics: A systematic literature review," *IEEE Access*, vol. 10, pp. 72480–72503, 2022, doi: [10.1109/ACCESS.2022.3188767](https://doi.org/10.1109/ACCESS.2022.3188767).
- [9] R. M. Mathew and D. R. Gunasundari, "A review on handling multi-class imbalanced data classification in education domain," in *Proc. Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, 2021, pp. 752–755.
- [10] K. L.-M. Ang, F. L. Ge, and K. P. Seng, "Big educational data & analytics: Survey, architecture and challenges," *IEEE Access*, vol. 8, pp. 116392–116414, 2020, doi: [10.1109/ACCESS.2020.2994561](https://doi.org/10.1109/ACCESS.2020.2994561).
- [11] X. S. Y. Zhang, Y. Yun, H. Dai, and J. Cui, "Graphs regularized robust matrix factorization and its application on student grade prediction," *Appl. Sci.*, vol. 10, no. 5, p. 1755, 2020, doi: [10.3390/app10051755](https://doi.org/10.3390/app10051755).
- [12] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of imbalanced data: Review of methods and applications," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, 2021, Art. no. 012077, doi: [10.1088/1757-899x/1099/1/012077](https://doi.org/10.1088/1757-899x/1099/1/012077).
- [13] Y. Sun, A. K. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 4, pp. 687–719, 2009, doi: [10.1142/S0218001409007326](https://doi.org/10.1142/S0218001409007326).
- [14] A. Al-Ashoor and S. Abdullah, "Examining techniques to solving imbalanced datasets in educational data mining systems," *Int. J. Comput.*, vol. 21, no. 2, pp. 205–213, Jun. 2022, doi: [10.47839/ijc.21.2.2589](https://doi.org/10.47839/ijc.21.2.2589).
- [15] D. Thammasiri, D. Delen, P. Meesad, and N. Kasap, "A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition," *Exp. Syst. Appl.*, vol. 41, no. 2, pp. 321–330, Feb. 2014, doi: [10.1016/j.eswa.2013.07.046](https://doi.org/10.1016/j.eswa.2013.07.046).
- [16] S. S. Athani, S. A. Kodli, M. N. Banavasi, and P. G. S. Hiremath, "Student performance predictor using multiclass support vector classification algorithm," in *Proc. Int. Conf. Signal Process. Commun. (ICSPC)*, Jul. 2017, pp. 341–346, doi: [10.1109/CSPC.2017.8305866](https://doi.org/10.1109/CSPC.2017.8305866).
- [17] I. Pratama, Y. Pristyanto, and P. T. Prasetyaningrum, "Imbalanced class handling and classification on educational dataset," in *Proc. 4th Int. Conf. Inf. Commun. Technol. (ICOIACT)*, Aug. 2021, pp. 180–185, doi: [10.1109/ICOIACT53268.2021.9563968](https://doi.org/10.1109/ICOIACT53268.2021.9563968).
- [18] A. Chaplot, N. Choudhary, and K. Jain, "A review on data level approaches for managing imbalanced classification problem," *Int. J. Sci. Res. Sci., Eng. Technol.*, vol. 6, no. 2, pp. 91–97, Mar. 2019, doi: [10.32628/ijsrset196225](https://doi.org/10.32628/ijsrset196225).

- [19] K. Upadhyay and P. Kaur, "A review on data level approaches to address the class imbalance problem," in *Proc. Int. Conf. Challenges Eng. Sci. Technol.*, Apr. 2021, pp. 152–158.
- [20] A. Fernandez, S. Garcia, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, "Algorithm-level approaches," in *Learning from Imbalanced Data Sets*. Cham, Switzerland: Springer, 2018, pp. 123–146, doi: [10.1007/978-3-319-98074-4_6](https://doi.org/10.1007/978-3-319-98074-4_6).
