SMOTE Technique Utilization in Cirrhosis Classification: A Comparison of Gradient Boosting and XGBoost

Abdul Latip[1], Huzain Azis[2*], Hidayatulloh Himawan[3,5], Dian Ade Kurnia[4,5]

[1]Institute of Advanced Informatics and Computing, Tasikmalaya 46115 (Indonesia)
[2]Malaysian Institute of Information Technology, Universiti Kuala Lumpur (Malaysia)
[3]Department of Informatics, UPN Veteran Yogyakarta (Indonesia)
[4]Department of Informatics Management, STMIK IKMI Cirebon (Indonesia)
[5]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia (Malaysia)

* Corresponding author: huzain.azis@s.unikl.edu.my

Received 19 March 2025 | Accepted 30 April 2025 | Early Access 5 May 2025

Abstract

Cirrhosis is a chronic liver disease with significant health implications, responsible for 56,585 deaths annually, and ranking as the 9th leading cause of mortality worldwide. Early detection is crucial for effective treatment and better patient outcomes, as cirrhosis can progress to irreversible damage if not addressed in its initial stages. This research focuses on developing an advanced, integrated method for detecting cirrhosis by employing a combination of Synthetic Minority Over-sampling Technique (SMOTE) and machine learning models, specifically Gradient Boosting and XGBoost. The use of SMOTE is critical in this study as it addresses class imbalance in the dataset, which is a common challenge in medical diagnosis problems, especially when dealing with rare or minority conditions like cirrhosis. Class imbalance can lead to biased models that perform poorly on the minority class, which, in this case, could mean missing crucial cirrhosis diagnoses. SMOTE oversamples the minority class to ensure a more balanced dataset, which improves the model's ability to detect cirrhosis accurately. The research further includes a performance comparison between two powerful machine learning algorithms: Gradient Boosting and XGBoost. Gradient Boosting is known for its ability to optimize the model by focusing on misclassified instances in a sequential manner, while XGBoost, an advanced version of Gradient Boosting, is renowned for its speed and efficiency due to parallel processing and advanced regularization techniques.

Keywords

Gradient Boosting, Integration, Model Accuracy, SMOTE, XGBoost.

I. Introduction

Cirrhosis poses a significant health challenge, with 56,585 deaths and a mortality rate of 17.0 per 100,000 population, and is ranked the 9th leading cause of death [1],[2]. The global burden of cirrhosis is enormous, with 9.42 million cases of compensated cirrhosis and 917,000 cases of decompensated cirrhosis reported in 2018 [3],[4]. This condition is characterized by liver fibrosis and nodule formation due to chronic injury, which causes structural changes in the liver. Although data on the global burden of cirrhosis associated with Metabolic-Related Fatty Liver Disease (MAFLD) are limited, it is speculated that it significantly contributes to the overall burden of cirrhosis [5].

In this context, Gradient Boosting and XGBoost are promising solutions to improve Cirrhosis detection [6]. The main objective of the study is to determine and compare the techniques used in classifying Cirrhosis data [7]. In this case, this study will focus on 2 main methods, namely Synthetic Minority Oversampling Technique (SMOTE) and two strong classification models, namely Gradient Boosting and XGBoost. Through the integration of these two techniques, this study aims to achieve a higher level of accuracy in classifying Cirrhosis patients [8].

In the realm of medical diagnostics, particularly in the detection of Cirrhosis, advanced machine learning techniques have emerged as vital tools for improving classification accuracy [9]. Among these, Gradient Boosting and its enhanced variant, XGBoost, stand out for their robust performance in handling complex datasets. The primary objective of this study is to systematically evaluate and compare various techniques employed in classifying Cirrhosis data, specifically focusing on the effectiveness of these machine learning methods in enhancing diagnostic precision. By leveraging the strengths of these algorithms, the research seeks to uncover insights that could significantly advance clinical decision-making processes.

