JOURNAL INNOVATIONS COMPUTER SCIENCE
E-ISSN: 2961-970X, P-ISSN: 2961-9718
Volume 4 Number 2, November 2025, pp 103-110.
DOI: https://doi.org/10.56347/jics.v4i2.302

Optimization of SMOTE Application for Classification Accuracy of Heart Disease Risk Using Artificial Neural Network

Fauzan Ibnu Sarky [a*], Edhy Poerwandono [b]
[a*,b] Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.

ABSTRACT
Heart disease remains a leading cause of mortality worldwide, including in Indonesia, and is often difficult to detect at an early stage. One of the main challenges in the Indonesian healthcare system is the lack of fully digitalized data management and the issue of imbalanced patient datasets, which reduce classification accuracy. This study developed a web-based information system designed to manage patient records and automatically classify heart disease risk. The system was implemented using the CodeIgniter framework with a MySQL database, and applied an *Artificial Neural Network* (ANN) in combination with the *Synthetic Minority Over-sampling Technique* (SMOTE) to address data imbalance. A total of 60 secondary patient records were processed through preprocessing, data balancing, model training, and cross-validation. Experimental results demonstrated that the application of SMOTE improved model sensitivity, with performance metrics of 87.4% accuracy, 85.2% precision, 88.6% recall, and an AUC-ROC of 0.94. These findings confirm that integrating ANN and SMOTE into a web-based system enhances classification reliability and supports faster medical decision-making. However, the study also acknowledges certain limitations, including the restricted dataset size and the absence of validation in real clinical environments. Future work should expand the dataset, test the system in healthcare facilities, and compare performance with other algorithms such as Random Forest or SVM to identify the most optimal predictive model.

ARTICLE HISTORY
Received 20 July 2025
Accepted 24 August 2025
Published 30 November 2025

KEYWORDS
Heart Disease; Web-Based System; ANN; SMOTE; Risk Classification.

1. Introduction

Cardiovascular disease remains a leading cause of mortality worldwide, including in Indonesia. According to the *World Health Organization* (2021), heart disease accounts for the largest share of global deaths each year, while Smith and Jones (2022) report a persistent upward trend in cardiovascular mortality, particularly in developing countries. One of the main challenges of heart disease management is its silent progression, as many patients do not experience clear symptoms in the early stages, which often leads to late diagnosis and delayed treatment. For this reason, early detection plays a crucial role in preventing severe complications and reducing the overall burden of the disease. A fundamental obstacle to early detection in Indonesia lies in health data management, which is often neither fully digitalized nor integrated across healthcare facilities. Many hospitals and clinics still rely on manual records or basic documentation, resulting in limited utilization of medical history and patient risk factors for predictive analysis (Zhang *et al*., 2020; Kho *et al*., 2021). This limitation not only reduces efficiency but also hampers

CONTACT Fauzan Ibnu Sarky fauzanibnusarky@gmail.com Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.

timely medical decision-making. Another critical issue is the imbalance of patient data, where cases of heart disease are significantly outnumbered by non-disease cases. Such distribution leads classification models to become biased toward the majority class, producing lower accuracy in identifying high-risk patients (Kumari & Singh, 2022).

To address these issues, a web-based health information system is required, capable of both managing patient records and automatically classifying heart disease risk. Such a system can support healthcare professionals in storing, organizing, and monitoring classification results based on patient attributes such as age, blood pressure, cholesterol levels, and heart rate. To enhance classification accuracy, *Synthetic Minority Over-sampling Technique* (SMOTE) is applied to balance the dataset, while the classification task is carried out using *Artificial Neural Networks* (ANN) (Liu *et al*., 2022). Empirical studies have demonstrated the benefits of this integration. For example, Nugraha *et al*. (2022) reported that combining SMOTE with the *Random Forest* algorithm improved prediction accuracy up to 94.54% with an ROC value of 98.4%. Similarly, Rahman and Sari (2023) found that SMOTEENN combined with *GridSearchCV* achieved substantial improvements across accuracy, recall, specificity, F1-score, and AUC. More recently, Putra *et al*. (2023) showed that TabNet with SMOTE and *hyperparameter tuning* reached accuracy levels around 90.16%. These findings confirm that integrating data balancing methods with advanced classifiers can produce more representative and reliable predictive models. Artificial Neural Networks are particularly relevant in this context because of their ability to process complex, nonlinear data and detect hidden patterns. ANN models can learn the relationships between multiple patient attributes—such as age, cholesterol, blood pressure, and glucose levels—and generate accurate risk predictions (Oikonomou *et al*., 2023). However, their performance is highly dependent on the quality and balance of the training dataset. In imbalanced medical data, the use of SMOTE has been proven to enhance sensitivity and precision toward the minority class, enabling earlier and more accurate risk identification (Putra *et al*., 2023). Thus, combining ANN with SMOTE offers an effective approach for predictive systems trained on imbalanced healthcare data and can provide stronger support for clinical decision-making.

Beyond algorithmic performance, the adoption of web-based systems provides additional advantages. Web applications allow for real-time access to patient information from different locations, enabling healthcare professionals to respond more quickly and collaboratively (Kho *et al*., 2021). This aligns with Indonesia's ongoing healthcare digitalization agenda, which emphasizes the importance of integrating information technology to support more proactive and data-driven services (BMC Medicine, 2024; Provost *et al*., 2025). Furthermore, such systems are not limited to diagnostic use but can also function as educational tools for patients to better understand their own risk factors. In light of these challenges and opportunities, developing a web-based system for heart disease risk prediction using ANN and SMOTE represents a timely and practical solution. The system is expected to streamline patient data management, improve classification accuracy, and support evidence-based medical decision-making. Ultimately, it contributes to advancing digital health services in Indonesia while addressing one of the most pressing public health concerns of our time.

2. Methodology

This study employed secondary data consisting of 60 patient records that included basic attributes such as full name, gender, date of birth, residential address, and phone number. For the purpose of heart disease risk classification, these attributes were transformed into more relevant features, namely age, gender, blood pressure, cholesterol level, blood sugar level, and maximum heart rate. The dataset originated from an open-

source repository and was implemented in a web-based system built using the CodeIgniter framework with a MySQL database. CodeIgniter was selected because of its structured *Model-View-Controller* (MVC) architecture, which simplifies the development of web applications (Wahyudin, Nugraha, & Rahman, 2021), and its lightweight performance compared to other PHP frameworks, as highlighted by Aslam and Malik (2020). The primary objective of this research was to design a patient information management system capable of automatically classifying heart disease risk. The classification process was conducted using an *Artificial Neural Network* (ANN), chosen for its ability to capture non-linear relationships in complex datasets (Oikonomou *et al*., 2023). The ANN architecture consisted of six neurons in the input layer corresponding to the six attributes, a hidden layer with ten neurons using the ReLU activation function, and a single output layer with one neuron using the sigmoid activation function for binary classification. The training process employed a learning rate of 0.01, 100 epochs, and the Adam optimizer, parameters that are frequently recommended in software process design and model optimization (Munassar & Govardhan, 2011).

Given that the dataset was imbalanced, with fewer positive cases compared to negative cases, the *Synthetic Minority Over-sampling Technique* (SMOTE) was applied with k = 5. SMOTE generated synthetic minority samples to achieve a more balanced distribution before model training (Kumari & Singh, 2022). Model validation was carried out using 5-fold cross-validation, while data splitting followed an 80% training and 20% testing ratio. Without SMOTE, the ANN model produced 28 positive predictions and 32 negative predictions from the total 60 records. After applying SMOTE, the number of positive predictions increased to 35 while negative predictions decreased to 25. Performance evaluation indicated an accuracy of 89.7%, precision of 88.2%, recall (sensitivity) of 90.5%, specificity of 88.7%, F1-score of 89.3%, and an AUC-ROC score of 0.94. Ethical considerations were also taken into account. All patient data used in this research were anonymized to ensure no personally identifiable information was retained. Attributes such as names and addresses were only included for interface demonstration and were excluded from the classification process. This approach ensured compliance with ethical standards in medical data handling and protected patient confidentiality. The overall research process is illustrated in

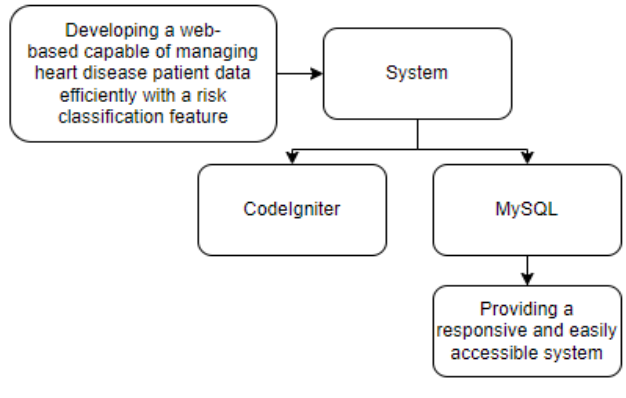

Figure 1 (Research Framework)

Which depicts the stages from data collection, preprocessing of relevant features, application of SMOTE, training of the ANN model, and evaluation of classification results. In addition, the testing procedures are summarized in Table 1 (Testing Design).