- [21] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: An experimental review," *J. Big Data*, vol. 7, no. 1, pp. 1–47, Dec. 2020, doi: [10.1186/s40537-020-00349-y](https://doi.org/10.1186/s40537-020-00349-y).
- [22] G. M. Tawfik, K. A. S. Dila, M. Y. F. Mohamed, D. N. H. Tam, N. D. Kien, A. M. Ahmed, and N. T. Huy, "A step by step guide for conducting a systematic review and meta-analysis with simulation data," *Tropical Med. Health*, vol. 47, no. 1, pp. 1–9, Dec. 2019, doi: [10.1186/s41182-019-0165-6](https://doi.org/10.1186/s41182-019-0165-6).
- [23] B. Kitchenham, "Procedures for performing systematic reviews, version 1.0," *Empir. Softw. Eng.*, vol. 33, no. 2004, pp. 1–26, 2004.
- [24] M. J. Page, D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, and R. Chou, "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," *BMJ*, vol. 372, pp. 1–37, Mar. 2021, doi: [10.1136/bmj.n160](https://doi.org/10.1136/bmj.n160).
- [25] S. D. A. Bujang, A. Selamat, R. Ibrahim, O. Krejcar, E. Herrera-Viedma, H. Fujita, and N. A. M. Ghani, "Multiclass prediction model for student grade prediction using machine learning," *IEEE Access*, vol. 9, pp. 95608–95621, 2021, doi: [10.1109/ACCESS.2021.3093563](https://doi.org/10.1109/ACCESS.2021.3093563).
- [26] S. Maldonado, J. Miranda, D. Olaya, J. Vásquez, and W. Verbeke, "Redefining profit metrics for boosting student retention in higher education," *Decis. Support Syst.*, vol. 143, Apr. 2021, Art. no. 113493, doi: [10.1016/j.dss.2021.113493](https://doi.org/10.1016/j.dss.2021.113493).
- [27] H. Waheed, M. Anas, S.-U. Hassan, N. R. Aljohani, S. Alelyani, E. E. Edifor, and R. Nawaz, "Balancing sequential data to predict students at-risk using adversarial networks," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107274, doi: [10.1016/j.compeleceng.2021.107274](https://doi.org/10.1016/j.compeleceng.2021.107274).
- [28] M. Tsiaikmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Fuzzy-based active learning for predicting student academic performance using autoML: A step-wise approach," *J. Comput. Higher Educ.*, vol. 33, no. 3, pp. 635–667, 2021.
- [29] A. A. Mubarak, H. Cao, and I. M. Hezam, "Deep analytic model for student dropout prediction in massive open online courses," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107271, doi: [10.1016/j.compeleceng.2021.107271](https://doi.org/10.1016/j.compeleceng.2021.107271).
- [30] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of students academic performance based on courses grades using deep neural networks," *IEEE Access*, vol. 9, pp. 140731–140746, 2021, doi: [10.1109/ACCESS.2021.3119596](https://doi.org/10.1109/ACCESS.2021.3119596).
- [31] S. B. Keser and S. Aghalarova, "HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students," *Educ. Inf. Technol.*, vol. 27, no. 4, pp. 4521–4552, 2021.
- [32] D. Hooshyar and Y. Yang, "Predicting course grade through comprehensive modelling of students learning behavioral pattern," *Complexity*, vol. 2021, pp. 1–12, May 2021, doi: [10.1155/2021/7463631](https://doi.org/10.1155/2021/7463631).
- [33] R. Ayienda, R. Rimuru, and W. Cheruiyot, "Predicting students academic performance using a hybrid of machine learning algorithms," in *Proc. IEEE AFRICON*, Sep. 2021, pp. 4–9, doi: [10.1109/AFRICON51333.2021.9571012](https://doi.org/10.1109/AFRICON51333.2021.9571012).
- [34] B. Alharbi, "Back to basics: An interpretable multi-class grade prediction framework," *Arabian J. Sci. Eng.*, vol. 47, no. 2, pp. 2171–2186, Feb. 2022, doi: [10.1007/s13369-021-06153-x](https://doi.org/10.1007/s13369-021-06153-x).