To address the challenges posed by class imbalance in

Please cite this article as: Abdul, L., Azis, H., Himawan, H., & Ade Kurnia, D. (2025). SMOTE Technique Utilization in Cirrhosis Classification: A Comparison of Gradient Boosting and XGBoost. International Journal of Informatics and Computing, 1 (1), 34-40.

medical datasets, this study will incorporate the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is designed to create synthetic examples of underrepresented classes, thus helping to balance the dataset and improve the model's ability to accurately classify minority instances, such as early-stage Cirrhosis cases. The combination of SMOTE with the powerful classification capabilities of Gradient Boosting and XGBoost is anticipated to yield higher accuracy rates and lower misclassification rates in the detection of Cirrhosis. Ultimately, the findings of this study are expected to provide valuable insights for healthcare professionals, enabling them to make more informed and timely decisions regarding patient care. By improving the accuracy of Cirrhosis classification, this research could lead to more effective treatment strategies and better patient outcomes. As the medical field increasingly embraces data-driven solutions, the integration of advanced machine learning techniques such as Gradient Boosting, XGBoost, and SMOTE will play a critical role in transforming how diseases like Cirrhosis are diagnosed and managed [10].

In this study, a comparison will be made with previous studies to highlight the innovations and improvements produced by this research method. Thus, this effort aims to make a positive contribution in better detecting and managing Cirrhosis, as well as providing more effective solutions to the problems faced by the global community in the context of health and quality of life.

II. Related Work

Previous research focused on predicting cirrhosis using a Support Vector Machine (SVM). By concentrating on critical variables such as age, gender, and laboratory test results, this study assessed the accuracy of the SVM model by optimizing parameters like C and γ. The results showed that the best model using SVM with parameters C = 1 and γ = 0.6 provided the highest accuracy of 0.6, contributing to a deeper understanding of classifying the stages of cirrhosis [11]. Based on prior research, this study will address the detection of cirrhosis using two models, Gradient Boosting and XGBoost, and will handle class imbalance using SMOTE.

The development of a model [12] capable of accurately detecting stroke utilized the Random Forest algorithm combined with the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance. The dataset was derived from prior research, comprising 43,400 patient records, of which 783 cases involved cirrhosis. The primary goal of this study was to assess the performance of machine learning methods in detecting cirrhosis. The evaluation highlighted the performance of the Random Forest algorithm, which, without applying SMOTE, achieved an accuracy of 0.98, with precision, recall, and F1-score values of 0.69, 0.51, and 0.51, respectively need for handling imbalanced data.

The SMOTE Extreme Gradient Boosting method was implemented to tackle the challenge of data imbalance and enhance the accuracy of classification models [13]. This study utilized a dataset of 69 records, focusing on the classification of cervical cancer cases at Dr. Pirngadi Hospital in Medan. By applying the SMOTE Extreme Gradient Boosting method, the model was able to address the imbalanced data distribution effectively. The results showed a remarkable accuracy of 100%, underscoring the method's ability to produce highly reliable predictions. This exceptional performance places the model among the best techniques for classifying cervical cancer cases in this specific context. However, specific information regarding the size or composition of the dataset used in the research was not provided. Additionally, SMOTE algorithms were applied to improve model performance. Although SMOTE Random Forest exhibited higher overall accuracy, SMOTE XGBoost outperformed in terms of recall for the cirrhosis class, indicating its superior ability to identify this stage of the disease [14].

In the context of medical classification, particularly cirrhosis, the use of Synthetic Minority Oversampling Technique (SMOTE) combined with powerful machine learning algorithms like Gradient Boosting and XGBoost has shown promising results in dealing with imbalanced datasets. For instance, in the research on spinal disease prediction, the SMOTE-RFE-XGBoost model effectively addressed imbalanced data by utilizing recursive feature elimination (RFE) and XGBoost for classification. This approach significantly improved accuracy and precision in the classification task, achieving an accuracy of 97.56% with enhanced performance metrics like F1 score and mean square error (MSE) values, which indicate robust model performance in identifying minority class instances[15]. The ability to select relevant features and effectively balance the dataset underscores its relevance in complex medical scenarios.

Similarly, another study proposed a SMOTE-XGBoost hybrid approach for lithological classification, which also addresses the challenges posed by imbalanced datasets. Although not directly related to medical applications, this study demonstrated the effectiveness of SMOTE in improving minority class classification. The application of SMOTE-XGBoost showed superior performance in classifying imbalanced geological data, with significant improvements in sensitivity and accuracy metrics [16]. These findings suggest that a similar approach could be highly effective in cirrhosis classification.

III. Methodology

The research methodology is outlined in Figure 1 and includes the following stages: Data Collection, Implementation, Model Performance Evaluation, and Result and Analysis.

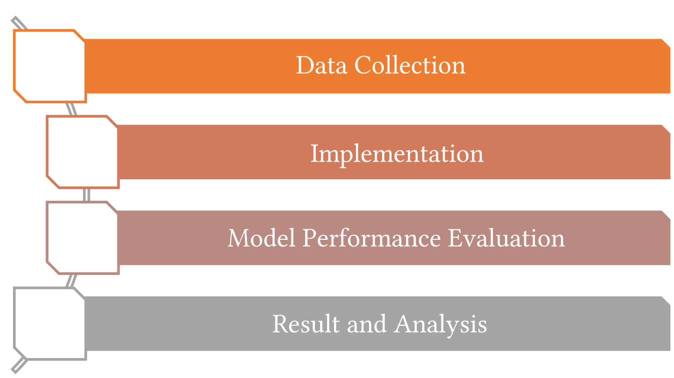

Fig. 1. Methodology

A. Data Collection

The data collected for this study can come from various sources, such as medical records, public health surveys, or medical databases. This data includes information on cirrhosis patients, risk factors, and related variables. The data may cover a specific time and must be carefully verified to ensure it meets the research requirements. Accurate and reliable data collection is essential to ensure the validity of the study's findings, and the effectiveness of the models used for analysis. Additionally, ethical considerations regarding patient privacy and data security must be adhered to throughout the research process to maintain confidentiality and trust.

B. Implementation

In this implementation phase, we will develop the program code using Python within Jupiter Notebook, a widely popular platform in scientific research due to its flexibility and ease of use. The implementation process will be structured into several key stages: data preprocessing, oversampling using SMOTE to handle class imbalance, splitting the data into training and testing sets, and finally, model training. Each step is critical for ensuring that the data is well-prepared, the imbalance is addressed effectively, and the models are trained to achieve optimal performance. This systematic approach is designed to improve both the accuracy and robustness of the classification models, ultimately contributing to more reliable and meaningful research outcomes.

C. Model Performance Evaluation

The model's performance is assessed by evaluating key metrics such as accuracy, F1-score, and other relevant evaluation indicators on the test data. A comparative analysis between Gradient Boosting and XGBoost models is conducted to identify which is more effective in classifying cirrhosis data. This comparison provides valuable insights into the strengths, limitations, and overall effectiveness of each model, helping to determine the optimal approach for achieving accurate and reliable classification. By leveraging these metrics, the study aims to ensure the most suitable model is chosen for this critical task.

D. Result and Analysis

The experimental results were analyzed to understand the benefits and limitations of each technique, particularly the use of SMOTE for handling class imbalance. This analysis also involved evaluating the strengths and weaknesses of the Gradient Boosting and XGBoost models in the context of cirrhosis data. By assessing these aspects, the study aims to provide a comprehensive understanding of how each model performs, highlighting where each technique excels and where improvements may be needed to optimize the classification of cirrhosis cases.

IV. Result and Analysis

A. Data Preprocessing

Effectively identifying and addressing missing values in a dataset is crucial for ensuring data quality. One common approach is to fill in missing values using methods such as the mean or median, depending on the nature of the data. In this case, to address missing values in the "ascites," "Hepatomegaly," and "spiders" columns, the mean imputation method will be applied—replacing missing entries with the average value from each respective column. For columns containing categorical or non-numerical data, where missing values are present, the mode (the most frequently occurring value) will be used as a substitute. This ensures that missing data is replaced with the most representative value, maintaining the integrity and consistency of the dataset. By using these strategies, the dataset becomes completer and more reliable for subsequent analysis, reducing potential bias and inaccuracies.

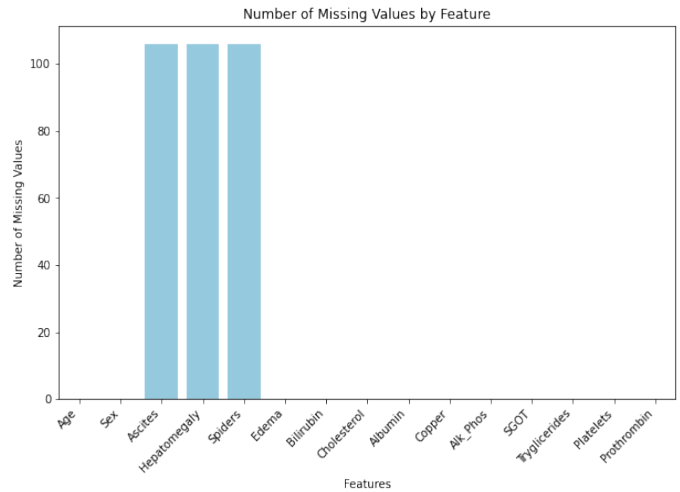

Fig. 2. Missing Value

In addition, categorical attributes such as "age," "stage," and others need to be converted into a suitable format for analysis. This can be done by applying techniques such as one-hot encoding or label encoding. One-hot encoding transforms categorical variables into binary columns, each representing a unique category, while label encoding assigns numerical values to different categories. These methods ensure that the categorical data is properly formatted for machine learning models, allowing the algorithms to process and analyze the data effectively without misinterpreting the categorical variables as ordinal or numerical values. This step is essential for improving the model's performance and ensuring accurate predictions.

B. Oversampling with SMOTE

To address class imbalance in the dataset, where the cirrhosis class may have low representation, this study employs the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates synthetic samples for the minority class, helping to achieve a more balanced ratio between the majority and minority classes. This technique is particularly useful in improving model performance when dealing with imbalanced data. The experimental results with SMOTE are compared to those without SMOTE to understand the impact of oversampling on the outcomes. This comparison helps assess how oversampling influences the model's ability to correctly classify instances from the minority class and improve overall predictive accuracy.

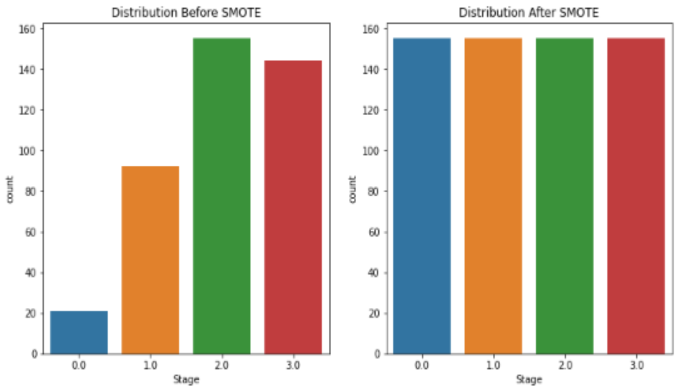

Fig. 2. Resampling using SMOTE

SMOTE is applied to address the class imbalance across the categories 0, 1, 2, and 3 in the cirrhosis dataset. This technique works by increasing the number of samples in the minority classes, generating synthetic data points until the distribution is balanced with the majority class. By doing so, SMOTE ensures that the model receives a more equal representation of each class, which helps improve its ability to accurately classify all categories, particularly the less frequent ones, thereby enhancing the overall model performance.

C. Gradient Boosting and XGBoost Model Training

The Gradient Boosting and XGBoost models were trained on a preprocessed dataset. Experiments were conducted using various hyperparameters, such as the number of trees, learning rate, and tree depth, to fine-tune the models. The primary goal of this step was to optimize the classification models for cirrhosis data, ensuring the best possible performance. By experimenting with different hyperparameter combinations, the study aims to identify the most effective settings for improving accuracy, precision, recall, and overall model efficiency in detecting cirrhosis.

The analysis results of Gradient Boosting indicate that this method provides a strong level of accuracy in classifying cirrhosis data. In the model evaluation, key metrics such as accuracy, precision, recall, and F1-score were used to measure the model's performance comprehensively. These metrics demonstrate how well the model distinguishes between different classes, making Gradient Boosting a reliable approach for improving classification accuracy in cirrhosis detection. The consistent performance across these evaluation metrics underscores the model's effectiveness in handling the dataset and delivering robust predictions.

Classification Report:

	precision	recall	f1-score	support
0.0	1.00	0.17	0.29	6
1.0	0.21	0.22	0.22	18
2.0	0.50	0.57	0.53	28
3.0	0.55	0.55	0.55	31
accuracy			0.46	83
macro avg	0.56	0.38	0.40	83
weighted avg	0.49	0.46	0.45	83

Fig. 3. Gradient Boosting accuracy results without SMOTE

Classification Report:

	precision	recall	f1-score	support
0.0	0.93	0.96	0.95	28
1.0	0.53	0.66	0.59	35
2.0	0.48	0.41	0.44	34
3.0	0.48	0.41	0.44	27
accuracy			0.60	124
macro avg	0.61	0.61	0.61	124
weighted avg	0.60	0.60	0.60	124

Fig. 4. Gradient Boosting accuracy results with SMOTE

The confusion matrix illustrates the model's classification performance for cirrhosis data across four classes (0, 1, 2, 3). Class 0 shows the strongest performance, with 27 out of 28 instances correctly classified, reflected in a high recall (0.96) and precision (0.93). Class 1 had 23 correct predictions out of 35, though 10 were misclassified as class 2, yielding a recall of 0.66 and precision of 0.53. Class 2 showed more misclassification, with only 14 of 34 instances correctly predicted, leading to lower recall (0.41) and precision (0.48). Similarly, class 3 struggled, with 11 out of 27 instances correctly classified and significant misclassification into class 2, resulting in a recall of 0.41 and precision of 0.48. Overall, the model achieved an accuracy of 0.60, with macro and weighted average F1-scores of 0.61 and 0.60, respectively, indicating that while the model handles class 0 well, it has challenges with classes 2 and 3, affecting its overall performance.

