Received 27 November 2023, accepted 28 December 2023, date of publication 8 January 2024, date of current version 19 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351569

RESEARCH ARTICLE

GLoW SMOTE-D: Oversampling Technique to Improve Prediction Model Performance of Students Failure in Courses

SUSANA LIMANTO[1,2], JOKO LIANTO BULIALI[1], (Member, IEEE), AND AHMAD SAIKHU[1]
[1]Informatics Department, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[2]Department of Informatics Engineering, Universitas Surabaya, Surabaya 60293, Indonesia

Corresponding author: Joko Lianto Buliali (joko@if.its.ac.id)

This work was supported by the Doctoral Dissertation Research Funds from the University of Surabaya.

ABSTRACT The percentage of passing courses is dependent on the assistance provided to students. To ensure the effectiveness of these efforts, identifying students at risk of course failure as early as possible is crucial. The list of students at risk can be generated through academic performance prediction based on historical data. However, the number of students failing (7%) is significantly lower than the number succeeding (93%), resulting in a class imbalance that hampers performance. A widely adopted technique for addressing class imbalance issues is synthetic sample oversampling. Many oversampling techniques neglect discrete features, whereas the existing technique for discrete features treats all features uniformly and does not select samples as a basis for generating synthetic data. This limitation is capable of introducing noise and borderline samples. As a result, this study introduced a novel discrete feature oversampling technique called GLoW SMOTE-D. This technique accelerated the improvement of minority sample learning by performing multiple selections and multiple weighting in order to effectively reduce noise. Experimental results showed that this technique significantly enhanced the performance of students' failure in the course prediction model when compared to various other techniques across a range of performance measures and classifiers.

INDEX TERMS Discrete, imbalanced dataset, oversampling, students' failure.

I. INTRODUCTION

The academic success of students is very important in shaping the quality of higher education and its perception by the public. Students indirectly serve as advocates for higher education because the information they share with friends, family, or the community can positively or negatively impact the reputation of these institutions. Consequently, it is necessary for higher education to enhance the quality of their students. One effective technique for achieving this goal is by increasing the percentage of students who successfully pass courses. This can be accomplished by offering support and assistance outside class to those who are at risk of failing courses [1], [2], [3]. Efforts to improve the percentage of passing courses by providing additional support outside the classroom have been carried out by Khan et al. [2]. The result showed that mentoring outside the classroom setting contributed to an increased percentage of passing courses [2].

To ensure that this support is both targeted and optimized, students at risk of failure need to be identified as early as possible [4]. This can be accomplished through the prediction of students' academic performance in their courses.

Prediction models rely on historical data on students' performance in courses. However, a significant imbalance often exists between the number of students who pass and those who fail. Data about failing students is scarcer than that of successful individuals, creating a class imbalance issue. In binary class imbalances, the category with fewer instances is termed the minority sample (positive sample), whereas the other is labeled the majority (negative sample) [5], [6], [7], [8]. This substantial disparity in sample sizes severely impacts the predictive capacity of the minority sample, leading to less reliable prediction [5], [6], [8], [9], [10], [12], [13], [14]. This bias toward the majority sample is due to classifiers

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

favoring it [12], [14], [15]. In the context of predicting failure in courses, it is vital to prioritize the predictive ability of the minority sample over the majority. Neglecting the prediction accuracy of minority samples results in an unusable list of students at risk of failure generated by the model due to a high number of misclassified minority samples.

Numerous studies have been conducted to address class imbalance issues and enhance the performance of minority samples. One effective technique to enhance the performance of minority samples is by increasing their learning capabilities through a technique called oversampling [5], [8], [9], [10], [11], [12]. Conceptually, oversampling is the process of increasing the number of samples, generally minority samples [6], [7], [8], [12], [14], [16].

In some scenarios, transforming continuous features into discrete ones through a process known as binning can help mitigate noise [17]. This transformation simplifies the mining process, rendering it more efficient and comprehensible [17], and often leading to improved performance [14], [17], [18]. This insight serves as a motivation for this study to explore datasets with discrete features. However, the number of oversampling techniques applicable to discrete features is limited.

Random Over Sampling (ROS), Synthetic Minority Oversampling TEchnique-Nominal (SMOTE-N), and SMOTE—Encoded Nominal and Continuous (SMOTE-ENC) are three oversampling techniques that can be utilized for discrete features. In ROS, minority samples are randomly duplicated to create synthetic samples. However, SMOTE-N and SMOTE-ENC generate synthetic samples for each minority sample based on the closest distance, with equal weighting for each feature. These techniques for generating synthetic samples may introduce noise whereas the uniform weighting can lead to non-representative samples, which can cause potential losses.

This study aims to enhance the performance of predicting students' failure in the course model. A prediction model is developed using a dataset comprising discrete features. Performance enhancement of minority samples is achieved through the application of the proposed synthetic data oversampling technique known as Global and Local Weighting on SMOTE with Discrete features (GLoW SMOTE-D). This study introduces several novel aspects:

1. Positive samples serving as the foundation for generating synthetic samples are obtained through two selection processes. The initial selection relies on predicted results, with a focus on positive samples, specifically the challenging ones that have been misclassified. To mitigate noise, a secondary selection is performed based on the number of majority samples within the k-Nearest Neighbors (kNN). Based on the available information, no existing oversampling technique for discrete features has implemented this selection process.
2. To address variations in correlation between individual features and the target class, two types of weightings are applied. The first is known as global weight, which is calculated using the entire dataset. However, relying solely on kNN for generating synthetic samples renders global weights less representative [19]. An adjustment is made to the global weight value based on n-Nearest Neighbors (nNN) to produce the second weight, known as local weight. These local weights are used in identifying the kNN, forming the basis for synthetic sample generation. It is important to acknowledge that no oversampling technique for discrete features has incorporated this weighting technique. Whereas AWH-SMOTE considers weighted features to determine kNN, these weights are calculated using the entire dataset, making them less representative.
3. The proposed oversampling technique is applied to a primary dataset for assessing students' academic performance in courses. Each of the proposed oversampling techniques has only been tested using publicly available data and their real-world benefits remain less evident.

The remainder of this paper is structured into four sections, with Section II presenting an overview of related work. Section III provides a detailed explanation of the proposed oversampling technique. Section IV explores the application of this technique in predicting students' failure in courses, offering analysis and discussion. Finally, Section V presents conclusions and outlines potential avenues for future exploration.

II. RELATED WORKS

In this section, several studies on predicting student failure in courses and techniques for oversampling synthetic data are presented.

A. PREDICTION MODEL FOR STUDENT FAILURE IN COURSES

The outcomes of predicting student failure in courses can be used to design interventions for supporting students at risk, in order to increase the percentage of passing courses. In reference to [2], a predictive model was developed to identify students at risk of failing the Phonetics and Phonology courses at Buraimi University College (BUC) in the Sultanate of Oman. This model used the academic performance data of participants in the Phonetics and Phonology courses over three semesters as a foundation for prediction. Prediction was generated at the end of the sixth week, following the first test score, which was included as one of the key features of the predictive model. Consequently, an impressive accuracy rate of 86.1% and an excellent precision of 92.7% were achieved. Moreover, the model was successfully applied to a course with 25 participants, revealing that five students were indeed at risk of failing the course. These five students received additional support outside of regular class hours from both their instructors and academic advisors. The end-of-semester evaluation revealed that four out of the five students made substantial progress and completed the course.

TABLE 1. Comparison with other related works on prediction times and oversampling techniques used.

Study	Prediction Time	Oversampling Technique
[2]	At the end of the sixth week	-
[20]	At the end of the third, sixth, ninth, twelfth, and fourteenth weeks	-
(This study)	**Before course begins**	**GLoW SMOTE-D**

Akçapınar et al. [20] conducted predictive analysis for students' failure in the Computer Hardware courses, considering multiple stages of prediction—specifically, at the end of the third, sixth, ninth, twelfth, and fourteenth weeks. The features used in this study were related to students' engagement with the Learning Management System. It was found that accuracy improved as the sample size was increased at each prediction stage.