- [35] N. Rachburee and W. Punlumjeak, "Oversampling technique in student performance classification from engineering course," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 4, pp. 3567–3574, 2021, doi: [10.11591/ijeee.v11i4.pp3567-3574](https://doi.org/10.11591/ijeee.v11i4.pp3567-3574).
- [36] H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: Automated machine learning approach," *Comput. Electr. Eng.*, vol. 89, Jan. 2021, Art. no. 106903, doi: [10.1016/j.compeleceng.2020.106903](https://doi.org/10.1016/j.compeleceng.2020.106903).
- [37] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Multi-split optimized bagging ensemble model selection for multi-class educational data mining," *Int. J. Speech Technol.*, vol. 50, no. 12, pp. 4506–4528, Dec. 2020, doi: [10.1007/s10489-020-01776-3](https://doi.org/10.1007/s10489-020-01776-3).
- [38] M. Ashraf, M. Zaman, and M. Ahmed, "An intelligent prediction system for educational data mining based on ensemble and filtering approaches," *Proc. Comput. Sci.*, vol. 167, pp. 1471–1483, Jan. 2020, doi: [10.1016/j.procs.2020.03.358](https://doi.org/10.1016/j.procs.2020.03.358).
- [39] H. Hassan, N. B. Ahmad, and S. Anuar, "Improved students performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining," *J. Phys., Conf.*, vol. 1529, no. 5, May 2020, Art. no. 052041, doi: [10.1088/1742-6596/1529/5/052041](https://doi.org/10.1088/1742-6596/1529/5/052041).
- [40] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students performance using machine learning techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020, doi: [10.1109/ACCESS.2020.2986809](https://doi.org/10.1109/ACCESS.2020.2986809).
- [41] H. Gull, M. Saqib, S. Z. Iqbal, and S. Saeed, "Improving learning experience of students by early prediction of student performance using machine learning," in *Proc. IEEE Int. Conf. Innov. Technol. (INOCON)*, Nov. 2020, pp. 20–23, doi: [10.1109/INOCON50539.2020.9298266](https://doi.org/10.1109/INOCON50539.2020.9298266).
- [42] M. Utari, B. Warsito, and R. Kusumaningrum, "Implementation of data mining for drop-out prediction using random forest method," in *Proc. 8th Int. Conf. Inf. Commun. Technol. (ICOICT)*, Jun. 2020, pp. 1–5, doi: [10.1109/ICOICT49345.2020.9166276](https://doi.org/10.1109/ICOICT49345.2020.9166276).
- [43] G. Dewantoro and N. Ardisa, "A decision support system for undergraduate students admissions using educational data mining," in *Proc. 7th Int. Conf. Inf. Technol., Comput., Electr. Eng. (ICITACEE)*, Sep. 2020, pp. 105–109, doi: [10.1109/ICITACEE50144.2020.9239244](https://doi.org/10.1109/ICITACEE50144.2020.9239244).
- [44] A. Saifudin and T. Desyani, "Forward selection technique to choose the best features in prediction of student academic performance based on Naive Bayes," *J. Phys., Conf.*, vol. 1477, no. 2, 2020, Art. no. 032007, doi: [10.1088/1742-6596/1477/3/032007](https://doi.org/10.1088/1742-6596/1477/3/032007).
- [45] K. Bouchard, L. Gonzales, J. Maitre, and S. Gaboury, "Features exploration for grades prediction using machine learning," in *Proc. 6th EAI Int. Conf. Smart Objects Technol. Social*, 2020, pp. 78–83, doi: [10.1145/3411170.3411123](https://doi.org/10.1145/3411170.3411123).
- [46] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381–407, Jun. 2019, doi: [10.1007/s10462-018-9620-8](https://doi.org/10.1007/s10462-018-9620-8).
