Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14347
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A Comparative Study of Ensemble Learning and Neural Networks for the Heart Disease Prediction

Gregorius Airlangga[1, *)], Oskar Ika Adi Nugroho[2)], Bobi Hartanto Pramudita Lim[3)]
[1)]Information System Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia,
[2)]Electrical Engineering Department, National Chung Cheng University, Chiayi, Taiwan,
[3)]Computer Science Department, National Chung Cheng University, Chiayi, Taiwan
[1,*)]gregorius.airlangga@atmajaya.ac.id, [2)]oskar@alum.ccu.edu.tw, [3)]lgl107u@alum.ccu.edu.tw

Submitted :Dec 15, 2024 | **Accepted** : Jan 22, 2025 | **Published** : Feb 10, 2025

Abstract: Heart disease continues to be a leading global cause of death, making the development of predictive models for early diagnosis a critical task. This study investigates the performance of various machine learning and deep learning models for heart disease prediction using a structured dataset of 918 observations and 11 features. The analysis includes ensemble methods like Random Forest, Gradient Boosting, and XGBoost, as well as neural networks such as Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs). Traditional classifiers, including Support Vector Machines (SVM) and Logistic Regression, are also considered for benchmarking. The dataset was preprocessed using label encoding, standardization, and the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance and ensure data consistency. Model evaluation was conducted using key metrics such as precision, recall, F1-score, and ROC-AUC. The results demonstrated that ensemble methods, particularly Random Forest (ROC-AUC: 0.9313) and Gradient Boosting (ROC-AUC: 0.9279), consistently delivered superior performance. Among neural networks, MLPs showed promising results (ROC-AUC: 0.9232), outperforming CNNs, which were less effective in handling tabular data. Meanwhile, TabNet was found to be unsuitable for this dataset, as it significantly underperformed across all metrics. This research highlights the effectiveness of ensemble methods and MLPs in heart disease prediction and the importance of proper preprocessing techniques. Future work could focus on integrating hybrid models or advanced optimization techniques to further enhance predictive accuracy in clinical settings.

Keywords: Heart Disease Prediction; Ensemble Learning; Neural Networks; Machine Learning Models; Data Preprocessing

INTRODUCTION

Cardiovascular diseases (CVDs) are the leading global cause of mortality, accounting for approximately 17.9 million deaths annually, or 31% of all global deaths (Bachheti et al., 2022; Hussain et al., 2024). Among these, heart attacks and strokes represent the predominant contributors, with one-third of these deaths occurring prematurely in individuals under the age of 70 (Jan et al., 2024; Ritchey, Wall, George, & Wright, 2020; Yahya et al., 2020). Heart failure, a critical and often fatal condition, emerges as a frequent consequence of CVDs, emphasizing the urgency for effective predictive mechanisms (Lourida & Louridas, 2022). Early detection and management of heart failure are imperative for reducing mortality rates and enhancing the quality of life for patients (Sapna et al., 2023). Machine learning has demonstrated immense potential in addressing this challenge by providing data-driven, accurate, and scalable solutions (Ahmad, Madonski, Zhang, Huang, & Mujeeb, 2022). Recent advancements in machine learning have paved the way for significant progress in disease prediction and healthcare diagnostics (Asif et al., 2024). Predictive models such as Random Forests, Gradient Boosting, Support Vector Machines, and deep learning techniques like Convolutional Neural Networks and Recurrent Neural Networks have been employed in various domains of medical diagnostics (Abdollahi, Nouri-Moghaddam, & Ghazanfari, 2021; Bhavsar et al., 2021; Ogunpola, Saeed, Basurra, Albarrak, & Qasem, 2024). Specifically, in prediction of heart disease, traditional statistical approaches such as logistic regression have been supplemented or replaced by these advanced methods due to their ability to capture complex, non-linear relationships in the data