Table 1 shows a comparison between this study and others in terms of predicting student failure in courses. This research involves making predictions before the course begins, allowing timely interventions for at-risk students to maximize benefits. To improve prediction model performance on minority data, a synthetic data oversampling process is incorporated during preprocessing.

B. OVERSAMPLING TECHNIQUES

ROS is known as the most straightforward technique as it randomly duplicates minority samples until a desired balance level is achieved [21]. However, this replication can lead to lower variance in the final samples compared to the initial ones, potentially causing overfitting issues [7], [8], [9], [10], [14], [21], [22]. To solve this problem, an alternative oversampling technique was devised to create synthetic samples.

One prominent synthetic oversampling technique is the Synthetic Minority Oversampling Technique (SMOTE), developed by Chawla et al. [9]. In SMOTE, new minority samples are generated by interpolating linearly between adjacent minority samples in the feature space. This technique expands the decision area, making it less precise, but it also frequently introduces noise [21] and borderline samples [23]. To address these issues, various advanced oversampling techniques were developed, including Borderline SMOTE [5], Advanced SMOTE (A-SMOTE) [23], Safe level SMOTE [10], The Adaptive Synthetic Sampling Technique for Imbalanced Learning (ADASYN) [11], Attribute Weighted and kNN Hub on SMOTE (AWH-SMOTE) [8], and Constrained Oversampling [21]. These techniques overcome these limitations by selecting positive samples as the basis for generating synthetic samples. The selection is often based on the number of majority samples in the kNN.

Most oversampling techniques assign equal weight to all features, but in some cases, the existing features have different correlations with the target class. Treating all features equally when finding nearest neighbors can result in non-representative neighbors, leading to losses [12]. To address this issue, [8] proposed a technique called AWH-SMOTE. AWH-SMOTE mitigates differences in correlation among features by assigning distinct weights to each existing feature. To enhance learning capabilities, synthetic samples are generated in regions with a higher concentration of negative samples. Whereas weights are calculated based on the entire dataset, the generation of synthetic samples relies solely on kNN, resulting in less representative weights.

These various oversampling techniques have proven their ability to enhance the performance of the prediction model. However, it is important to note that these techniques predominantly focus on continuous features [24]. In practical scenarios, datasets sometimes consist of categorical features, and applying the mentioned techniques to datasets with such features can lead to failures [24].

The SMOTE-N is an oversampling technique tailored for handling discrete features. In SMOTE-N, synthetic samples are generated from each positive sample, similar to the original SMOTE technique. However, the feature set of these synthetic samples is created based on majority voting of the positive sample features and their kNN [9], determined using the Value Difference Metric (VDM) for distance calculation. It is important to note that SMOTE-N shares a similar limitation with SMOTE, potentially introducing noise into the data.

Another oversampling technique designed to address discrete features is SMOTE-ENC, pioneered by Mukherjee and Khushi [24]. In this technique, prior to generating synthetic samples, each discrete feature goes through numerical encoding. A higher numerical value signifies a stronger association with the minority class [24]. This encoding is derived from the principles of Pearson's chi-squared test. The generation of synthetic samples is then based on the kNN technique. Continuous features within synthetic samples are generated in relation to SMOTE technique, while discrete features follow the same technique as SMOTE-N. The performance of the proposed technique is evaluated using five public datasets, showing good prediction results.

SMOTE-N and SMOTE-ENC generate synthetic samples from each positive sample using kNN technique, which can introduce noise and borderline samples. Additionally, assigning equal weight to each feature during the nearest neighbor search can yield non-representative neighbors, potentially leading to losses.

Table 2 shows the comparison of the oversampling techniques of this study and other related works. The proposed technique focuses on discrete features and assigns varying weights to each feature based on the correlation with the target class. Positive samples are selected to generate synthetic samples to reduce noise.

III. A NEW OVERSAMPLING METHOD: GLOW ON DISCRETE FEATURE OVERSAMPLING

The main objective of this study was to predict academic performance in courses and identify students at risk of failure to facilitate the implementation of timely preventive measures.

TABLE 2. **Comparison with other related works on specification of oversampling techniques.**

Techniques	Features	Candidate	Feature weighting
ROS	Continuous and discrete	Random	No
SMOTE	Continuous	All positive samples	No
Borderline SMOTE	Continuous	Positive samples are in borderline areas	No
A-SMOTE	Continuous	All positive samples	No
Safe level SMOTE	Continuous	Positive samples that are not noise with a safe level ratio not equal to ∞	No
ADASYN	Continuous	Positive samples that have negative samples as nearest neighbors	No
AWH-SMOTE	Continuous	Positive samples are in the borderline area and have the lowest safe value	Yes
Constrained Oversampling	Continuous	Positive samples are in the overlapping area	No
SMOTE-N	Discrete	All positive samples	No
SMOTE-ENC	Continuous and discrete	All positive samples	No
GLoW SMOTE-D	**Discrete**	**Misclassified positive samples that are not noise**	**Yes**

A substantial class imbalance issue existed, with significantly fewer failing students compared to those who succeeded. This imbalance could detrimentally affect the performance of the predictive model, particularly for the minority sample, which constituted the main focus of the study. As a result, there is a need to rectify the distribution of target classes.

To address the class imbalance problem, synthetic data oversampling was used in this investigation. The proposed oversampling technique, named GLoW SMOTE-D, was visually shown in Fig. 1, and comprised three key stages, namely extraction of the misclassified minority sample set, computation of global weights, and synthetic minority oversampling.

A. THE EXTRACTION OF THE MISCLASSIFIED MINORITY SAMPLE SET

The process of generating synthetic data started with the acquisition of a misclassified minority sample set, accomplished using the Decision Tree technique—referred to as the base learner. Using a 10-fold cross-validation technique, this base learner yielded ten sets of inaccurately predicted data. The final misclassified minority sample set used for synthetic data generation combined these ten sets from each fold.

B. GLOBAL WEIGHT CALCULATION

Given that each feature possesses distinct correlations with the target class, it became essential to assign unique weights based on these correlations. This weight assignment significantly influenced the performance of kNN classification model [8]. The computation of feature weights followed the Weights Optimizing Distance algorithm proposed by Wojna [25], described in Algorithm 1. These weight computations comprised the entire dataset, termed global weights, and relied on the Misclassification Ratio—a metric representing the ratio between the sums of distances to the nearest neighbors $\rho(\mathbf{x}, \text{nearest}(\mathbf{x}))$ for misclassified samples and all training objects [19], [25].

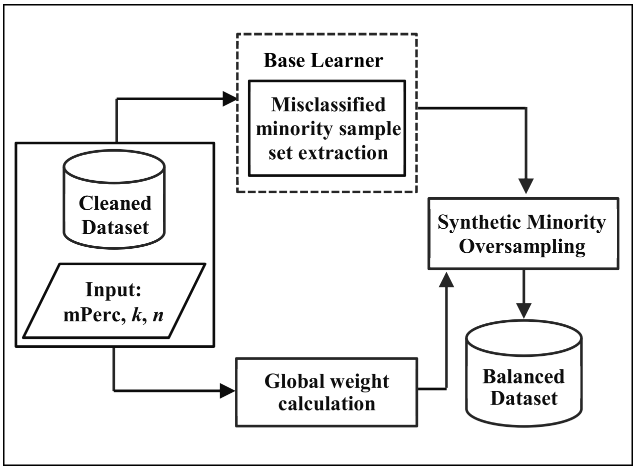

FIGURE 1. **Steps of GLoW SMOTE-D.**

Algorithm 1 Weight Optimizing Distance

U_{trn} = training dataset
dec($\mathbf{x}$) = class label of $\mathbf{x}$

for each attribute $w_i = 1.0$
modifier = 0.9
convergence = 0.9
repeat l times
  S_{trn} = a random training sample from U_{trn}
  S_{tst} = a random test sample from U_{trn}
  $\text{MR} = \frac{\sum_{\mathbf{x}\in S_{tst}:\, \text{dec}(\mathbf{x})\neq\text{dec}(\text{nearest}(\mathbf{x}))} \rho(\mathbf{x},\text{nearest}(\mathbf{x}))}{\sum_{\mathbf{x}\in S_{tst}} \rho(\mathbf{x},\text{nearest}(\mathbf{x}))}$
  for each attribute a_i
    $\text{MR}(a_i) = \frac{\sum_{\mathbf{x}\in S_{tst}:\, \text{dec}(\mathbf{x})\neq\text{dec}(\text{nearest}(\mathbf{x}))} \rho_i(\mathbf{x}_i,\text{nearest}(\mathbf{x})_i)}{\sum_{\mathbf{x}\in S_{tst}} \rho_i(\mathbf{x}_i,\text{nearest}(\mathbf{x})_i)}$
  END for
  for each attribute a_i
    if $\text{MR}(a_i) > \text{MR}$ then $w_i = w_i +$ modifier
  END for
  modifier = modifier * convergence
END repeat