- [47] L. M. A. Zohair, "Prediction of student's performance by modelling small dataset size," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, pp. 1–18, Dec. 2019, doi: [10.1186/s41239-019-0160-3](https://doi.org/10.1186/s41239-019-0160-3).
- [48] I. Khan, A. Al Sadiri, A. R. Ahmad, and N. Jabeur, "Tracking student performance in introductory programming by means of machine learning," in *Proc. 4th MEC Int. Conf. Big Data Smart City (ICBDSC)*, Jan. 2019, pp. 1–6, doi: [10.1109/ICBDSC.2019.8645608](https://doi.org/10.1109/ICBDSC.2019.8645608).
- [49] G. Czibula, A. Mihailescu, and L. M. Criveti, "S PRAR: A novel relational association rule mining classification model applied for academic performance prediction," *Proc. Comput. Sci.*, vol. 159, pp. 20–29, Jan. 2019, doi: [10.1016/j.procs.2019.09.156](https://doi.org/10.1016/j.procs.2019.09.156).
- [50] I. E. Livieris, T. Kotsilieris, V. Tampakas, and P. Pintelas, "Improving the evaluation process of students' performance utilizing a decision support software," *Neural Comput. Appl.*, vol. 31, no. 6, pp. 1683–1694, 2019, doi: [10.1007/s00521-018-3756-y](https://doi.org/10.1007/s00521-018-3756-y).
- [51] T. W. Lim, K. C. Khor, and K. H. Ng, "Dimensionality reduction for predicting student performance in unbalanced data sets," *Int. J. Adv. Soft Comput. Appl.*, vol. 11, no. 2, pp. 76–86, 2019.
- [52] A. H. Filho, F. Concatto, J. Nau, H. A. D. Prado, D. O. Imhof, and E. Ferneda, "Imbalanced learning techniques for improving the performance of statistical models in automated essay scoring," *Proc. Comput. Sci.*, vol. 159, pp. 764–773, Jan. 2019, doi: [10.1016/j.procs.2019.09.235](https://doi.org/10.1016/j.procs.2019.09.235).
- [53] G. Dimic, D. Rancic, O. Pronic-Rancic, and D. Milosevic, "An approach to educational data mining model accuracy improvement using histogram discretization and combining classifiers into an ensemble," in *Smart Education and e-Learning* (Smart Innovation, Systems and Technologies), vol. 144. Singapore: Springer, 2019, pp. 267–280, doi: [10.1007/978-981-13-8260-4_25](https://doi.org/10.1007/978-981-13-8260-4_25).
- [54] M. Hlosta, Z. Zdrahal, and J. Zendulka, "Are we meeting a deadline? Classification goal achievement in time in the presence of imbalanced data," *Knowl.-Based Syst.*, vol. 160, pp. 278–295, Nov. 2018, doi: [10.1016/j.knosys.2018.07.021](https://doi.org/10.1016/j.knosys.2018.07.021).
- [55] X. Zhang, R. Xue, B. Liu, W. Lu, and Y. Zhang, "Grade prediction of student academic performance with multiple classification models," in *Proc. 14th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Jul. 2018, pp. 1086–1090, doi: [10.1109/FSKD.2018.8687286](https://doi.org/10.1109/FSKD.2018.8687286).

- [56] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araujo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students academic failure in introductory programming courses," *Comput. Hum. Behav.*, vol. 73, pp. 247–256, Aug. 2017, doi: [10.1016/j.chb.2017.01.047](https://doi.org/10.1016/j.chb.2017.01.047).
- [57] H. Al-Shehri, A. Al-Qarni, L. Al-Saati, A. Batoaq, H. Badukhen, S. Alrashed, J. Alhiyafi, and S. O. Olatunji, "Student performance prediction using support vector machine and K-nearest neighbor," in *Proc. IEEE 30th Can. Conf. Electr. Comput. Eng. (CCECE)*, Apr. 2017, pp. 28–31, doi: [10.1109/CCECE.2017.7946847](https://doi.org/10.1109/CCECE.2017.7946847).