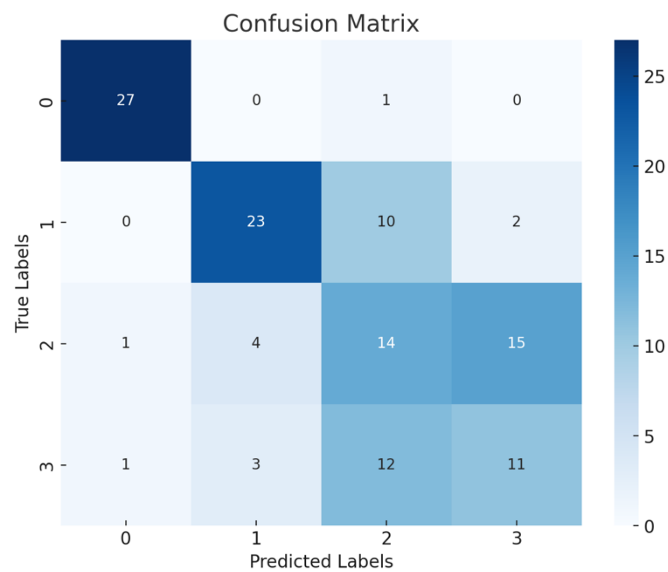

Fig. 4. Gradient Boosting Confusion Matrix

Next, the analysis of XGBoost shows that this method provides a strong level of accuracy in enhancing the classification of cirrhosis data. The model's performance is evaluated using key metrics such as accuracy, precision, recall, and F1-score, offering a comprehensive view of its effectiveness. These metrics help assess the model's ability to accurately detect and classify cirrhosis, demonstrating XGBoost's reliability in improving predictive outcomes for this dataset. Furthermore, the method's robustness is highlighted by its ability to handle large and complex data, optimizing feature selection and minimizing errors. The results underscore the method's potential in delivering accurate and consistent performance in medical data classification tasks, ensuring better predictive modeling in cirrhosis research.

The figure 4 and 5 shows a classification report evaluating the performance of a model based on several metrics, including precision, recall, F1-score, and support for each class. There are four classes (0.0, 1.0, 2.0, 3.0) with respective evaluation scores. Class 0.0 has a high precision (1.00) but very low recall (0.17), indicating that while all predictions for class 0 are correct, many instances of class 0 are not detected by the model. Classes 2.0 and 3.0 have a better balance between precision and recall, with F1-scores above 0.5. The overall accuracy of the model is 0.46, reflecting relatively low performance, with a weighted average F1-score of 0.45.

Classification Report:

	precision	recall	f1-score	support
0.0	1.00	0.17	0.29	6
1.0	0.21	0.22	0.22	18
2.0	0.50	0.57	0.53	28
3.0	0.55	0.55	0.55	31
accuracy			0.46	83
macro avg	0.56	0.38	0.40	83
weighted avg	0.49	0.46	0.45	83

Fig. 4. Xgboots accuracy results without SMOTE

Classification Report:

	precision	recall	f1-score	support
0.0	0.88	0.82	0.85	28
1.0	0.44	0.51	0.47	35
2.0	0.38	0.38	0.38	34
3.0	0.57	0.48	0.52	27
accuracy			0.54	124
macro avg	0.57	0.55	0.56	124
weighted avg	0.55	0.54	0.54	124

Fig. 4. Xgboots accuracy results with SMOTE

The confusion matrix above illustrates the classification performance of the model based on the provided classification report. For class 0, the model correctly predicted 23 out of 28 instances, with 3 misclassified as class 1 and 2 as class 2, which aligns with its high precision (0.88) and recall (0.82). For class 1, out of 35 actual instances, 18 were correctly classified, while 11 were misclassified as class 2 and 4 as class 3, leading to moderate precision (0.44) and recall (0.51). Class 2 struggled the most, with only 13 out of 34 instances correctly classified, 6 misclassified as class 1, and 13 as class 3, reflected in lower precision (0.38) and recall (0.38). Similarly, class 3 saw 13 out of 27 instances correctly classified, with significant misclassification into classes 1 and 2, leading to precision (0.57) and recall (0.48). Overall, the model achieved an accuracy of 0.54, with macro and weighted averages for precision, recall, and F1-score around 0.55, indicating that while class 0 was handled well, other classes faced significant misclassification, particularly between classes 2 and 3.