When representing samples as discrete value vectors $\mathbf{x} = \{x_1, \cdots, x_m\}$, where m is the number of features, the distance between two samples $\mathbf{x} = \{x_1, \cdots, x_m\}$, and $\mathbf{y} = \{y_1, \cdots, y_m\}$, is defined by (1).

$$\rho(\text{x}, \text{y}) = \sum_{i=1}^{m} w_i \times \rho_i(x_i, y_i) \tag{1}$$

where $\rho_i(x_i, y_i)$ is a measure of similarity between two feature values and w_i represents feature weight whose value is updated during the global weight calculation process. The first time w_i are initialized with 1 and two discrete values x_i and y_i are considered to be similar when they imply similar decision distribution. The value of $\rho_i(x_i, y_i)$ is calculated using the VDM as seen in (2) [19].

$$\rho_i(x_i, y_i) = \sum_{c=1}^{C} \left| P_{i,x_i,c} - P_{i,y_i,c} \right| = \sum_{c=1}^{C} \left| \frac{N_{i,x_i,c}}{N_{i,x_i}} - \frac{N_{i,y_i,c}}{N_{i,y_i}} \right| \tag{2}$$

where C is the number of target class labels.

C. SYNTHETIC MINORITY OVERSAMPLING

In the Synthetic Minority Oversampling phase, the insight provided by Skowron and Wojna [19] suggested that creating a global model from existing data was often infeasible due to the intricate nature of real-world phenomena. Instead, Skowron and Wojna [19] proposed the use of a local model based on test samples, a technique seamlessly incorporated into this study for oversampling purposes.

Normalization of global weights was applied to construct the local model (distance function) as shown in (3). For each minority sample with an incorrect prediction ($\mathbf{x}$), a local model was constructed by selecting its nNN ($N(\mathbf{x}, n)$). Local weight calculations based on $N(\mathbf{x}, n)$ were then executed, leveraging Algorithm 1 to derive these weights, with initial values drawn from global weights.

$$Z_i = \frac{w_i}{\sum_{i=1}^{m} w_i} \tag{3}$$

Oversampling for sample $\mathbf{x}$ comprises the selection of kNN from $N(\mathbf{x}, n)$, with normalized local weights guiding this process. When all kNN possess class labels differing from the class label of $\mathbf{x}$, $\mathbf{x}$ is categorized as noise [5], [10], rendering it ineligible for synthetic sample generation.

In cases where $\mathbf{x}$ is not classified as noise, the features of the synthetic sample are generated based on the mode between $\mathbf{x}$ and its kNN. In scenarios with multiple modes, random selection is implemented. The quantity of synthetic positive samples produced is adaptable, determined by the defined mPerc parameter—specifying the percentage of minority samples relative to majority samples post-synthetic sample generation.

The produced synthetic sample combines all individual synthetic samples, with the specific process described in Algorithm 2.

IV. EXPERIMENT

The steps taken in this study were in line with the flowchart shown in Fig. 2. The focus of the study was to develop an oversampling technique to enhance the performance of students' failure prediction model for enrolled courses.

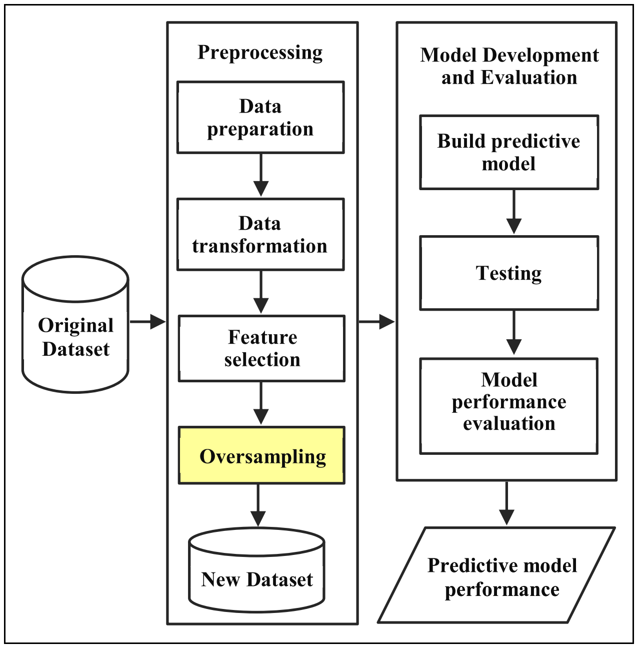

FIGURE 2. Research methodology.

A. DATASET

The dataset used comprised 3,712 samples from 246 students at a reputable private university in Surabaya, Indonesia. These students gained admission into the university for the past three years, from 2019 to 2021. Each sample included 21 features consisting of demographic information, high school academic records, and higher education performance.

Demographic and academic performance data were the two primary types of features widely used and were acknowledged for their substantial impact in predicting student academic performance [26], [27]. As influential factors, these features offered benefits in improving prediction model performance, as well as motivating the incorporation of demographic data and academic performance as dataset features. However, it was essential to recognize that features considered significant in one study might not necessarily maintain the same influence when applied to a different dataset. Therefore, this study used feature selection to identify features more correlated with the target class.

After data preprocessing, the dataset was refined to contain 2,530 samples with 13 selected features. Among these, 2,346 samples were labeled as "Pass" (92.73%), while 184 samples were considered "Failed" (7.27%). The balance level for this dataset was 12.75, calculated as the ratio of majority to minority samples.

B. PREPROCESSING

The initial phase included tasks such as filling in missing features, eliminating irrelevant ones, and removing samples with infeasible missing data. Additionally, continuous features went through binning, and all features were discretized.

Algorithm 2 GLow SMOTE-D

m = the number of features in the dataset
SCleaned = preprocessing result dataset
Input:
Smin = dataset containing incorrectly predicted minority samples
$\mathbf{z}$ = normalized global weights
mPerc = the percentage of the number of minority samples to the majority samples after the generation of synthetic samples
k = the number of nearest neighbors that will be used to generate synthetic data ($k \leq n$)
n = the number of nearest neighbors that will be used to form a local model
Process:
for $i = 1$: number of samples in Smin
  $\mathbf{x}$ = the i-th sample from Smin
  $N(\mathbf{x}, n)$ = the set of n nearest neighbors of $\mathbf{x}$ based on ρ and $\mathbf{z}$
  $\mathbf{w_x}$ = local weighted VDM metric induced from the neighborhood $N(\mathbf{x}, n)$
  $\mathbf{z_x}$ = normalization result of $\mathbf{w_x}$
  $S(\mathbf{x}, k)$ = the set of k nearest neighbors of $\mathbf{x}$ based on ρ and $\mathbf{z_x}$
  if $\mathbf{x}$ is not noise then
    $\mathbf{x}$ is stored in the notNoise list
  END if
END for
nMinority = the number of synthetic minority samples generated from each notNoise member based on mPerc
for $i = 1$: number of samples in notNoise
  for $j = 1$: nMinority
    for $l = 1$: m
      $\text{Syn}_{ijl} = \text{mode}(x_{il},\ S(\mathbf{x},\ k)_{il})$
    END for
  END for
END for
SBalanced = Merger SCleaned and Syn
Output:
SBalance = balanced dataset

The results of this discretization process were subsequently used in the feature selection stage.