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14347
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(Miller, Panneerselvam, & Liu, 2022; Soltani & Lee, 2024; Zhou, Yan, & Zhang, 2024). However, existing research often suffers from limitations. Many studies rely on smaller datasets, which limit the generalizability of their findings (Gangwal, Ansari, Ahmad, Azad, & Sulaiman, 2024; Goyal & Mahmoud, 2024; Tang et al., 2024). Model comparisons are frequently conducted without considering critical aspects such as hyperparameter tuning, cross-validation, and appropriate handling of imbalanced datasets (Montesinos López, Montesinos López, & Crossa, 2022). The choice of evaluation metrics also varies widely, complicating the comparison of performance across studies. Furthermore, the lack of explainability in advanced models remains a significant barrier to their adoption in clinical practice (Amann et al., 2020). The integration of hybrid modeling approaches, such as ensemble techniques and deep learning architectures, has shown potential to address these limitations (Rane, Choudhary, & Rane, 2024). Models like XGBoost and LightGBM, known for their computational efficiency and high accuracy, have become popular choices for structured tabular data (Shehadeh, Alshboul, Al Mamlook, & Hamedat, 2021). Similarly, CNNs and Multi-Layer Perceptrons are frequently employed for feature extraction and classification tasks in unstructured or high-dimensional datasets (Anitha, Varshini, Mahalakshmi, & Jishnu, 2024). Nonetheless, the effectiveness of these models in heart disease prediction using comprehensive datasets such as the one employed in this study has not been extensively explored.

Despite the wealth of research in CVD prediction, a substantial gap exists in creating a unified framework that systematically compares the performance of state-of-the-art machine learning and deep learning models using rigorous evaluation techniques. Previous studies have often employed limited datasets or focused on a narrow range of algorithms, which restricts the applicability of their findings in real-world clinical scenarios (Osaba et al., 2021). Moreover, the lack of focus on domain-specific preprocessing techniques, such as handling categorical and numerical features effectively, further diminishes the predictive power of these models (Mumuni & Mumuni, 2024). Another critical limitation is the underutilization of specialized loss functions tailored to address class imbalances, a common issue in healthcare datasets (Wang, 2023). Traditional binary cross-entropy loss may not sufficiently penalize models for misclassifying minority classes, which can lead to suboptimal decision boundaries (Hossain, Betts, & Paplinski, 2021). In this context, employing loss functions such as focal loss can significantly improve the sensitivity of predictive models. This research seeks to address these gaps by proposing a comparative study for heart disease prediction. The study evaluates traditional machine learning models, including Random Forest, Gradient Boosting, Support Vector Machines, Logistic Regression, and XGBoost, alongside advanced deep learning architectures such as Multi-Layer Perceptrons and Convolutional Neural Networks, on a unified dataset. Rigorous cross-validation techniques, including stratified k-fold validation, are employed to ensure robust performance evaluation across models. The study explores the impact of specialized loss functions, such as focal loss, on the performance of deep learning models in handling imbalanced data and utilizes neural architecture search frameworks to identify optimal configurations for Multi-Layer Perceptrons and Convolutional Neural Networks. It incorporates advanced models like TabNet, which have shown promise in tabular data analysis, and benchmarks their performance against other architectures. Comprehensive visualizations of performance metrics, including precision, recall, F1 score, and ROC-AUC, facilitate an in-depth understanding of the strengths and limitations of each architecture. The remainder of this article is organized into several sections. The next section describes the materials and methods, including dataset preprocessing, model architectures, and evaluation metrics. This is followed by the results and discussion, which analyze and interpret the performance of various models. Finally, the conclusion highlights the key findings, clinical implications, and potential directions for future research. By addressing existing gaps and offering a robust framework for heart disease prediction, this study contributes significantly to advancing machine learning applications in healthcare.

LITERATURE REVIEW

Research into heart failure prediction has evolved significantly, transitioning from traditional statistical approaches to advanced machine learning and deep learning techniques (Olsen, Mentz, Anstrom, Page, & Patel, 2020). Early studies predominantly relied on logistic regression due to its simplicity and interpretability (La Cava et al., 2023). These models, however, are constrained by their assumptions of linear relationships and the independence of predictors, which are often violated in complex medical datasets (Leeuwenberg et al., 2022). For instance, studies using logistic regression for heart disease prediction reported limited performance when applied to datasets with high variability in demographic and clinical features, underscoring the need for more flexible and robust approaches (Islam, Majumder, Miah, & Jannaty, 2024). Ensemble learning methods, including Random Forests and Gradient Boosting algorithms such as XGBoost and LightGBM, have emerged as powerful alternatives (Mienye & Sun, 2022). Random Forests aggregate the predictions of multiple decision trees, offering improved robustness against overfitting and providing insights into feature importance (Aria, Cuccurullo, & Gnasso, 2021). Studies applying Random Forests to heart failure prediction have demonstrated their ability to model non-linear relationships and interactions among features, yielding high predictive accuracy (Khan, Anwar, & Sikandar, 2023). Similarly, Gradient Boosting methods iteratively refine predictions by minimizing residual errors, often outperforming traditional statistical models in structured datasets (V. