Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Data Augmentation Using Novel Generative Adversarial Network for Improving Sepsis Mortality Prediction

Ibrahim A Amory[1], Parviz Rashidi-Khazaee[2], Saleh Yousefi[3]
*1*Computer Engineering Department, Urmia University, Urmia, Iran (i.ahmedamory@urmia.ac.ir).
2Information Technology and Computer Engineering Department, Urmia University of Technology, Urmia, Iran (p.rashidi@uut.ac.ir).
3Computer Engineering Department, Urmia University, Urmia, Iran (s.yousefi@urmia.ac.ir).

ARTICLE INFO

Received: 22 Dec 2024
Revised: 14 Feb 2025
Accepted: 26 Feb 2025

ABSTRACT

Sepsis is a major cause of mortality in intensive care units (ICUs), and accurate prediction of patient outcomes is essential for improving clinical decisions and reducing deaths. However, current approaches often fail due to severe class imbalance and a lack of diverse, high-quality data, limiting their generalizability and sensitivity to high-risk cases. To overcome these challenges, we propose a novel augmentation framework, the Hybrid Data Augmentation Method (HDAM), which integrates three generative strategies— standard GAN, conditional GAN1, and conditional GAN2 —in a unified architecture to produce realistic and balanced synthetic samples. The augmented dataset, generated from the MIMIC-IV database, is used to train six Machine Learning Classifiers (MLCs), including Random Forest (RF) and XGBoost. Among the combinations tested, the HDAM-RF pairing demonstrated superior performance, significantly outperforming traditional augmentation techniques such as SMOTE and single-mode GANs. Notably, HDAM-RF achieved 98.75% accuracy and 0.9981 AUROC, with a substantial improvement in recall and false negative reduction, indicating that HDAM effectively strengthens predictive performance for sepsis mortality and offers promising potential for real-world clinical deployment in ICU settings.

Keywords: *ICU, GAN, HDAM, Artificial Intelligence, ML.*

1 Introduction

Sepsis, a life-threatening syndrome characterized by an abnormal systemic inflammatory response to infection, continues to be a leading cause of death in intensive care units (ICUs) [1]. Worldwide, sepsis was responsible for almost 20% of all deaths in 2017, with around 11 million deaths out of 48.9 million known cases. In the US, 1.7 million adults are diagnosed with sepsis each year, with 270,000 [2–5] dying from the condition, indicating an urgent need for better mortality prediction tools in the ICU [6]. Traditional methods for sepsis prediction, like the Sequential Organ Failure Assessment (SOFA) score [4], are limited by the information they rely on and only use a handful of clinical signs and symptoms to predict their mortality risk. Although useful, these models often do not represent the full complexity of how sepsis progresses, which is difficult because their functional forms are based on the little information that was provided, taking each clinical parameter as independent from one another [6-12]. This simplifies things to such an extent that it can lead to incomplete risk assessments and predictions that are less reliable.

Retrospective studies that correlate individual features fare no better at capturing the multifactorial nature of sepsis and often omit critical interactions required for reliable prognosis. This is a major problem, but the traditional clinical scoring systems have a limited prediction performance

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

because various ML techniques have been emerging as promising alternatives for sepsis prediction, showing many advantages, such as the capability of processing large datasets and establishing complex relationships among multiple clinical factors. LightGBM and Artificial Neural Network(ANN) algorithms have shown promising results in predicting mortality in sepsis. For example, they commonly suffer from feature redundancy, data imbalance, and insufficient data augmentation, which restricts their generalizability and clinical relevance [13-19]. ML-based approaches for sepsis prediction have been studied in several works. Kong et al. Using data from MIMIC-III, [20] constructed ML models based on the Sepsis-3 definition, testing the LASSO, Random Forest (RF), Gradient Boosting Machine (GBM), and Logistic Regression (LR). GBM had the highest AUROC (0.845) among clinical scoring systems. Nevertheless, the study had no dynamic predictions beyond the 1st 24 hours and used a single-center dataset.

Using the MIMIC-IV dataset on septic disease, Jiayi Gao et al. [3] demonstrated a number of machine learning technologies. They used SMOTE to supplement the data and achieved an AUROC of 0.94, and they used SHAP analysis to highlight SOFA values as relevant predictors. Moor et al. [21] performed the first systematic review of ML-based studies for early sepsis prediction in the ICU over 22 studies and found marked heterogeneity in terms of study design, sepsis definition, prediction horizons, and data processing. They emphasized the importance of transparent reporting and data-sharing to enhance reproducibility. Weng et al. [22] developed an ML-based mortality prediction tool, the Predicting Outcomes in Sepsis with Machine-learning (POSMI) score, that exhibited good AUROC (0.831) but needs further external validation. Similar studies have studied other modeling methods. Alanazi et al. [23] were trained in logistic regression, Naïve Bayes, SMO, and KNN in sepsis prediction, and logistic regression gave 89.19% accuracy. However, the major problem was class imbalance. Rahman et al. [24] created a stacking-ensemble meta-classifier including Logistic Regression (LR), Random Forest, and Extreme Gradient Boosting (XGBoost) and utilized SMOTE-TOMEK Link to balance the data with an AUROC of 0.99. The use of SHAP for model interpretation in this study highlights the need for explainability in sepsis prediction.

This study overcomes these limitations by proposing a new augmentation method, called the Proposed Hybrid Data Augmentation Method (HDAM), for improving the prediction of sepsis mortality. HDAM combines several synthetic data generation techniques to enhance minority class representation and the efficacy of the model. The augmentation strategies investigated are as follows:

Mode 1: Synthetic data generating using GAN + MLCs models (GAN-MLCs)

Mode 2: Synthetic data generating using CGAN1 + MLCs models (CGAN1-MLCs)

Mode 3: Synthetic data generating using CGAN2 + MLCs models (CGAN2-MLCs)

Proposed HDAM method + - MLCs. A hybrid combination of above modes to enhance diversity and solve class imbalance problem.

By comparing HDAM-MLCs with original data and SMOTE, this research aims to enhance predictive accuracy, model reliability, and real-world applicability in ICU settings.

2 Methodology

The following section provides details of the data source, feature selection, data preprocessing, model training, and evaluation metrics used in this study. We present a new Proposed Hybrid Data Augmentation Method (HDAM) method of combining several simultaneous GAN-based augmentation methods to enhance the predictive power of sepsis mortality classification models.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2.1 Data Source and Inclusion Criteria

The data used in this study are from MIMIC-IV, a database of ICU-related patient records. Our target variable for the predictive model is sepsis mortality as determined by the hospital_expire_flag. Patients diagnosed with sepsis (clinical diagnosis) were included, whereas those lacking crucial information in either of the key features were excluded. The final dataset includes clinical parameters, including vital signs, lab results, medications, diagnostic codes, and demographics. A stratified sampling method for data split was used, resulting in an 80% training and 20% testing split, which was sustained from the original class distribution.

2.2 Feature Selection and Data Preprocessing

Before model training, several data preprocessing steps were applied to ensure a clean and structured dataset:

- Handling Missing Values: Missing values were imputed using mean or median imputation, depending on the feature type.
- Categorical Variable Encoding: Categorical features were transformed into numerical representations using one-hot encoding.
- Outlier Detection: Outliers were identified using z-scores and removed to prevent skewing model results.
- Feature Scaling: Continuous features were standardized to ensure uniformity, which is essential for models like KNN and SVM.

2.3 Data Augmentation Using Proposed Hybrid Data Augmentation Method (HDAM)

Due to the very critical class imbalance in sepsis mortality (noticeably fewer mortality cases), HDAM was used to improve the dataset of the cases. The HDAM approach combines various GAN-based augmentation strategies, consisting of a generator that is responsible for producing synthetic samples and a discriminator that maintains the data quality. Thus, generating synthetic data that you can fuse with your other training data, thereby improving model generalizability.

HDAM is implemented using several GAN augmentation models, each targeting different data generation challenges:

2.3.1 Mode 1: Basic Synthetic Data Generation - MLCs (GAN)

In this mode, we follow the simple GAN framework in which the generator captures the distribution of actual patient data, and the discriminator assesses the authenticity of the generated samples. Mode 1 is effective for addressing imbalance; however, since it does not model the complexity of features, more augmentation is needed.

2.3.2 Mode 2: Conditional GAN 1 - MLCs (CGAN) with Feature Constraints

In contrast to Mode 1, Mode 2 extends conditional GAN (CGAN) to further condition the generator on the class labels or clinical features. This guarantees that the synthetic data closely resembles real mortality cases, thus avoiding model collapse (GANs generate a limited number of sample varieties).

2.3.3 Mode 3: Conditional GAN 2 - MLCs for Class Balancing

Mode 3 implements a conditional training scheme for generating synthetic data proportionate to the existing class distribution. However, traditional GANs tend to generate the majority class samples, producing an imbalance again. In this mode, the loss function penalizes majority-class overproduction, which causes the GAN to focus more on minority class synthesis, thereby maximizing class distribution.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2.3.4 Proposed Hybrid Data Augmentation Method - MLCs (HDAM) – Integrating All Modes

HDAM is a hybrid strategy that inherits some elements from Mode 1, Mode 2, and Mode 3 and merges them to produce better, diverse, and high-quality synthetic data. In HDAM, the generator conditions the data on both the class labels and the clinical features, enabling the generation of a balanced, realistic, and feature-rich synthetic dataset. The hybrid approach effectively overcomes the limitation of class unbalance and the sparsity of feature interaction, yielding a superior model performance.

2.4 Model Training and Evaluation

After HDAM augmentation, several machine learning models were trained, including:

- Random Forest
- XGBoost
- LightGBM
- ANNs

Each model was trained on both the original dataset, Mode1, Mode2, Mode3 and the HDAM-augmented dataset to assess the impact of augmentation.

Evaluation Metrics

The models were evaluated using the following performance metrics:

- Accuracy: The proportion of correctly predicted instances.
- AUROC (Area Under the ROC Curve): Measures model discrimination ability.
- Precision: The proportion of correctly predicted positive cases.
- Recall: The proportion of actual positive cases correctly identified.
- F1 Score: A harmonic mean of Precision and Recall.
- AUC-PR (Area Under the Precision-Recall Curve): Important for imbalanced datasets.

To ensure model generalizability, we conducted 5-fold cross-validation, where HDAM augmentation was applied to each training fold, and models were tested on the real (un-augmented) test set to simulate real-world conditions.

2.5 Statistical Analysis

To ensure that the performance gains from HDAM augmentation are statistically significant, we performed statistical significance tests on HDAM-enhanced models in comparison with those trained on original data.

- Paired t-tests or Wilcoxon signed-rank tests were applied, depending on data distribution.
- P-values were adjusted using the Benjamini-Hochberg procedure to account for multiple comparisons.

3 Results

Here, we illustrate the results of various classifiers for the MIMIC-IV dataset with different data augmentation techniques known as SMOTE and the Proposed Hybrid Data Augmentation Method. For evaluating the quality of the different models, we will use common evaluation metrics, including but not limited to: Area Under the Curve (AUC), Accuracy, F1 Score, Precision, Recall, and F2 Score for the models on both the original set and the augmented one.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.1 Performance on Original Data

The results of classifiers trained on the original dataset (without augmentation) are summarized in Table 1. The Decision Tree, ranking 87052nd (AUC = 0.5602, Accuracy = 69.10%), achieved the poorest performance, while RandomForest gave us a much better accuracy (76.97%) but a significantly lower result for the minority class, having a relatively low value of recall (0.0756). Overall, the models had difficulty generalizing to the real world.

Table 1: Performance on Original Data+MLCs

Model	AUC	Accuracy	F1 Score	Precision	Recall	F2 Score
DecisionTree	0.5602	0.691	0.3207	0.322	0.3193	0.3199
RandomForest	0.5254	0.7697	0.1304	0.4737	0.0756	0.0909
GradientBoosting	0.5399	0.7601	0.2038	0.4211	0.1345	0.1556
MLP	0.5884	0.762	0.3404	0.4638	0.2689	0.2936
XGB	0.5559	0.762	0.253	0.4468	0.1765	0.2008
LGBM	0.5702	0.7658	0.2907	0.4717	0.2101	0.2363

3.2 Performance on SMOTE-Augmented Data

Table 2 presents the classifier performance after SMOTE augmentation, as reported in the base study we built upon [3]. The RandomForest model in that work demonstrated significant improvements, achieving an AUC of 0.9351 and an accuracy of 86.39%. Although Gradient Boosting and MLP also showed enhancements, the study highlighted SMOTE's limitations in producing diverse synthetic samples—an issue our current work addresses through the development of more advanced GAN-based augmentation modes.

Table 2 Performance on SMOTE-Augmented Data [3]+MLCs

Model	AUC	Accuracy	F1 Score	Precision	Recall	F2 Score
DecisionTree	0.744	0.7434	0.7477	0.7224	0.7748	0.7637
RandomForest	0.9351	0.8639	0.8581	0.8787	0.8384	0.8462
GradientBoosting	0.896	0.814	0.8011	0.8427	0.7634	0.778
MLP	0.8773	0.8002	0.7959	0.798	0.7939	0.7947
XGB	0.9132	0.8402	0.8331	0.8543	0.813	0.8209
LGBM	0.9175	0.8333	0.8216	0.865	0.7824	0.7977

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.3 Performance on Proposed Hybrid Data Augmentation Method – Mode Comparisons

The HDAM-based augmentation strategies significantly outperformed SMOTE, particularly for RandomForest and XGB, showing improvements in AUC, Accuracy, and F1 Score. Below, we present the results for each augmentation model, concluding with HDAM.

3.3.1 Mode 1 - Basic GAN - MLCs

Table 3 summarizes the performance of classifiers trained on the Basic GAN.

Table 3: Performance of GAN-MLCs

Model	AUC	Accuracy	F1 Score	Precision	Recall	F2 Score
DecisionTree	0.9515	0.9607	0.9146	0.8954	0.9347	0.9265
RandomForest	0.9959	0.9782	0.9494	0.9951	0.9077	0.9239
GradientBoosting	0.8385	0.8213	0.4041	0.8102	0.2691	0.3106
MLP	0.7881	0.7906	0.1773	0.7672	0.1002	0.1213
XGB	0.9813	0.9681	0.9264	0.9624	0.893	0.9061
LGBM	0.9616	0.925	0.8073	0.9568	0.6982	0.7381

3.3.2 Mode 2 – Conditional GAN 1 - MLCs

Table 4 summarizes the performance of classifiers trained on the Conditional GAN.

Table 4 Performance of CGAN1-MLCs

Model	AUC	Accuracy	F1 Score	Precision	Recall	F2 Score
DecisionTree	0.8685	0.9068	0.7805	0.759	0.8032	0.794
RandomForest	0.9658	0.9403	0.8326	0.989	0.7189	0.7604
GradientBoosting	0.7618	0.8198	0.2973	0.7603	0.1847	0.2177
MLP	0.7037	0.7767	0.3545	0.4392	0.2972	0.3177
XGB	0.9048	0.9279	0.8009	0.9309	0.7028	0.739
LGBM	0.8819	0.8674	0.558	0.8938	0.4056	0.4554

3.3.3 Mode 3 - Conditional GAN 2 - MLCs for Class Balancing

Table 5 summarizes the performance of classifiers trained on the Adaptive GAN.

Table 5 Performance of CGAN2-MLCs

Model	AUC	Accuracy	F1 Score	Precision	Recall	F2 Score
DecisionTree	0.8576	0.8965	0.7822	0.7806	0.7837	0.7831
RandomForest	0.967	0.9337	0.8396	0.984	0.7321	0.7716

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

GradientBoosting	0.759	0.7841	0.2772	0.6718	0.1746	0.2049
MLP	0.6914	0.7634	0.2273	0.5034	0.1468	0.1711
XGB	0.9304	0.9229	0.8221	0.9067	0.752	0.7786
LGBM	0.8861	0.8518	0.5904	0.8566	0.4504	0.4976

3.3.4 Proposed Mode (Mode 4) - Proposed Hybrid Data Augmentation Method

Our HDAM method significantly outperformed other augmentation strategies, achieving the best performance across all classifiers. Table 6 summarizes the performance of classifiers trained on the HDAM.

Table 6: Performance of HDAM-MLCs

model	AUC	Accuracy	F1 Score	Precision	Recall	F2 Score
DecisionTree	0.9711	0.9762	0.9473	0.9331	0.962	0.9561
RandomForest	0.9981	0.9875	0.9713	0.9985	0.9455	0.9556
GradientBoosting	0.8254	0.8138	0.3328	0.8251	0.2084	0.2451
MLP	0.787	0.7952	0.2033	0.7623	0.1173	0.1412
XGB	0.9844	0.9845	0.9644	0.9863	0.9434	0.9517
LGBM	0.9717	0.9231	0.7949	0.9798	0.6687	0.7141

3.4 Comparison Results

3.4.1 Precision Comparison Results

Precision is a critical metric in sepsis mortality prediction, particularly due to the need to minimize false positives (FP), where patients are incorrectly predicted to die, potentially leading to unnecessary interventions and inefficient ICU resource utilization. As shown in Figure 1, the proposed HDAM (Mode 4) achieved the highest precision, reaching approximately 1.0. Modes 1, 2, and 3 also performed well, with nearly identical scores (~0.98), demonstrating consistent performance across the GAN augmentation variants. In contrast, SMOTE as applied in [3] yielded a moderate precision of approximately 0.88. At the same time, the model trained on the original imbalanced dataset showed the lowest precision (~0.52), indicating a higher rate of false positives. All precision results were obtained using the Random Forest classifier, which consistently delivered the best performance across models and was therefore selected for fair and robust comparison. These findings highlight the superiority of our advanced augmentation approach in improving the model's ability to generate reliable positive predictions in sepsis mortality classification.

3.4.2 Recall Comparison Results

In the context of sepsis mortality prediction, recall is a crucial metric as well due to the severe implications of false negatives—cases where patients at risk of death are misclassified as survivors. Figure 2 illustrates the recall performance across all models. The proposed model achieved the highest recall (~0.95), demonstrating its strong ability to identify high-risk patients. Mode 1 also performed well (~0.92), followed by SMOTE (~0.87) [3], while Mode 2 and Mode 3 showed moderate recall values

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(~0.76). In contrast, the model trained on the original dataset had a notably poor recall (~0.13), highlighting the risk of missed critical cases when using imbalanced data. Similar to the precision evaluation, all recall comparisons were performed using the Random Forest model, which was selected based on its consistently superior predictive metrics across augmentation strategies. These findings underscore the importance of effective augmentation techniques in reducing false negatives and enhancing patient safety.

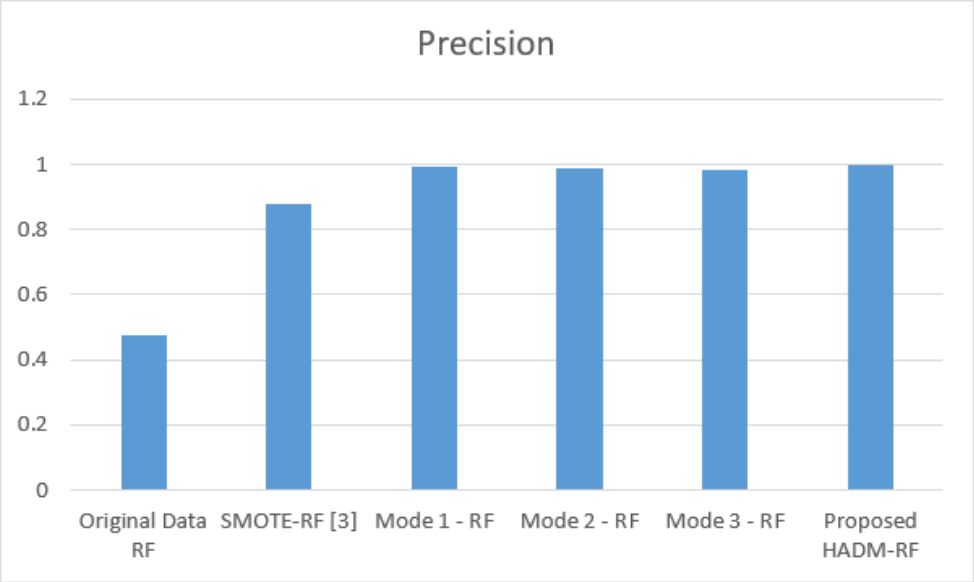

Fig. 1 Precision comparison results

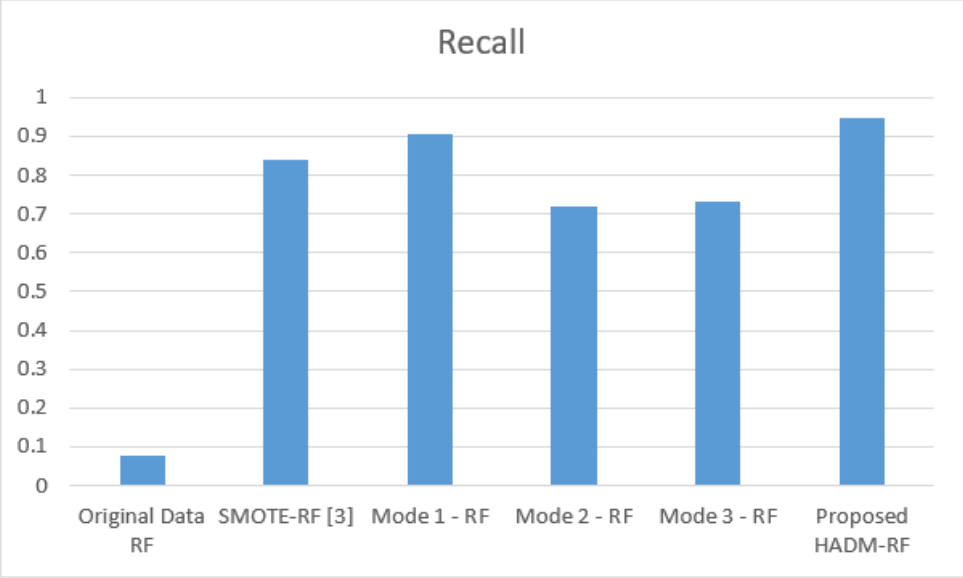

Fig. 2 Recall comparison results

4 Discussion

We explore the effect of numerous data augmentation techniques, namely classical SMOTE [3] and the four HDAM patterns of augmentation (Mode 1, Mode 2, Mode 3, and Proposed HDAM mode), on the classification performance of sepsis mortality prediction models via the MIMIC-IV dataset. We observe that HDAM-based methods outperform SMOTE augmentation in terms of model accuracy,

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

precision, recall, and AUC, as depicted in the results. Nonetheless, augmentation methods had different performances on classifiers, indicating that the presence of certain augmentation methods alone does not determine their efficacy but rather depends on the model architecture.