Feature selection relied on the Chi-square technique and the selected features were shown in Table 3. The dataset derived from this feature selection was termed the "cleaned dataset." Subsequently, this cleaned dataset was used to generate synthetic samples using the GLoW SMOTE-D algorithm. The combination of these synthetic samples with the original dataset resulted in a new dataset, which was used to evaluate the performance of the oversampling technique.

C. MODEL DEVELOPMENT AND EVALUATION

To assess the effectiveness of the oversampling technique in predicting students' failure in courses, three established classification techniques—DT, Naïve Bayes (NB), and Support Vector Machine (SVM)—were used. Training and testing were executed using a 10-fold cross-validation technique. The dataset was divided into ten equally sized partitions or folds. One-fold was allocated for testing, while the remaining nine were used for training. This process was iterated ten times, with each repetition using a different fold for testing. Prediction model performance metrics were calculated based on average values.

While total accuracy is a commonly used performance metric for prediction models, it often leads to misunderstandings in class-imbalanced datasets [4], [20]. In these cases, high accuracy can be achieved even when numerous minority samples are misclassified. Consequently, alternative measures were adopted, including recall, precision, F-measure, and Area Under the Receiver Operating Characteristic Curve (AUC).

The recall, precision, and F-measure metrics assessed prediction model performance from the minority sample perspective. Recall measured completeness, precision assessed exactness, and F-measure combined both completeness and exactness into a single value [2]. In predicting student failure, prioritizing the performance of minority samples was crucial. As a result, these three metrics were considered more appropriate for this exploration.

The Receiver Operating Characteristic (ROC) curve is a graphical representation illustrating the balance between the True Positive Rate (TPR) and False Positive Rate (FPR) of a classifier across varying threshold values [9], [14], [23]. The key characteristic of ROC curve was its resilience to test data imbalance [14], making it a crucial metric for evaluating imbalanced datasets. Meanwhile, AUC is a metric used in measuring the performance of ROC curve [9]. Based on the imbalanced nature of the dataset, AUC metric was particularly considered for use.

The computation of accuracy, recall, precision, and F-measure relied on the confusion matrix derived from the prediction process. In Table 4, True Positive (TP) represented correctly predicted positive samples, True Negative (TN) signified correctly predicted negative samples, False Positive (FP) reflected negative samples incorrectly classified as positive and False Negative (FN) denoted positive samples incorrectly classified as negative.

Accuracy, recall, precision, and F-measure values were calculated using (4) to (7). Greater values for recall, precision, and F-measure signified improved model performance for minority samples, while accuracy for the majority sample was determined using (8). In this scenario, (5) and (9) were used to calculate TPR and FPR, respectively. Model performance relied on the AUC metric, with a broader AUC indicating superior model performance.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{4}$$

$$\text{Recall} = \text{TPR} = \text{ACC}^{+} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{F} - \text{measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{7}$$

TABLE 3. Predictor variables used.

Name	Possible Value
Program	1 (Informatics Engineering), 2 (Business Information Systems), 3 (Multimedia)
The specialization chosen in high school	1 (Natural Sciences), 2 (Social Sciences), 3 (Languages), 4 (others)
Average math score at 12th grade (MScore)	1 (66 <= MScore < 73) 2 (73 <= MScore < 80) 3 (80 <= MScore < 87) 4 (87 <= MScore < 94) 5 (94 <= MScore < 101)
Average English score at 12th grade (EScore)	1 (66 <= EScore < 73) 2 (73 <= EScore < 80) 3 (80 <= EScore < 87) 4 (87 <= EScore < 94) 5 (94 <= EScore < 101)
Average score at 12th grade (AScore)	1 (66 <= AScore < 73) 2 (73 <= AScore < 80) 3 (80 <= AScore < 87) 4 (87 <= AScore < 94) 5 (94 <= AScore < 101)
Cumulative credits (CCredit)	1 (0 ≤ CCredit < 20) 2 (20 ≤ CCredit < 40) 3 (40 ≤ CCredit < 60) 4 (60 ≤ CCredit < 80) 5 (80 ≤ CCredit < 100) 6 (100 ≤ CCredit < 120) 7 (120 ≤ CCredit < 140)
Cumulative Grade Point Average (CGPA)	1 (0 < CGPA ≤ 1) 2 (1 < CGPA ≤ 2) 3 (2 < CGPA ≤ 3) 4 (3 < CGPA ≤ 4)
Number of course participants (num)	1 (1 <= num < 23) 2 (23 <= num < 45) 3 (45 <= num < 67) 4 (67 <= num < 89) 5 (89 <= num < 111)
Previous course grade	0 (E), 1 (D), 2 (C), 3 (BC), 4 (B), 5 (AB), 6 (A), 7 (if it is the first time he is taking the course)
Length of time repeat	0, 1, 2, …
Pass percentage (pass)	1 (31 ≤ pass < 41) 2 (41 ≤ pass < 51) 3 (51 ≤ pass < 60) 4 (61 ≤ pass < 71) 5 (71 ≤ pass < 81) 6 (81 ≤ pass < 91) 7 (91 ≤ pass < 101)
Prerequisite grade	0 (E), 1 (D), 2 (C), 3 (BC), 4 (B), 5 (AB), 6 (A), 7 (if there are no prerequisites)
Prerequisite time taken	0, 1, 2, …

TABLE 4. Confusion matrix for binary class problem.

Actual	Predicted	
	+	-
+	TP	FN
-	FP	TN

$$\mathrm{Acc}^{-} = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}} \tag{8}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{9}$$

To evaluate the performance of the proposed technique, a comparison was made with four oversampling techniques, namely the original dataset without oversampling (origin), ROS, SMOTE-N, and SMOTE-ENC. The selection of these three oversampling techniques for comparison was based on the distinctive focus of SMOTE-N on discrete features, while ROS and SMOTE-ENC could be applied to discrete features. Other oversampling techniques were unsuitable for discrete features because they concentrated on continuous features and used Euclidean distance calculations.

Statistical tests were conducted to facilitate the statistical analysis of experimental results. Two nonparametric statistical tests, namely Friedman Rank Test and Holm Post Hoc Tests [23], were used for hypothesis testing. Specifically, the Friedman Test determined whether there was a statistically significant difference between the mean performance values of prediction models from the four oversampling techniques. The null hypothesis posited that the mean performance values of prediction models from the four oversampling techniques were all equal. Meanwhile, the alternative hypothesis suggested that at least one population mean was different. In case the Friedman Test rejected the null hypothesis, the Holm Post Hoc Test would be applied to show that the mean performance value of the prediction model from GLoW SMOTE-D differed from the others. These tests were applied to the five performance metrics used in this study.

D. RESULT

In this study, consideration was given to four distinct minority sample rates, including 30%, 50%, 70%, and 100%. The minority sample rate signified the proportion of minority samples relative to the majority samples after generating synthetic samples. For each minority sample rate, the application of five oversampling techniques, namely origin, ROS, SMOTE-N, SMOTE-ENC, and GLoW SMOTE-D, was added to the existing dataset. With these oversampling techniques, predictions of student failure in courses were carried out using three classifiers including DT, NB, and SVM. A detailed account of the prediction model performance was shown in Table 6, while a clearer performance comparison at each minority sample rate was shown in Fig. 3 to Fig. 5.

In this study, the proposed technique was the top-ranking method, as shown in Table 5. The table showed that the technique had the lowest average ranking among oversampling techniques, indicating superior performance compared to others. The ranking was determined by evaluating each oversampling technique's performance across various classifiers and minority rates.