- [58] M. Sivasakthi, "Classification and prediction based data mining algorithms to predict students' introductory programming performance," in *Proc. Int. Conf. Inven. Comput. Informat. (ICICI)*, 2017, pp. 3–7, doi: [10.1145/3167486.3167520](https://doi.org/10.1145/3167486.3167520).
- [59] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," *Comput. Educ.*, vol. 103, pp. 1–15, Dec. 2016, doi: [10.1016/j.compedu.2016.09.005](https://doi.org/10.1016/j.compedu.2016.09.005).
- [60] A. Polyzou and G. Karypis, "Grade prediction with models specific to students and courses," *Int. J. Data Sci. Anal.*, vol. 2, nos. 3–4, pp. 159–171, Dec. 2016, doi: [10.1007/s41060-016-0024-z](https://doi.org/10.1007/s41060-016-0024-z).
- [61] G. Badr, A. Algobail, H. Almutairi, and M. Almutery, "Predicting students performance in university courses: A case study and tool in KSU mathematics department," *Proc. Comput. Sci.*, vol. 82, pp. 80–89, Jan. 2016, doi: [10.1016/j.procs.2016.04.012](https://doi.org/10.1016/j.procs.2016.04.012).
- [62] P. Kaur, M. Singh, and G. S. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," *Proc. Comput. Sci.*, vol. 57, pp. 500–508 Mar. 2015, doi: [10.1016/j.procs.2015.07.372](https://doi.org/10.1016/j.procs.2015.07.372).
- [63] S. T. Jishan, R. I. Rasha, N. Haque, and R. M. Rahman, "Improving accuracy of students final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique," *Decis. Anal.*, vol. 2, no. 1, pp. 1–25, Dec. 2015, doi: [10.1186/s40165-014-0010-2](https://doi.org/10.1186/s40165-014-0010-2).
- [64] G. Akçapinar, A. Altun, and P. Askar, "Modeling students academic performance based on their interactions with the online learning environment," *Elem. Educ. Online*, vol. 14, no. 2, pp. 815–824, Aug. 2015, doi: [10.17051/io.2015.03160](https://doi.org/10.17051/io.2015.03160).
- [65] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Hum. Behav.*, vol. 104, Mar. 2020, Art. no. 106189, doi: [10.1016/j.chb.2019.106189](https://doi.org/10.1016/j.chb.2019.106189).
- [66] D. Hernández-Leo, R. Martínez-Maldonado, A. Pardo, J. A. Muñoz-Cristóbal, and M. J. Rodríguez-Triana, "Analytics for learning design: A layered framework and tools," *Brit. J. Educ. Technol.*, vol. 50, no. 1, pp. 139–152, Jan. 2019, doi: [10.1111/bjet.12645](https://doi.org/10.1111/bjet.12645).
- [67] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
- [68] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: Overview study and experimental results," in *Proc. 1st Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2020, pp. 243–248, doi: [10.1109/ICICS49469.2020.239556](https://doi.org/10.1109/ICICS49469.2020.239556).
- [69] F. J. Ferri, J. V. Albert, and E. Vidal, "Considerations about sample-size sensitivity of a family of edited nearest-neighbor rules," *IEEE Trans. Syst., Man, Cybern., B*, vol. 29, no. 5, pp. 667–672, Oct. 1999, doi: [10.1109/3477.790454](https://doi.org/10.1109/3477.790454).
- [70] I. Tomek, "Two modifications of CNN," *IEEE Trans. Syst., Man, Cybern., B*, vol. SMC-6, no. 11, pp. 769–772, Nov. 1976, doi: [10.1109/TSMC.1976.4309452](https://doi.org/10.1109/TSMC.1976.4309452).