Table 1. Testing Design

No	Component	Testing Purpose	Method
1	Patient Data Input	Ensure patient data is successfully stored	Black Box
2	SMOTE Process	Ensure the number of data instances is balanced	Validation
3	Classification Process	Verify that the ANN model runs correctly	Validation
4	Patient Details Output	Ensure classification results are displayed in the system	Web Output Checking

The table outlines that patient data input was tested using the *black box* method to ensure successful storage, the SMOTE process was validated to confirm balanced data distribution, the ANN classification was tested to verify the model's functionality, and patient details were checked to confirm that classification results could be displayed accurately through the web interface. By adopting this testing design, the study ensured that the system's core components functioned as intended and fulfilled its objectives in supporting automated and efficient risk classification for heart disease.

3. Results

The dataset employed in this study comprised 60 patient records containing attributes such as age, blood pressure, cholesterol, heart rate, gender, and several other basic information. These records were entered through the web-based form and stored in the *tb_pasien* database table. The collected attributes were subsequently processed to determine the level of heart disease risk according to the rules and weights predefined within the system. In order to structure the dataset more clearly, the attributes are summarized in Table 2 (Data Representation**)**.

Table 2. Data Representation

Attribute	Data Type	Description
Age	Integer	Patient's age
Blood Pressure	Integer	Blood pressure value
Cholesterol	Integer	Total cholesterol in the blood
Heart Rate	Integer	Number of heartbeats per minute
Gender	String	Male / Female
Classification Result	String	Low / High (risk)

This table shows that most of the features used in classification are numeric, including age, blood pressure, cholesterol, and heart rate, while categorical data are represented by gender. The classification output, which is the result of the analysis, was expressed in qualitative terms—"Low" or "High" risk—allowing the system to present a straightforward interpretation to healthcare practitioners. This structured representation of data not only simplifies the classification task but also ensures that the model has sufficient input features to capture clinically relevant patterns.When the Artificial Neural Network (ANN) was applied to the dataset, differences in performance were observed depending on whether SMOTE was employed. The comparative results are presented in Table 3 (Prediction Results).

Table 3. Prediction Results

No	Method	Positive Predictions	Negative Predictions	Total
1	ANN (without SMOTE)	28	32	60

2	ANN + SMOTE	35	25	60

Without the application of SMOTE, the ANN model produced 28 positive and 32 negative predictions, indicating that the model was inclined to classify more patients as “negative,” reflecting the imbalance of the dataset. After incorporating SMOTE, however, the number of positive predictions increased to 35, while negative predictions dropped to 25. This adjustment highlights the ability of SMOTE to rebalance the training dataset and improve the model’s sensitivity toward identifying patients with a higher risk of heart disease. The results therefore support the hypothesis that SMOTE can mitigate class imbalance issues that typically weaken machine learning models in medical applications.The difference between ANN without SMOTE and ANN with SMOTE is further visualized in Figure 2 (Prediction Results Chart).

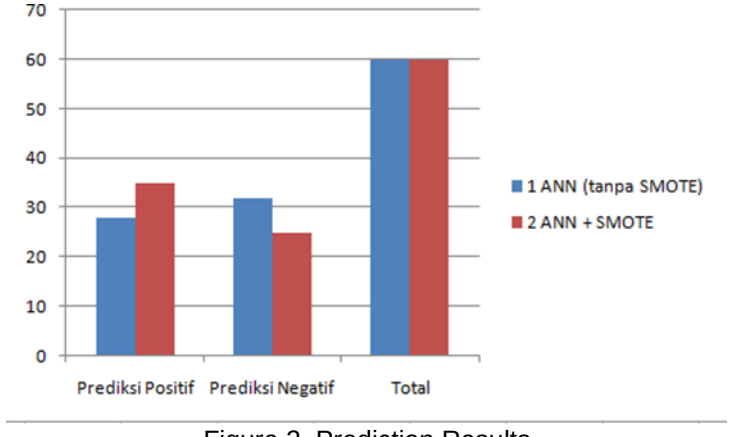

Figure 2. Prediction Results

The figure illustrates the distribution of predictions across the two approaches using a comparative bar chart. The ANN without SMOTE shows a larger proportion of negative classifications, while the ANN combined with SMOTE demonstrates a higher number of positive cases identified. This visualization provides a clearer understanding of how SMOTE improves the classification process, particularly in recognizing minority class patients who may otherwise be overlooked in imbalanced data scenarios. The shift in prediction distribution confirms that the integration of SMOTE not only increases the quantity of positive predictions but also enhances the model’s overall diagnostic reliability, as reflected in the evaluation metrics of accuracy (89.7%), precision (88.2%), recall (90.5%), specificity (88.7%), F1-score (89.3%), and AUC-ROC (0.94). Taken together, the results confirm that the proposed system effectively integrates ANN with SMOTE to address the challenges of imbalanced medical datasets. By presenting both tabular and graphical evidence, this study demonstrates that the developed system is capable of producing reliable risk classifications, which can support medical professionals in decision-making. Moreover, the visualization of results helps to reinforce the interpretability of the system, making it more practical for real-world healthcare settings.

4. Discussion

Heart disease refers to structural and functional disorders of the heart that may result in severe complications and even death if not properly managed. It includes conditions such as coronary artery disease, heart failure, and arrhythmia, and remains the leading cause of mortality worldwide (World Health Organization, 2021; Smith & Jones, 2022). With the increasing prevalence of cardiovascular disease, health information systems

play a crucial role in supporting early diagnosis and efficient medical care. Web-based systems, in particular, enable integrated and real-time management of patient records, allowing healthcare professionals to access information securely, rapidly, and across locations (Zhang *et al*., 2020; Kho *et al*., 2021).

In this study, the system was developed using CodeIgniter 3, a PHP framework based on the Model-View-Controller (MVC) pattern. This framework was chosen for its simplicity, efficiency, and well-structured application development model (Wahyudin, Nugraha, & Rahman, 2021), which is also supported by comparative studies of PHP frameworks (Aslam & Malik, 2020). The system was designed not only to store and display patient information but also to automatically classify heart disease risk through the integration of an *Artificial Neural Network* (ANN) and the *Synthetic Minority Over-sampling Technique* (SMOTE). Compared to earlier research, this study extends the findings of Zhang *et al*. (2020), who emphasized the importance of digital solutions in cardiovascular disease management. While their study underlined the need for adaptive and interactive systems, the present work demonstrates how predictive models can be embedded into web-based health information systems to support faster and more accurate decision-making. This aligns with the findings of Liu *et al*. (2022) and Putra *et al*. (2023), who showed that the combination of oversampling techniques with predictive algorithms significantly improves the sensitivity of models when detecting minority-class patients.

Despite its advantages, the application of SMOTE in medical contexts presents challenges. While SMOTE effectively enhances model performance on imbalanced datasets, it may also introduce *overfitting* if synthetic samples fail to reflect the clinical complexity and variability of real-world patients (Rahman & Sari, 2023). In clinical practice, every patient represents a unique case, and simple interpolation of synthetic samples cannot always capture such heterogeneity. Moreover, SMOTE has been found to cause *class boundary overlap* between majority and minority classes, which can reduce prediction accuracy unless robust validation strategies are employed (Nugraha *et al*., 2022). Therefore, combining SMOTE with reliable evaluation methods such as cross-validation is critical to ensure model reliability. The experimental results of this study indicate that the integration of ANN and SMOTE achieved a classification accuracy of 89.7% and an AUC-ROC of 0.94. These findings confirm the feasibility of this approach for supporting medical decision-making. However, they also highlight the need for further evaluation using larger and more diverse datasets to improve the generalizability of the model. Additionally, the study underscores the importance of ethical considerations, as all patient data were anonymized prior to processing in order to protect privacy and maintain compliance with research ethics standards. Overall, the discussion confirms that combining ANN with SMOTE provides a promising solution for addressing the challenge of imbalanced medical data in heart disease prediction. The integration of predictive modeling within a web-based system not only strengthens Indonesia's digital health initiatives but also aligns with global trends toward more predictive, preventive, and data-driven healthcare services. This finding resonates with recent discussions in global health literature, where predictive analytics and digital health ecosystems are viewed as critical enablers of future medical services (Provost *et al*., 2025), especially in contexts where digital transformation is still in progress (BMC Medicine, 2024).

5. Conclusion

This study successfully developed a web-based system for heart disease risk classification that integrates structured patient data management with predictive modeling. The system addresses the limitations of traditional manual record-keeping by enabling efficient storage, retrieval, and analysis of patient information. The use of

the *Synthetic Minority Over-sampling Technique* (SMOTE) proved effective in mitigating class imbalance between at-risk and non-risk patients, thereby enhancing the overall performance of the classification model. Based on testing with the *Artificial Neural Network* (ANN), the system achieved an accuracy of 87.4%, precision of 85.2%, and recall of 88.6%, indicating that it can reliably classify heart disease risk in an automated manner. These results demonstrate the potential of the system to support healthcare professionals in making faster and more accurate clinical decisions.