Hypothesis testing results, using the Friedman Test on five performance metrics, showed a difference in mean performance values among oversampling techniques. Consequently, multiple comparisons were conducted using the Holm Post Hoc Test to identify the oversampling technique rejecting the hypothesis of equal mean performance values. In this test, the p-value of each pair of oversampling techniques was compared with that of Holm adjusted p-value.

TABLE 5. Performance comparison for the five oversampling technique and three classifiers on minority sample rate 30%, 50%, 70%, and 100%.

Minority Sample Rate	Classifier	Oversampling Technique	Performance Metric				
			Accuracy	Recall	Precision	F-measure	AUC
30%	DT	Origin	**92.81**	26.72	51.35	33.76	62.42
		SMOTE-N	87.83	75.5	74.31	74.73	83.57
		SMOTE-ENC	86.89	68.4	74.9	71.32	80.57
		ROS	87.48	67.16	77.74	71.71	80.5
		GLoW-SMOTE-D	92.1	**76.23**	**88.48**	**81.73**	**86.52**
	NB	Origin	**89.64**	39.93	32.92	35.63	66.75
		SMOTE-N	81.05	59.01	60.41	59.64	73.42
		SMOTE-ENC	80.66	49.74	61.53	54.96	69.99
		ROS	80.4	49	61.12	54.24	69.63
		GLoW-SMOTE-D	88.66	**79.42**	**73.65**	**76.36**	**85.44**
	SVM	Origin	**92.73**	0	0	0	50
		SMOTE-N	84.72	51.36	77.14	61.58	73.23
		SMOTE-ENC	84.94	51.11	78.06	61.55	73.29
		ROS	83.06	42.49	76.14	54.45	69.13
		GLoW-SMOTE-D	89.9	**70.96**	**83.19**	**76.42**	**83.28**
50%	DT	Origin	**92.81**	26.72	51.35	33.76	62.42
		SMOTE-N	87.86	**84.71**	78.88	81.63	87.02
		SMOTE-ENC	86.81	80.57	78.68	79.57	85.17
		ROS	84.99	79.45	75.25	77.14	83.57
		GLoW-SMOTE-D	92.18	82.86	**92.97**	**87.56**	**89.85**
	NB	Origin	**89.64**	39.93	32.92	35.63	66.75
		SMOTE-N	78.29	65.3	66.29	65.76	74.82
		SMOTE-ENC	78.61	56.06	71.05	62.62	72.65
		ROS	76.46	49.77	68.14	57.45	69.39
		GLoW-SMOTE-D	88.66	**83.55**	**82.64**	**83.05**	**87.38**
	SVM	Origin	**92.73**	0	0	0	50
		SMOTE-N	82.41	63.35	77.55	69.65	77.37
		SMOTE-ENC	82.43	64.5	76.97	70.1	77.72
		ROS	80.78	59.48	75.18	66.36	75.15
		GLoW-SMOTE-D	90.11	**81.82**	**87.66**	**84.58**	**88.04**
70%	DT	Origin	**92.81**	26.72	51.35	33.76	62.42
		SMOTE-N	88.98	89.8	84.64	87.11	89.07
		SMOTE-ENC	87.85	**90.98**	81.78	86.12	88.31
		ROS	85.48	90.16	78.24	83.69	86.18
		GLoW-SMOTE-D	92.5	88.65	**92.93**	**90.7**	**91.96**
	NB	Origin	**89.64**	39.93	32.92	35.63	66.75
		SMOTE-N	77.46	72.08	72.99	72.52	76.63
		SMOTE-ENC	76.56	61.08	77.62	68.28	74.3
		ROS	73.69	54.41	75.24	63.12	70.87
		GLoW-SMOTE-D	89.24	**84.44**	**89.01**	**86.58**	**88.58**
	SVM	Origin	**92.73**	0	0	0	50
		SMOTE-N	80.93	72.91	79.43	75.88	79.76
		SMOTE-ENC	83.21	79.06	80.13	79.52	82.6
		ROS	80.79	73.32	78.8	75.91	79.72
		GLoW-SMOTE-D	90.75	**86.28**	**90.91**	**88.46**	**90.14**
100%	DT	Origin	92.81	26.72	51.35	33.76	62.42
		SMOTE-N	89.49	92.44	87.46	89.88	89.47
		SMOTE-ENC	88.54	93.31	85.42	89.15	88.5
		ROS	86.58	**94.89**	81.53	87.7	86.48
		GLoW-SMOTE-D	**93.93**	91.89	**95.8**	**93.78**	**93.92**
	NB	Origin	89.64	39.93	32.92	35.63	66.75
		SMOTE-N	75.6	73.92	76.87	75.35	75.61
		SMOTE-ENC	75.41	66.17	81.69	73.09	75.5
		ROS	71.93	59.99	79.42	68.29	72.08
		GLoW-SMOTE-D	**90.43**	**88.62**	**92**	**90.25**	**90.44**
	SVM	Origin	**92.73**	0	0	0	50
		SMOTE-N	82.86	83.83	82.51	83.15	82.85
		SMOTE-ENC	82.29	84.46	81.22	82.8	82.27
		ROS	80.81	81.15	80.89	81.02	80.81
		GLoW-SMOTE-D	92.35	**90.9**	**93.63**	**92.23**	**92.36**

The hypothesis of equal mean performance values rejection occurred if the adjusted p-value was less than alpha. Both hypothesis tests used a 5% alpha level, and the results of the Holm Post Hoc Test were shown in Table 7.