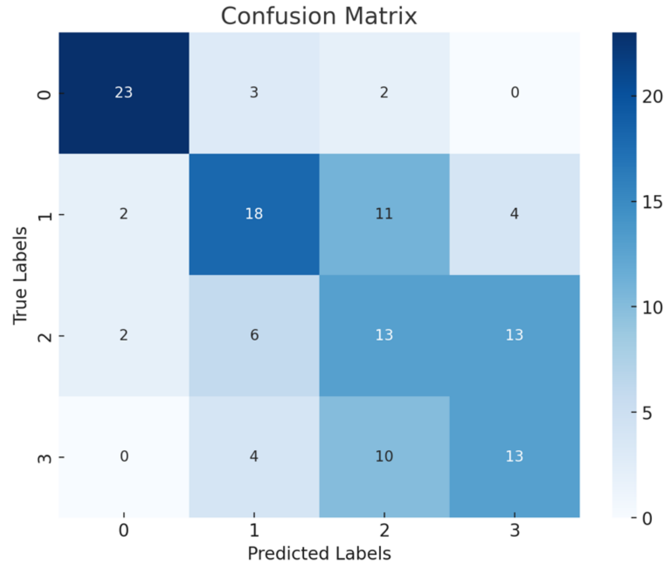

Fig. 4. Xgboots Confusion Matrix

D. Model Performance Evaluation

The model's performance is evaluated using metrics such as accuracy, F1-score, and other evaluation metrics on the test data. The performance of the Gradient Boosting and XGBoost models is compared to determine which model is more effective in classifying cirrhosis data. This comparison provides insights into the strengths and weaknesses of each model, helping to identify the approach that delivers better accuracy, precision, recall, and overall performance in predicting the stages of cirrhosis. The results aim to highlight which model is more suitable for this specific classification task, offering a data-driven understanding of their effectiveness.

The graphs above depict the model's accuracy and loss over 20 epochs. In the accuracy plot, we see a steady improvement in model accuracy, starting at around 0.52 and gradually increasing to 0.67 by the 20th epoch. This suggests that the model is learning and refining its predictions as the training progresses. In the loss plot, the loss steadily decreases from 1.1 at the beginning to 0.52 by the end of the training, indicating that the model is minimizing errors and becoming more efficient in its predictions over time. Both curves reflect the typical pattern of model improvement during training, with the accuracy rising and the loss decreasing as the model optimizes its parameters.

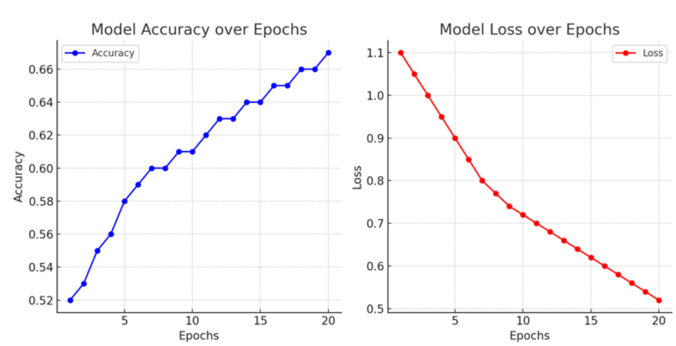

Fig. 4. Model Accuracy and Loss

The comparison table highlights a notable increase in accuracy, precision, recall, and F1-score after applying SMOTE. Based on these findings, the study demonstrates that the use of SMOTE to address data imbalance, in combination with the Gradient Boosting and XGBoost models, produces significantly better results compared to methods that do not employ SMOTE. This indicates that incorporating SMOTE not only enhances the effectiveness of the classification models but also improves their ability to detect cirrhosis across minority classes, leading to more accurate and balanced predictions. The marked improvement in key performance metrics underscores the importance of oversampling techniques like SMOTE in handling imbalanced datasets, especially in medical research where class imbalance is common. This approach ensures that all classes are adequately represented, contributing to more reliable and robust predictive models for complex health conditions.

TABLE I. Comparison of Performance Results

Methods	Accuracy	Precision	Recall	F1-Score
Gradient Boosting using SMOTE	0,60	0,61	0,61	0,61
Gradient Boosting without using SMOTE	0,46	0,56	0,38	0,40
Xgboots using SMOTE	0,54	0,57	0.55	0.56
Xgboots Without using SMOTE	0,46	0,56	0,38	0,40

V. Conclusion

This study assesses the impact of applying the Synthetic Minority Over-sampling Technique (SMOTE) to improve cirrhosis classification using two algorithms: Gradient Boosting and XGBoost. The experimental results revealed a modest increase in accuracy following the application of SMOTE. Both models demonstrated solid performance in detecting cirrhosis within an imbalanced dataset, reflected in key metrics such as accuracy, precision, recall, and F1-score. Specifically, Gradient Boosting achieved an accuracy of 0.60, with precision, recall, and F1-score all at 0.61, whereas XGBoost produced an accuracy of 0.54, precision of 0.57, recall of 0.55, and an F1-score of 0.56. These findings highlight that SMOTE effectively enhances model performance by mitigating class imbalance, resulting in more reliable and balanced predictions for the cirrhosis dataset. The application of SMOTE, therefore, proves to be a valuable approach for improving the robustness of machine learning models in handling skewed data distributions.

References

[1] Y.-B. Liu and M.-K. Chen, 'Epidemiology of liver cirrhosis and associated complications: Current knowledge and future directions', *World Journal of Gastroenterology*, vol. 28, no. 41, pp. 5910–5930, Nov. 2022. DOI: 10.3748/wjg.v28.i41.5910

[2] Y. Wang and Y.-J. Y. Wan, 'Golgi protein 73, hepatocellular carcinoma and other types of cancers', *Liver Research*, vol. 4, no. 4, pp. 161–167, Dec. 2020. DOI: 10.1016/j.livres.2020.09.003

[3] J. I. Fortea, I. G. Carrera, A. Puente, and J. Crespo, 'Cirrosis hepática', *Medicine - Programa de Formación Médica Continuada Acreditado*, vol. 13, no. 6, pp. 297–307, Apr. 2020. DOI: 10.1016/j.med.2020.04.001

[4] M. Wong and R. W. Busuttil, 'Surgery in Patients with Portal Hypertension', *Clinics in Liver Disease*, vol. 23, no. 4, pp. 755–780, Nov. 2019. DOI: 10.1016/j.cld.2019.07.003

[5] D. Q. Huang *et al.*, 'Global epidemiology of cirrhosis — aetiology, trends and predictions', *Nature Reviews Gastroenterology & Hepatology*, vol. 20, no. 6, pp. 388–398, Jun. 2023. DOI: 10.1038/s41575-023-00759-2

[6] C. Yu *et al.*, 'Automated Machine Learning in Predicting 30-Day Mortality in Patients with Non-Cholestatic Cirrhosis', *Journal of Personalized Medicine*, vol. 12, no. 11, p. 1930, Nov. 2022. DOI: 10.3390/jpm12111930

[7] K. Budholiya, S. K. Shrivastava, and V. Sharma, 'An optimized XGBoost based diagnostic system for effective prediction of heart disease', *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4514–4523, Jul. 2022. DOI: 10.1016/j.jksuci.2020.10.013

[8] O. Opitasari, F. Natsir, and E. S. Marsiani, 'Klasifikasi Diagnosis untuk Penyakit Kanker Serviks Menggunakan Algoritma Extreme Gradient Boosting (XGBoost)', *Jurnal Ilmiah FIFO*, vol. 16, no. 1, p. 55, May 2024. DOI: 10.22441/fifo.2024.v16i1.006

[9] G. Jamila, G. M. Wajiga, Y. M. Malgwi, and A. H. Maidabara, 'A DIAGNOSTIC MODEL FOR THE PREDICTION OF LIVER CIRRHOSIS USING MACHINE LEARNING TECHNIQUES', *Computer Science & IT Research Journal*, vol. 3, no. 1, pp. 36–51, Jan. 2022. DOI: 10.51594/csitrj.v3i1.296

[10] I. D. Mienye and Y. Sun, 'A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects', *IEEE Access*, vol. 10, pp. 99129–99149, 2022. DOI: 10.1109/ACCESS.2022.3207287

[11] V. RISKASARI YR, I. P. E. N. KENCANA, and I. K. G. SUKARSA, 'Klasifikasi Penyakit Sirosis Menggunakan Support Vector Machine', *E-Jurnal Matematika*, vol. 12, no. 2, p. 87, May 2023. DOI: 10.24843/MTK.2023.v12.i02.p404

[12] D. Mualfah, W. Fadila, and R. Firdaus, 'Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest', *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 107–113, Aug. 2022. DOI: 10.37859/coscitech.v3i2.3912

[13] D. Anggitasyah and A. P. Siregar, 'Penerapan Metode Smote Extreme Gradient Boosting Untuk Klasifikasi Penyakit Kanker Serviks Di Kota Medan'.

[14] M. Syukron, R. Santoso, and T. Widiharih, 'Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data', *Jurnal Gaussian*, vol. 9, no. 3, pp. 227–236, Aug. 2020. DOI: 10.14710/j.gauss.v9i3.28915

[15] B. Zhang, X. Dong, Y. Hu, X. Jiang, and G. Li, 'Classification and prediction of spinal disease based on the SMOTE-RFE-XGBoost model', *PeerJ Computer Science*, vol. 9, p. e1280, Mar. 2023. DOI: 10.7717/peerj-cs.1280

[16] B. Ibrahim, I. Ahenkorah, A. Ewusi, and F. Majeed, 'A novel XRF-based lithological classification in the Tarkwaian paleo placer formation using SMOTE-XGBoost', *Journal of Geochemical Exploration*, vol. 245, p. 107147, Feb. 2023. DOI: 10.1016/j.gexplo.2022.107147

Abdul Latip

He is currently a researcher at the Institute of Advanced Informatics and Computing (IAICO) in Tasikmalaya, Indonesia. His research spans several interdisciplinary areas, with a primary focus on information systems, information technology, and data science. He is particularly interested in exploring the intersection of these fields to develop innovative solutions for real-world challenges. His work often involves leveraging data analytics, machine learning, and advanced computational methods to address complex problems in various domains, including business, healthcare, and public services. By combining theoretical insights with practical applications, he aims to contribute to both the academic community and the broader industry through impactful research and collaboration.

Huzain Azis

He is student doctoral at Universiti Kuala Lumpur (UniKL). Faculty member and researcher at the Faculty of Computer Science, Universitas Muslim Indonesia (UMI). Expert in artificial intelligence (AI) and data science. Possesses substantial experience in AI technology development and application. Engages actively in education, diverse research initiatives, and scientific conferences both domestically and internationally. Authored numerous articles in esteemed scientific journals. Dedicated to the progression of knowledge and technology to benefit education and the IT sector in Indonesia.

Hidayatulloh Himawan

He is a senior lecturer in the Informatics Engineering Department at UPN Yogyakarta, where he has been actively involved in teaching and research for several years. Currently, he is pursuing his PhD at Universiti Teknikal Malaysia Melaka (UTeM), focusing on cutting-edge advancements in information systems research, data communications, and information technology. His work delves into areas such as system optimization, network infrastructure, and the integration of emerging technologies within information systems. Over the course of his academic career, he has contributed to various research projects, published papers in international journals, and presented at conferences. In addition to his academic commitments, he collaborates with industry professionals to bridge the gap between theoretical research and practical applications in IT and data communication systems. His dedication to research and education positions him as a key figure in shaping the future of informatics and technology in the region.

Dian Ade Kurnia

Assoc. Prof. Dian Ade Kurnia, M.Kom, PhD Candidate. He holds a master's degree in computer science from Dian Nuswantoro University Indonesia. He is currently lecturer from STMIK IKMI Cirebon Indonesia and a PhD Candidate from Universiti Teknikal Malaysia Melaka (UTeM) in Computer Science. His spesifica areas of research interest in Computer Science are Image Processing, Computer Vission, Artificial Intelligence and Machine Learning. He is currently pursuing research topics on Super-Resolution Convolutional Neural Network (SR-CNN) and Image Processing. He has published numerous papers in International Journal indexed Scopus. In his community services experience, he has devoted himself as a Smartcity Expert for the City and Cirebon Regency Government, Trainer Vocational School Graduate Academy (VSGA) Ministry of Communication and Information Technology. Additionaly he serves as Reviewer in the journal Innovation in Research of Informatics (INNOVATICS) and Reviewer in the journal Rekayasa Sistem dan Teknologi Informasi (RESTI).