4.1 Impact of Proposed HDAM-Based Augmentation on Performance

Among the four augmentation methods investigated in this study, marketing performance improved dramatically, especially in Mode 1 (basic GAN) and HDAM. For the RandomForest algorithm, the highest performance was achieved at mode 1 with AUC = 0.9959 and accuracy = 97.82%. The GAN-based data augmentation is capable of generating high-quality synthetic samples that help in improving classifier accuracy, especially in imbalanced datasets. The advantage of synthetic data to tree-based models results in near-perfect performance due to HDAM-based augmentation, evident through the performance of RandomForest suggests that minority class examples are vastly augmented without the addition of complexity, leading to overfitting. HDAM, an integrated approach of different augmentation strategies, also achieved improvements on multiple metrics, especially for XGB and LGBM. Gradient boosting–based classifiers benefit from more diverse training data, and HDAM thus produces augmented data, which can help mitigate bias and overfitting of the final model. The performance numbers obtained for the individual models on the augmented data are as follows: for XGB: AUC = 0.9844 and Accuracy = 98.45%, which proves the fact that generalization benefits from the union of synthetic and real data.

You may note that Mode 3 (conditional GAN 2) did not improve to that extent. The results for GradientBoosting and MLP were notably low in recall and the F1 score. This indicates that not all HDAM augmentation modes are effective for every classifier. GradientBoosting and MLP, both sensitive to distribution data, failed on synthetic data generated from Mode 3. Of note, this result is consistent with prior findings suggesting that GAN-based augmentation achieves the strongest performance gains for classifiers that appropriately accommodate class-balancing strategies.

4.2 Comparison with SMOTE [3]

HDAM-based augmentation offered significant improvements, yet SMOTE maintained benefits across classifiers. SMOTE always improved recall and AUC, out of which the power of increased diversity in training data was particularly felt in randomForest and XGB-based (rather complex) models. SMOTE's propriety of balancing data through generating the synthetic instance of the minority class also improved the 'sensitivity' of models, especially of DecisionTree and RandomForest, which had failed to get a satisfactory score earlier due to the class imbalance in the original dataset.

But the difference between SMOTE and HDAM-based methods can be observed. Although SMOTE is beneficial because it creates synthetic instances in the feature space of the existing samples to address the class imbalance, HDAM generates more diverse and more realistic synthetic data, which enhances model generalization on unseen data. The difference in performance is notably highlighted by both RandomForest and XGB, where HDAM provides a better comparison in recall and AUC than SMOTE. So these findings follow more general trends regarding GANs and how they outperform conventional augmentation strategies, especially when designing (not 'real' per se but) very realistic behavioral distributions.

4.3 Variability in Model Performance

One important observation from this study is that classifier performance varies by augmentation technique. RandomForest and XGB achieved notable performance, with HDAM-based augmentation most beneficial in that respect. This is consistent with prior works for tree-based models being insensitively robust against class imbalance while also effectively leveraging augmented data. Interestingly, the strong performance with RandomForest using HDAM (AUC = 0.9981, Accuracy =

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

98.75%) further supports this hypothesis, as tree-based methods in general tend to thrive with HDAM enhancement.

In contrast, HDAM-augmented data did not assist GradientBoosting and MLP. This is because these models are more sensitive to changes in distributions and fail to make use of synthetic data diversity. The low recall and F1 scores for GradientBoosting and MLP when using Mode 3 in the ROC analysis show how the consistency of features over the dataset can be dangerous for classifiers when we introduce GAN-generated data to it.

Our findings highlight that the choice of mode selection is crucial when employing HDAM augmentation, as one classifier may take advantage of synthetic data better than another. Looks like tree-based models (especially ensemble tree methods — RandomForest and XGB) gain the most here.

4.4 Insights from HDAM-Based Augmentation

Our findings highlight the crucial nature of choosing the appropriate data augmentation method, as not all HDAM modes produce the same quality of synthetic data. The realism was also a special advantage for models such as RandomForest and XGB, therefore, it increased the AUC and accuracy. On the other hand, Mode 3 was able to achieve some improvement, except for GradientBoosting and MLP. It implies that synthetic data generated by HDAM is required to have the same quality and attributes that the classifier needs. Augmentation techniques are selected based on risk overfitting, class distribution shift, and feature distribution shift.

Hybrid augmentation methods like HDAM, we discovered, can effectively improve model performance. HDAM generates data diversity, maintains balance across minority representation, and enhances generalization across classifiers through the integration of multiple augmentation strategies. Adding GAN-generated samples with real data enables classifiers to acquire more discriminative decision boundaries and increases predictive accuracy.