- [71] Y. Pristyanto, N. A. Setiawan, and I. Ardiyanto, "Hybrid resampling to handle imbalanced class on classification of student performance in classroom," in *Proc. 1st Int. Conf. Informat. Comput. Sci. (ICICoS)*, Nov. 2017, pp. 207–212, doi: [10.1109/ICICoS.2017.8276363](https://doi.org/10.1109/ICICoS.2017.8276363).
- [72] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: [10.1109/TKDE.2008.239](https://doi.org/10.1109/TKDE.2008.239).
- [73] K. Deepika and N. Sathyalarayana, "Hybrid model for improving student academic performance," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 10, pp. 768–779, 2020.
- [74] N. Aslam, I. U. Khan, L. H. Alamri, and R. S. Almuslim, "An improved early student's performance prediction using deep learning," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 12, pp. 108–122, 2021, doi: [10.3991/ijet.v16i12.20699](https://doi.org/10.3991/ijet.v16i12.20699).
- [75] W. Intayoad, C. Kamayod, and P. Temdee, "Synthetic minority oversampling for improving imbalanced data in educational web usage mining," *ECTI Trans. Comput. Inf. Technol. (ECTI-CIT)*, vol. 12, no. 2, pp. 118–129, Feb. 2019, doi: [10.37936/ecti-cit.2018122.133280](https://doi.org/10.37936/ecti-cit.2018122.133280).
- [76] L. Wang, M. Han, X. Li, N. Zhang, and H. Cheng, "Review of classification methods on unbalanced data sets," *IEEE Access*, vol. 9, pp. 64606–64628, 2021, doi: [10.1109/ACCESS.2021.3074243](https://doi.org/10.1109/ACCESS.2021.3074243).
- [77] A. Jain, "Multi-class classification to track students academic outcome," *Tech. Rep.*, 2018.
- [78] B. Predic, G. Dimic, D. Rancic, P. Strbac, N. Macek, and P. Spalevic, "Improving final grade prediction accuracy in blended learning environment using voting ensembles," *Comput. Appl. Eng. Educ.*, vol. 26, no. 6, pp. 2294–2306, Nov. 2018, doi: [10.1002/cae.22042](https://doi.org/10.1002/cae.22042).



SITI DIANAH ABDUL BUJANG received the B.S. degree in science (computer science) and the M.S. degree in science from the Universiti Teknologi Malaysia (UTM), Kuala Lumpur, in 2006 and 2010, respectively, where she is currently pursuing the Ph.D. degree in software engineering with the Malaysia-Japan International Institute of Technology. Her thesis focuses on the application of predictive analytics in a higher education institution. From 2010 to 2019, she was a Senior Lecturer with the Information and Communication Technology Department, Polytechnic Sultan Idris Shah, Sabak Bernam, Selangor, Malaysia. She has experience in developing the Polytechnic Curriculum for Diploma in Information Technology (Technology Digital) 2.5 years' program. She is one of the book authors that contribute for the Department of Polytechnic and Community College Education. Her research interests include machine learning, data analytics, predictive analytics, learning analytics, and educational data mining.

Lecturer with the Information and Communication Technology Department, Polytechnic Sultan Idris Shah, Sabak Bernam, Selangor, Malaysia. She has experience in developing the Polytechnic Curriculum for Diploma in Information Technology (Technology Digital) 2.5 years' program. She is one of the book authors that contribute for the Department of Polytechnic and Community College Education. Her research interests include machine learning, data analytics, predictive analytics, learning analytics, and educational data mining.



ALI SELAMAT (Member, IEEE) has been with the Dean of the Malaysia Japan International Institute of Technology (MJIT), Universiti Teknologi Malaysia (UTM), Malaysia, since 2018. He is currently a Full Professor with the UTM, where he is also a Professor with the Department of Software Engineering, Faculty of Computing. An academic institution established under the cooperation of the Japanese International Cooperation Agency (JICA) and the Ministry of Education Malaysia (MOE) to provide the Japanese style of education in Malaysia. He has published more than 60 IF research papers. His H-index is 20, and his number of citations in WoS is more than 800. His research interests include software engineering, software process improvement, software agents, web engineering, information retrievals, pattern recognition, genetic algorithms, neural networks, soft computing, computational collective intelligence, strategic management, key performance indicator, and knowledge management. He has been serving as the Chair for the IEEE Computer Society Malaysia, since 2018. He is on the Editorial Board of the *Journal Knowledge-Based Systems* (Elsevier).