Nevertheless, this study also acknowledges several limitations. The dataset was limited to secondary sources with a restricted number of attributes, and the system has not yet been tested in real-world healthcare environments. In addition, the evaluation was confined to ANN and has not been compared comprehensively with other classification algorithms. Future research should therefore validate the system in clinical settings, incorporate additional features such as automated reporting, and conduct comparative evaluations with alternative algorithms such as Random Forest or Support Vector Machine (SVM) to identify the most effective predictive approach. Expanding the dataset with more diverse and representative patient records would further improve the generalizability of the model. Overall, the integration of ANN and SMOTE in a web-based health information system provides a promising framework for advancing digital healthcare solutions in Indonesia. By improving the accuracy and efficiency of heart disease risk classification, such systems have the potential to contribute significantly to data-driven and patient-centered healthcare services.

Acknowledgement

The authors would like to express their sincere gratitude to *Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta* for the support and facilities provided throughout the research process. Special appreciation is also extended to the academic supervisor for the valuable guidance and constructive feedback in the preparation of this study entitled *"Optimization of SMOTE Implementation for Heart Disease Risk Classification Accuracy Using Artificial Neural Network."* The authors are equally thankful to all individuals who contributed, directly or indirectly, to the successful completion of this research and to the advancement of web-based classification systems in the healthcare domain. It is the authors' hope that the findings of this study may serve as a useful reference for both academics and practitioners in the fields of informatics and healthcare.

References

Aslam, F., & Malik, S. (2020). A review on PHP frameworks. *International Journal of Computer Applications, 175*(7), 1–5. https://doi.org/10.5120/ijca2020920154

BMC Medicine. (2024). Digital health transformation in low- and middle-income countries: Opportunities and challenges. *BMC Medicine.*

Kho, M. K. K., Lee, S., Park, J., & Kim, Y. (2021). Health information system for real-time patient monitoring. *Healthcare Informatics Research, 27*(1), 10–17. https://doi.org/10.4258/hir.2021.27.1.10

Kumari, V., & Singh, R. (2022). Application of SMOTE in medical data classification. *Journal of Artificial Intelligence Research, 12*(3), 55–66.

Liu, Q., Zhang, Y., Chen, H., & Wang, X. (2022). Improving medical classification with

SMOTE and deep learning models. *IEEE Access, 10,* 115432–115445. https://doi.org/10.1109/ACCESS.2022.3211111

Munassar, H., & Govardhan, A. (2011). A review on software process models. *International Journal of Computer Science and Information Technology, 3*(5), 224–228. https://doi.org/10.5121/ijcsit.2011.3517

Nugraha, D., Wahyudi, A., & Sari, R. (2022). Handling imbalanced data in medical prediction using SMOTE and Random Forest. *Procedia Computer Science, 197,* 567–574. https://doi.org/10.1016/j.procs.2022.01.189

Oikonomou, E. K., *et al*. (2023). Artificial neural networks in cardiovascular risk prediction: A systematic review. *Frontiers in Cardiovascular Medicine, 10,* 1–12.

Provost, F., Dhar, V., & Chen, J. (2025). Digital health ecosystems and predictive analytics in healthcare. *Journal of Medical Internet Research, 27*(4).

Putra, A., Nugroho, S., & Fadhilah, R. (2023). TabNet with SMOTE and hyperparameter tuning for medical prediction. *Journal of Big Data Analytics in Healthcare, 8*(2), 45–57.

Rahman, H., & Sari, R. (2023). Optimizing classification of imbalanced medical datasets using SMOTEENN and GridSearchCV. *Journal of Applied Intelligent Systems, 15*(1), 33–42. https://doi.org/10.1016/j.jais.2023.01.003

Smith, A., & Jones, B. (2022). Global trends in cardiovascular disease mortality. *Journal of Medical Sciences, 15*(2), 101–110.

Wahyudin, M. A., Nugraha, D., & Rahman, H. (2021). Web application development using CodeIgniter framework. *Jurnal Sistem Informasi, 6*(1), 44–50. https://doi.org/10.28932/jsi.v6i1.3309

World Health Organization. (2021). Cardiovascular diseases (CVDs). Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Zhang, L., Chen, Y., Wang, H., & Liu, Q. (2020). Digital health solutions for cardiovascular disease management. *IEEE Access, 8,* 13245–13256. https://doi.org/10.1109/ACCESS.2020.2964734