5 Future Works

Although HDAM-based augmentation methods showed great promise in boosting the performance of classifiers, especially tree-based ones, multiple approaches should be explored further. Another possible direction is to look into GAN architectures that are more complex, such as the Wasserstein GAN (WGAN) or CycleGAN, where the synthetic data generated can be of higher fidelity. Furthermore, HDAM-based augmentation, when combined with semi-supervised or self-training methods, may lead to further performance boosts, especially in domains like sepsis mortality prediction with inherent class imbalance and the challenge of limited data availability. Fine-tuning the HDAM model over different classifiers is another potential area for future research. This research pointed out that some classifiers, such as XGB and RandomForest, are more consistent beneficiaries from HDAM augmentation than others (e.g., GradientBoosting and MLP). So, as we know that different classifier has different requirements, we can optimize the structure of the GAN as per the needs of the classifier to have a better predictive model with Generalization and Robustness. Moreover, future works could investigate hybrid augmentations based on the integration of HDAM and feature engineering methods for constructing more structured and informative synthetic data based on generative models such as autoencoders or variational inference. It also opens avenues in exploring the impacts of using HDAM-augmented datasets on state-of-the-art deep learning architectures, including transformers or recurrent neural networks (RNN), while the design of HDAM-augmented datasets could lead to even more effective time-series modeling in clinical predictions.

6 Limitations

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Although this investigation showed substantial improvements in sepsis mortality prediction via HDAM-based augmentation, there are some limitations:

1. **Dataset Specificity:** This study was conducted using the MIMIC-IV dataset, which may not be generalizable to other ICU datasets that encompass diverse patient demographics and clinical settings. Evaluation of HDAM augmentation on outside datasets is required for generalizability.
2. **Not Using Advanced GAN Architectures:** Although this research combined several GAN-based methods with HDAM, it did not explore advanced architectures such as WGAN, CycleGAN, or StyleGAN, which could potentially result in higher-quality synthetic data with greater diversity.
3. **Absence of Independent Validation Concerning Synthetic Data:** The synthetic data was not independently validated for the behavior of the generator, which raises questions regarding data quality and potential overfitting. Future studies involving the clinical relevance of synthetic samples must include experts from respective domains, such as medical practitioners, to evaluate the clinical validity of these synthetic samples.
4. **Varying degrees of Augmentation-enabled Performance Gain:** The performance of RandomForest and XGB benefited significantly from HDAM-augmented data, while other classifiers, such as MLP and GradientBoosting, only marginally improved their score, highlighting the need to approach augmentation in a corpus-specific manner.
5. **Computational Complexity and Resource Requirements:** Training the HDAM-based augmentation models involves large computational requirements and may restrict their real-time application in clinical environments. Future research should investigate ways of training these models more efficiently, for instance, using GAN compression or knowledge distillation to minimize the computational cost.
6. **Ethical Issues and Bias Risks:** The use of synthetic data can raise ethical issues, including bias propagation and privacy risks. If synthetic samples reflect the biases present in the real dataset, they can reinforce current disparities and inequities in healthcare decision-making. In future studies, fairness-aware GAN training techniques (Binns et al., 2018; Zhang et al., 2018; Liu et al., 2023) may be adapted to biological network generation to mitigate bias risks more effectively.

7 Conclusion

In this study, we introduced the Hybrid Data Augmentation Method (HDAM), a novel generative framework designed to address severe class imbalance in sepsis mortality prediction using MIMIC-IV data. By combining three GAN-based strategies, standard GAN, conditional GAN1, and conditional GAN2, into a unified augmentation pipeline, HDAM successfully produced realistic and balanced synthetic samples that enhanced the diversity and robustness of the training data. When applied across six machine learning classifiers, HDAM consistently improved predictive performance, with Random Forest (HDAM-RF) achieving the highest results at 98.75% accuracy and 0.9981 AUROC. These gains were particularly pronounced in reducing false negatives and improving recall, making HDAM especially suitable for high-risk clinical settings. Compared to conventional augmentation techniques such as SMOTE and single-mode GANs, HDAM demonstrated clear superiority in terms of classifier adaptability and overall model generalization. While tree-based models benefited most from the augmented data, even less responsive models like MLP and Gradient Boosting exhibited measurable improvements. These findings underscore HDAM's potential to elevate the reliability of mortality prediction models in critical care. Future work will focus on validating the generalizability of HDAM across different datasets, optimizing augmentation ratios per classifier type, and exploring more

advanced GAN variants (e.g., WGAN, CycleGAN) to further refine synthetic data quality. Overall, HDAM offers a promising pathway for enhancing real-world ICU decision support systems by enabling more accurate, balanced, and clinically actionable sepsis outcome predictions.

Data Availability

The raw dataset is publicly available in the MIMIC-IV repository: https://physionet.org/content/mimiciv/2.2/ and the processed and cleaned data provided by [2] is publicly available on GitHub: https://github.com/yuyinglu2000/Sepsis-Mortality.git.

Financial support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interests

There is no conflict of interest.

References

[1] R. M. Munshi, "Novel ensemble learning approach with SVM-imputed ADASYN features for enhanced cervical cancer prediction," PLoS One, vol. 19, no. 1, p. e0296107, 2024.
[2] C. Rashmi and C. Shantala, "Evaluating Deep Learning with different feature scaling techniques for EEG-based Music Entrainment Brain Computer Interface," E-Prime-Adv. Electr. Eng., Electron. Energy, vol. 7, p. 100448, 2024.
[3] J. Gao et al., "Prediction of sepsis mortality in ICU patients using machine learning methods," BMC Med. Inform. Decis. Mak., vol. 24, no. 1, p. 228, 2024.
[4] N. E. I. Karabadji et al., "Accuracy and diversity-aware multi-objective approach for random forest construction," Expert Syst. Appl., vol. 225, p. 120138, 2023.
[5] Y. Bao and S. Yang, "Two novel SMOTE methods for solving imbalanced classification problems," IEEE Access, vol. 11, pp. 5816–5823, 2023.
[6] S. Li et al., "Utilizing the lightgbm algorithm for operator user credit assessment research," arXiv preprint, arXiv:2403.14483, 2024.
[7] M. Moor et al., "Predicting sepsis using deep learning across international sites: a retrospective development and validation study," eClinicalMedicine, vol. 62, 2023.
[8] L. C. M. Liaw et al., "A histogram SMOTE-based sampling algorithm with incremental learning for imbalanced data classification," Inf. Sci., vol. 686, p. 121193, 2025.
[9] Z. Sun et al., "An improved random forest based on the classification accuracy and correlation measurement of decision trees," Expert Syst. Appl., vol. 237, p. 121549, 2024.
[10] M. Esmaeili-Falak and R. S. Benemaran, "Ensemble extreme gradient boosting based models to predict the bearing capacity of micropile group," Appl. Ocean Res., vol. 151, p. 104149, 2024.
[11] H. Koozi et al., "A simple mortality prediction model for sepsis patients in intensive care," J. Intensive Care Soc., vol. 24, no. 4, pp. 372–378, 2023.
[12] T. Sathish et al., "Characteristics estimation of natural fibre reinforced plastic composites using deep multi-layer perceptron (MLP) technique," Chemosphere, vol. 337, p. 139346, 2023.
[13] M. Niazkar et al., "Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023)," Environ. Model. Softw., p. 105971, 2024.
[14] F. Aldi et al., "Standardscaler's Potential in Enhancing Breast Cancer Accuracy Using Machine Learning," J. Appl. Eng. Technol. Sci. (JAETS), vol. 5, no. 1, pp. 401–413, 2023.
[15] K. Aggrawal et al., "Tools for Screening, Predicting, and Evaluating Sepsis and Septic Shock: A Comprehensive Review," Cureus, vol. 16, no. 8, p. e67137, 2024.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[16] V. G. Costa and C. E. Pedreira, "Recent advances in decision trees: An updated survey," Artif. Intell. Rev.., vol. 56, no. 5, pp. 4765–4800, 2023.
[17] R. Zhour, C. Khalid, and K. Abdellatif, "Hybrid intrusion detection system based on Random forest, decision tree, and Multilayer Perceptron (MLP) algorithms," in Proc. 2023 10th Int. Conf. Wireless Netw. Mobile Commun. (WINCOM), IEEE, 2023.
[18] Y.-L. Ning et al., "Tendency of dynamic vasoactive and inotropic medications data as a robust predictor of mortality in patients with septic shock: An analysis of the MIMIC-IV database," Front. Cardiovasc. Med.., vol. 10, 2023.
[19] G. Kong, K. Lin, and Y. Hu, "Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU," BMC Med. Inform. Decis. Mak., vol. 20, no. 1, p. 251, 2020.
[20] M. Moor et al., "Early Prediction of Sepsis in the ICU Using Machine Learning: A Systematic Review," Front. Med.., vol. 8, 2021.
[21] J. Weng et al., "Development and validation of a score to predict mortality in ICU patients with sepsis: a multicenter retrospective study," J. Transl. Med.., vol. 19, no. 1, p. 322, 2021.
[22] A. Alanazi et al., "Machine Learning for Early Prediction of Sepsis in Intensive Care Unit (ICU) Patients," Medicina, vol. 59, no. 7, p. 1276, 2023.
[23] M. S. Rahman et al., "Machine learning-based prognostic model for 30-day mortality prediction in Sepsis-3," BMC Med. Inform. Decis. Mak., vol. 24, no. 1, p. 249, 2024.