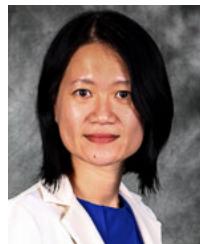
ONDREJ KREJCAR received the Ph.D. degree in technical cybernetics from the Technical University of Ostrava, Czech Republic, in 2008. He is a Full Professor in systems engineering and informatics with the University of Hradec Kralove, Faculty of Informatics and Management, Center for Basic and Applied Research, Czech Republic; and Research Fellow at Malaysia-Japan International Institute of Technology, University Technology Malaysia, Kuala Lumpur, Malaysia and a Vice-Rector for science and creative activities of the University of Hradec Kralove from June 2020 and also the Director of the Center for Basic and Applied Research at the University of Hradec Kralove. He is also a guarantee of the doctoral study program in Applied Informatics with the University of Hradec Kralove, where he is focusing on lecturing on smart approaches to the development of information systems and applications in ubiquitous computing environments. From 2016 to 2020, he was the Vice-Dean for science and research at the Faculty of Informatics and Management, UHK. His H-index is 25 according to Web of Science, with more than 2700 citations received in the Web of Science, where more than 160 IF journal articles is indexed in JCR index (h-index 29 at SCOPUS with more than 3500 citations). His research interests include technical cybernetics, ubiquitous computing, control systems, smart sensors, wireless technology, biomedicine, image segmentation and recognition, biometrics, biotelemetric system architecture (portable device architecture, wireless biosensors), development of applications for mobile / remote devices with use of remote or embedded biomedical sensors. He has been a Vice-Leader and Management Committee member at WG4 at project COST CA17136, since 2018. He has also been a Management Committee member substitute at project COST CA16226 since 2017. Since 2019, he has been a Chairperson of the Program Committee of the KAPPA Program, Technological Agency of the Czech Republic as a regulator of the EEA/Norwegian Financial Mechanism in the Czech Republic (2019 to 2024). Since 2020, he has been a Chairperson of the Panel 1 (Computer, Physical and Chemical Sciences) of the ZETA Program, Technological Agency of the Czech Republic. From 2014 to 2019, he was the Deputy Chairperson of the Panel 7 (Processing Industry, Robotics, and Electrical Engineering) of the Epsilon Program, Technological Agency of the Czech Republic. In 2018, he was the 14th top peer Reviewer in Multidisciplinary in the World according to Publons and a Top Reviewer in the Global Peer Review Awards 2019 by Publons. Currently, he is on the editorial board of the MDPI Sensors IF journal (Q1/Q2 at JCR), and several other ESCI indexed journals.



FARHAN MOHAMED (Senior Member, IEEE) received the B.Sc. degree in computer science from the Universiti Teknologi Malaysia, in 2003, the Master of Science degree in web performance optimization, in 2006, and the Ph.D. degree in computer science (data visualization) from Swansea University, in 2014. He was appointed as a Chief Technology Officer (CTO) of MaGICX Sdn. Bhd., from 2015 to 2017. He is currently the Deputy Director of Corporate Affairs, Publication and Research with MaGICX, UTM. His current work focuses on digital branding, gamification, data visualisation and augmented, and virtual reality. His research interests include visual analytics, visual metaphors, human-computer interactions, virtual environment, and procedural computer graphics. He is also a Committee Member of the IEEE Computer Society Malaysia, and a member of ACM MyHCI-UX Malaysia. He was a member of the 100 ACM Global Practitioner Advisory Committee, from 2017 to 2018. He was a fellow of the CEO@Faculty 2.0 (2017–2018) attached to Huawei Malaysia. He received the Royal Academy of Engineering U.K., and Leaders in Innovation Fellowship, in 2017. He received the Excellence Student Award by Juita-UTM, in 2003, for his B.Sc. degree. He has been working on Web Performance Optimization for his Master of Science study which the work has received Gold Award at Malaysian Technology Expo 2006 and Gold Award at the International Exhibition of Inventions, Geneva, in 2006.