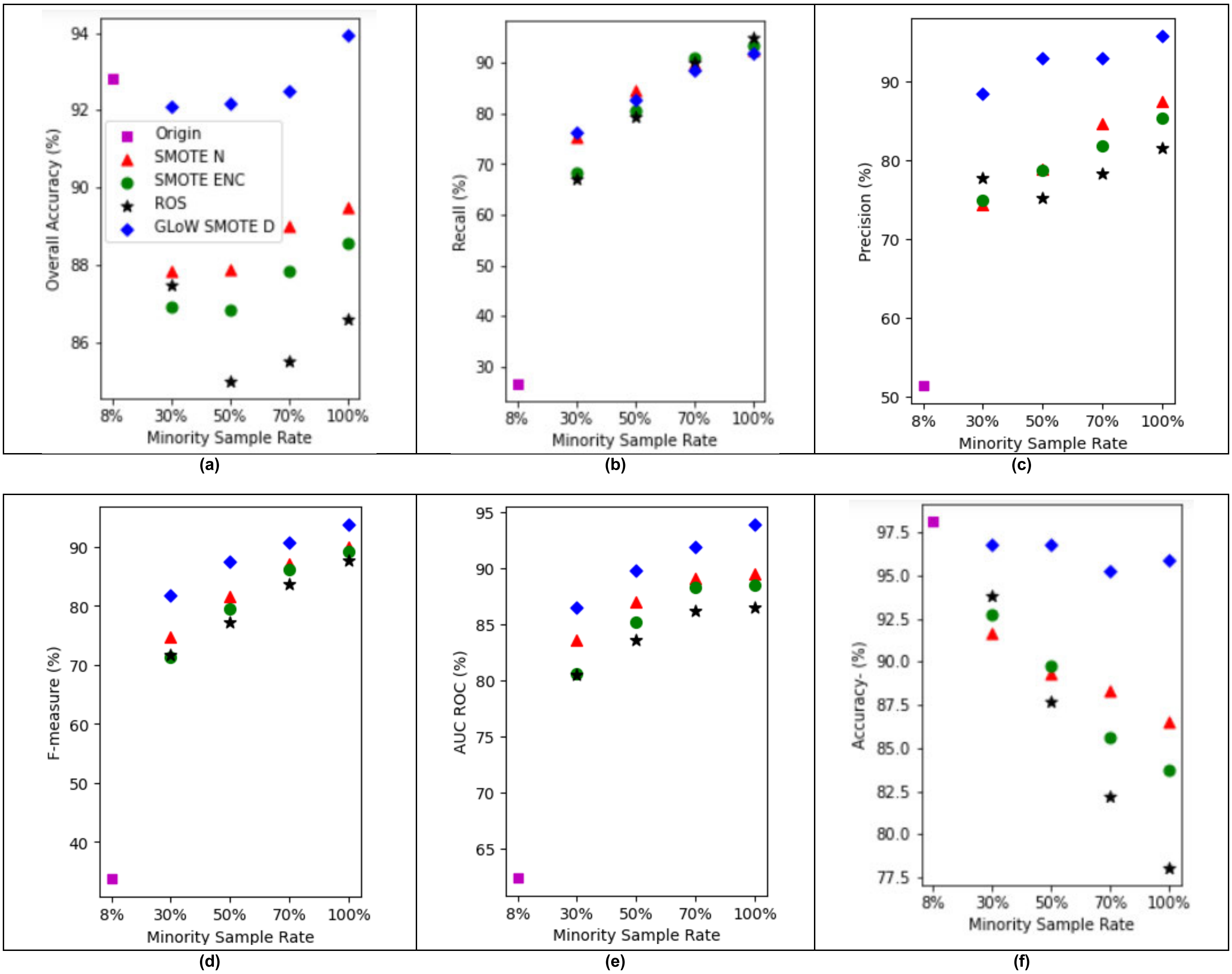

FIGURE 3. Experimental results using DT (a) Accuracy (b) Recall (c) Precision (d) F- measure (e) AUC (f) Acc $^{-}$.

TABLE 6. The average rank obtained by each oversampling technique in the friedman test.

Technique	Ranking				
	Accuracy	Recall	Precision	F-measure	AUC
SMOTE-N	2.33	2.42	3	2.25	2.25
SMOTE-ENC	2.75	2.5	2.5	2.92	2.75
ROS	3.92	3.5	3.5	3.83	4
GLoW SMOTE-D	**1**	**1.58**	**1**	**1**	**1**

E. DISCUSSION

This study showed that the accuracy rates of the original data were high. However, when scrutinized individually for each class, accuracy for the majority class (Acc^{-}) was exceedingly high, with some even achieving 100%, and the minority class (Acc^{+}) was low, with select results even reaching zero (Fig. 3a to Fig. 5a and Fig. 3f to Fig. 5f). This observation showed the robust learning abilities of the majority samples, compared with the challenges faced by the minority samples. As the minority sample rate increased, there was a tendency for the learning abilities of the minority samples to improve. However, this often coincided with a decline in

TABLE 7. Holm post Hoc test results with GLoW SMOTE-D as control technique.

Metric	Technique	P-value	Adjusted p-value	Hypothesis
Accuracy	SMOTE-N	0.0000112	0.0000448	Reject
	SMOTE-ENC	0.0000044	0.0000267	Reject
	ROS	0.0000084	0.0000422	Reject
Recall	SMOTE-N	0.0021308	0.0085232	Reject
	SMOTE-ENC	0.0009816	0.0049080	Reject
	ROS	0.0004816	0.0028895	Reject
Precision	SMOTE-N	0.0000001	0.0000004	Reject
	SMOTE-ENC	0.0000004	0.0000017	Reject
	ROS	0.00000003	0.0000002	Reject
F-measure	SMOTE-N	0.0000092	0.0000458	Reject
	SMOTE-ENC	0.0000212	0.0000849	Reject
	ROS	0.0000071	0.0000426	Reject
AUC	SMOTE-N	0.0000207	0.0000829	Reject
	SMOTE-ENC	0.0000051	0.0000254	Reject
	ROS	0.0000015	0.0000092	Reject

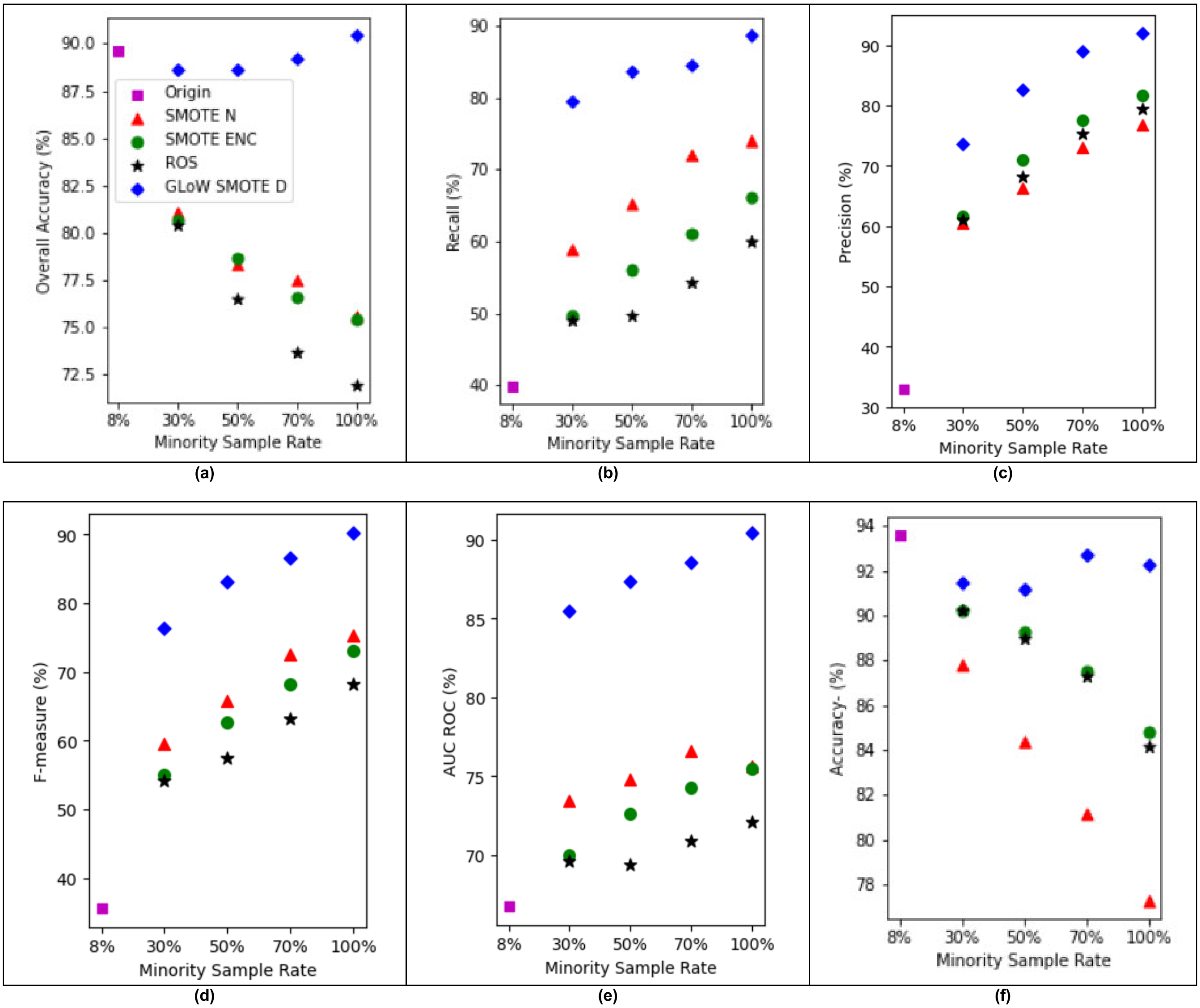

FIGURE 4. **Experimental results using NB (a) Accuracy (b) Recall (c) Precision (d) F- measure (e) AUC (f) Acc $^-$.**

the learning abilities of the majority samples. These findings validated and strengthened previous studies conducted by Liu et al. [21]. The inconsistency observed could be attributed to the inclusion of certain synthetic minority samples as noise within the majority class, leading to misclassification. On the other hand, the incorporation of a noise selection process and feature selection criteria during the creation of synthetic minority samples mitigated the noise's impact on the majority domain. This presented a distinct advantage to GLoW SMOTE-D, curbing a high decline in accuracy for the majority samples when compared to ROS, SMOTE-N, and SMOTE-ENC as shown in Fig. 3f to Fig. 5f.