LIM KOK CHENG (Member, IEEE) is currently pursuing the Ph.D. degree in usability engineering with the Department of Computing, School of Engineering, Universiti Teknologi Malaysia. He started his career as an IT Analyst Consultant for a New Zealand, Singapore, and Malaysian-based company. It is within this period of time that he obtained specific experience and field knowledge in software marketing, testing, and training. With his clear interest in academics, he expanded his knowledge by continuing his master's degree while at the same time conducted multiple research plus development works with the Ministry of Science, Technology and Innovation (MOSTI), in the midst of also tutoring in the Universiti Tenaga Nasional (UNITEN), Malaysia, and Asia Pacific University (APU). He joined UNITEN's permanent academic force as a Lecturer, in 2010, and has been contributing to the university to date. He is also a Lecturer with the College of Computing and Informatics, UNITEN. He is also working as the Head of Software Engineering Program, the Head of External Relations Unit, UNITEN, and a Treasurer for the IEEE Malaysia Section Computer Chapter. He has accumulated 15 years of teaching and training experiences, produces more than a 100 student projects and won multiple awards in international innovation competitions.



PO CHAN CHIU received the M.Sc. degree in information technology from the Universiti Malaysia Sarawak (UNIMAS), in 2010. She is currently pursuing the Ph.D. degree in computer science with the Universiti Teknologi Malaysia (UTM). She started her career as a Software Engineer for three years. She is also working as a Lecturer with the UNIMAS. She worked on several consultancy projects and developed software solutions to meet the needs of the woodworking industry. Her research interests include artificial intelligence, data analytics, optimization, and neural networks.



HAMIDO FUJITA (Life Senior Member, IEEE) received the Doctor Honoris Causa degree from Óbuda University, Budapest, Hungary, in 2013, and the Doctor Honoris Causa degree from Timisoara Technical University, Timisoara, Romania, in 2018. He is currently an Emeritus Professor with Iwate Prefectural University, Takizawa, Japan. He is also the Executive Chairman of i-SOMET Incorporated Association, Morioka, Japan. He received the title of an Honorary Professor from Óbuda University, in 2011. He is a Distinguished Research Professor with the University of Granada, and an Adjunct Professor with Stockholm University, Stockholm, Sweden; University of Technology Sydney, Ultimo, NSW, Australia; National Taiwan Ocean University, Keelung, Taiwan; and others. He has supervised Ph.D. students jointly with the University of Laval, Quebec City, QC, Canada; University of Technology Sydney; Oregon State University, Corvallis, OR, USA; University of Paris 1 Pantheon-Sorbonne, Paris, France; and the University of Genoa, Italy. He has published more than 400 highly cited papers. He has four international patents in software systems and several research projects with Japanese industry and partners. He was a recipient of the Honorary Scholar Award from the University of Technology Sydney, in 2012. He is the Emeritus Editor-in-Chief of *Knowledge-Based Systems*. He is also the Editor-in-Chief of *Applied Intelligence* (Springer), he headed a number of projects, including intelligent HCI, a project related to mental cloning for healthcare systems as an intelligent user interface between human-users and computers, and SCOPE project on virtual doctor systems for medical applications. He is a Highly Cited Researcher in Cross-Field for the year 2019 and 2020 in computer science field, respectively from Clarivate Analytics. He collaborated with several research projects in Europe, and recently he is collaborating in OLIMPIA project supported by Tuscany region on Therapeutic monitoring of Parkinson disease.