The oversampling technique developed was initiated by selecting positive samples with limited learning abilities for training. Specifically, the training consisted of assigning weights based on the correlation level with the target class within the nearest neighbor group. This accelerated the augmentation of the minority samples' learning capacities in comparison to the alternative oversampling technique, a phenomenon borne out by the recall, precision, F-measure, and AUC values in Fig. 3 to Fig. 5. The expansion of the decision area for minority samples improved heterogeneity and reduced noise, supporting performance of the predictive model. However, ROS omitted consideration of such expansion, terminating in diminished recall, precision, F-measure, and AUC values.

In the original dataset, DT exhibited pre-eminence in terms of accuracy and precision but in other performance measures, NB took the lead. According to the application of oversampling with varied techniques and minority sample rates, DT consistently outperformed NB and SVM.

Performance trends shown by DT, NB, and SVM were the same and the original technique produced the highest accuracy. As the minority sample rate increased to 30%, accu-

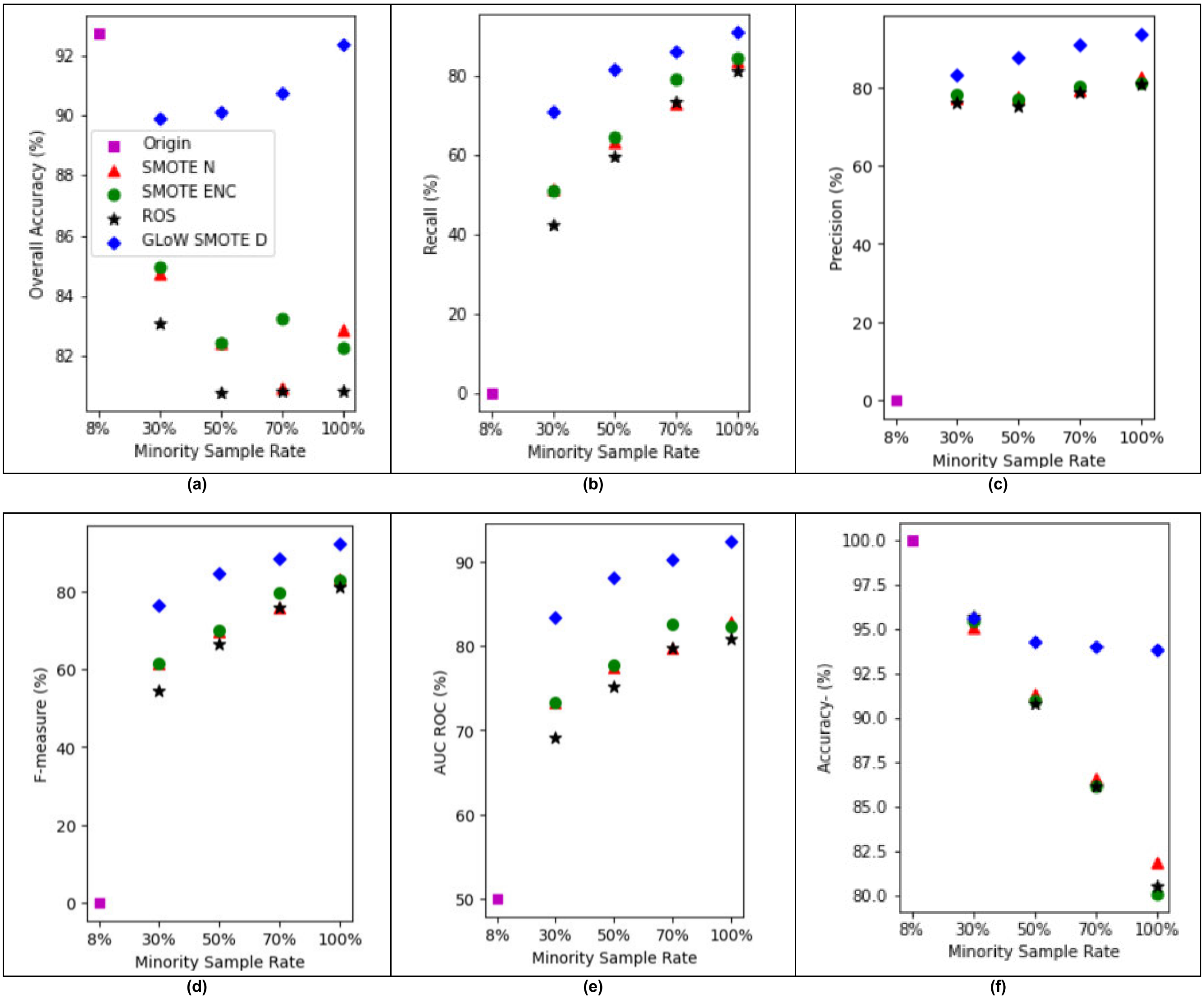

FIGURE 5. Experimental results using SVM (a) Accuracy (b) Recall (c) Precision (d) F- measure (e) AUC (f) Acc $^{-}$.

racy exhibited a decrement, but it subsequently exhibited an ascending trajectory with further increments in the minority sample rate. However, recall, precision, F-measure, and AUC values were at their lowest for the original technique. After enhancing the minority sample rates, there was a tendency for performance values to increase.

Further statistical analysis showed the dominance of GLoW SMOTE-D performance over other techniques. GLoW SMOTE-D secured the top ranking in all performance metrics, as seen in Table 5. Despite the recall ranking in Table 5 showing occasional less dominance in some classifiers and minority rates, the recall ranking of GLoW SMOTE-D still outperformed other oversampling techniques. This conclusion was supported by the Holm Post Hoc Test results, showing a significant difference in mean performance between GLoW SMOTE-D and other techniques across all metrics.

Enhancing the performance of minority samples is crucial in effectively predicting students' failure in courses. As more individuals at risk of failing courses became identifiable, accurate prediction offered a valuable window to reduce failure rates. This was achieved through proactive measures taken by both students and lecturers. However, even a slight decline in the performance of the majority samples could also yield benefits. When those who had the potential to succeed were predicted to fail, it opened up opportunities for them to improve their abilities. These individuals could receive guidance and support from their lecturers or take initiative. This led to an increase in the percentage of students passing their courses and an improvement in academic performance. On the other hand, it was crucial to emphasize that all of these positive outcomes depended on strong initiatives and cooperation between students and lecturers.

V. CONCLUSION

This study aimed to enhance the performance of students' failure prediction models in courses. This improvement was achieved through the introduction of a novel oversampling technique, GLoW SMOTE-D, specifically designed to address class-imbalanced datasets. The techniques were carried out in three crucial stages, namely the extraction of the misclassified minority sample set, the computation of global weights, and the execution of synthetic minority oversampling. This selection of minor samples to be oversampled and weighted enhanced the learning capabilities among the minority samples while minimizing noise infiltration into the majority area. Moreover, when synthetic data was generated using GLoW SMOTE-D, it expanded the decision boundaries for minority samples, leading to increased diversity, and decreased noise. As a result, GLoW SMOTE-D exhibited superior performance in terms of metrics such as accuracy, recall, precision, F-measure, AUC, and Acc^{-} when compared to ROS, SMOTE N, and SMOTE ENC across different classifiers.

Further analysis revealed that, under ideal conditions, the identification of more students at risk of predictable failure enhanced the total pass rates. The misidentification of students with the potential to succeed and the provision of support contributed to improved academic performance.

The proposed oversampling technique focused primarily on discrete features and comprised higher computational costs compared to alternative techniques. This implies that there is still an opportunity to develop the proposed oversampling technique in the future. Opportunities for improvement in the Misclassified Minority Sample Set search process were identified. Currently, this set is generated using the DT method. In future research, several classification algorithms will be integrated to mitigate the variance introduced by the DT method Additionally, the time-consuming weight calculation process presented an avenue for developing a more efficient weighting technique.

REFERENCES

[1] O. Daramola, O. Emebo, I. Afolabi, and C. Ayo, "Implementation of an intelligent course advisory expert system," *Int. J. Adv. Res. Artif. Intell.*, vol. 3, no. 5, pp. 6–12, May 2014, doi: 10.14569/ijarai.2014.030502.

[2] I. Khan, A. R. Ahmad, N. Jabeur, and M. N. Mahdi, "An artificial intelligence approach to monitor student performance and devise preventive measures," *Smart Learn. Environ.*, vol. 8, no. 1, p. 27, Sep. 2021, doi: 10.1186/s40561-021-00161-y.

[3] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: 10.1109/ACCESS.2020.2981905.

[4] L. M. A. Zohair, "Prediction of student's performance by modelling small dataset size," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, p. 18, Aug. 2019, doi: 10.1186/s41239-019-0160-3.

[5] H. Han, W. Wang, and B. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput. (ICIC)*, Hefei, China: Springer, 2005, pp. 878–887, doi: https://doi.org/10.1007/11538059_91.

[6] V. S. Spelmen and R. Porkodi, "A review on handling imbalanced data," in *Proc. Int. Conf. Current Trends Converging Technol. (ICCTCT)*. Coimbatore, India: IEEE, Mar. 2018, pp. 1–11, doi: 10.1109/ICCTCT.2018.8551020.

[7] N. K. Mishra and P. K. Singh, "Feature construction and smote-based imbalance handling for multi-label learning," *Inf. Sci.*, vol. 563, pp. 342–357, Jul. 2021, doi: 10.1016/j.ins.2021.03.001.

[8] T. Fahrudin, J. L. Buliali, and C. Fatichah, "Enhancing the performance of smote algorithm by using attribute weighting scheme and new selective sampling method for imbalanced data set," *Int. J. Innov. Comput. Inf. Control*, vol. 15, no. 2, pp. 423–444, Apr. 2019, doi: 10.24507/ijicic.15.02.423.

[9] 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, doi: 10.1613/jair.953.

[10] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-Level-SMOTE: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *Proc. 13th Pacific–Asia Conf. Knowl. Discovery Data Mining (PAKDD)*, Bangkok, Thailand: Springer, Apr. 2009, pp. 475–482, doi: 10.1007/978-3-642-01307-2_43.

[11] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw., IEEE World Congr. Comput. Intell.* Hong Kong: IEEE, Jun. 2008, pp. 1322–1328, doi: 10.1109/IJCNN.2008.4633969.

[12] S. Cinaroglu, "The impact of oversampling with 'ubSMOTE' on the performance of machine learning classifiers in prediction of catastrophic health expenditures," *Oper. Res. Health Care*, vol. 27, Dec. 2020, Art. no. 100275, doi: 10.1016/j.orhc.2020.100275.

[13] M. R. Ayyagari, "Classification of imbalanced datasets using one-class SVM, k-Nearest neighbors and CART algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, pp. 1–5, Nov. 2020, doi: 10.14569/IJACSA.2020.0111101.

[14] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining*, 2nd ed. New York, NY, USA: Pearson Education, 2019.

[15] M. Roknizadeh and H. M. Naeen, "Hybrid ensemble optimized algorithm based on genetic programming for imbalanced data classification," 2021, *arXiv:2106.01176*.

[16] K. Li, B. Ren, T. Guan, J. Wang, J. Yu, K. Wang, and J. Huang, "A hybrid cluster-borderline SMOTE method for imbalanced data of rock groutability classification," *Bull. Eng. Geol. Environ.*, vol. 81, no. 1, pp. 1–15, Jan. 2022, doi: 10.1007/s10064-021-02523-9.

[17] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.

[18] S. T. Jishan, R. I. Rashu, 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, Mar. 2015, doi: 10.1186/s40165-014-0010-2.

[19] A. Skowron and A. Wojna, "*K* nearest neighbor classification with local induction of the simple value difference metric," in *Proc. Int. Conf. Rough Sets Current Trends Comput. (RSCTC)*. Uppsala, Sweden: Springer, 2004, pp. 229–234, doi: 10.1007/978-3-540-25929-9_27.

[20] G. Akçapınar, A. Altun, and P. Aşkar, "Using learning analytics to develop early-warning system for at-risk students," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, p. 40, Oct. 2019, doi: 10.1186/s41239-019-0172-z.

[21] C. Liu, S. Jin, D. Wang, Z. Luo, J. Yu, B. Zhou, and C. Yang, "Constrained oversampling: An oversampling approach to reduce noise generation in imbalanced datasets with class overlapping," *IEEE Access*, vol. 10, pp. 91452–91465 2022, doi: 10.1109/ACCESS.2020.3018911.

[22] A. Kirshners, S. Parshutin, and H. Gorskis, "Entropy-based classifier enhancement to handle imbalanced class problem," *Proc. Comput. Sci.*, vol. 104, pp. 586–591, Jan. 2017, doi: 10.1016/j.procs.2017.01.176.

[23] A. S. Hussein, T. Li, C. W. Yohannese, and K. Bashir, "A-SMOTE: A new preprocessing approach for highly imbalanced datasets by improving SMOTE," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 2, p. 1412, Nov. 2019, doi: 10.2991/ijcis.d.191114.002.

[24] M. Mukherjee and M. Khushi, "SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features," *Appl. Syst. Innov.*, vol. 4, no. 1, p. 18, Mar. 2021, doi: 10.3390/asi4010018.

[25] A. Wojna, "Analogy-based reasoning in classifier construction," Ph.D. dissertation, Dept. Math. Inform. Mech., Warsaw Univ., Warsaw, Poland, 2016. [Online]. Available: https://www.researchgate.net/publication/238650300

[26] E. Tjandra, S. S. Kusumawardani, and R. Ferdiana, "Student performance prediction in higher education: A comprehensive review," in *Proc. Int. Conf. Informat., Technol., Eng. (InCITE)*. New York, NY, USA: AIP, 2022, pp. 1–9, doi: 10.1063/5.0080187.

[27] E. Alyahyan and D. Düştegör, "Predicting academic success in higher education: Literature review and best practices," *Int. J. Educ. Technol. Higher Educ.*, vol. 17, no. 1, pp. 1–21, Feb. 2020, doi: 10.1186/s41239-020-0177-7.

SUSANA LIMANTO received the S.T. degree from Universitas Surabaya, Indonesia, in 1995, and the M.Sc. degree from Institut Teknologi Sepuluh Nopember (ITS), Indonesia, in 2005, where she is currently pursuing the Ph.D. degree with the Department of Informatics. She joined Universitas Surabaya, in 1997, where she is currently a Lecturer with the Informatics Engineering Study Program. Her primary research interests include information systems, machine learning, and optimization.

JOKO LIANTO BULIALI (Member, IEEE) received the bachelor's degree from Institut Teknologi Sepuluh Nopember (ITS), Indonesia, in 1991, and the M.Sc. and Ph.D. degrees from the University of Manchester Institute of Science and Technology (UMIST), England, in 1995 and 1998, respectively. He is currently a Lecturer with the Informatics Department, ITS. His primary research interests include modeling and simulation, optimization, and data analytics. He actively engages in informatics research and serves as a reviewer for several journals and conferences. He is an active member of ACM and Indonesian Association of Higher Education in Informatics and Computing.

AHMAD SAIKHU received the bachelor's degree from the Statistics Department, Institut Teknologi Sepuluh Nopember (ITS), Indonesia, in 1994, the M.T. degree from the Informatics Department, ITS, in 2000, and the Ph.D. degree from the Computer Science Department, ITS, in 2019. He has been a Lecturer with the Informatics Department, ITS, since 2006. His primary research interests include data mining and time series. He actively engages in informatics research and serves as a reviewer for several journals and conferences.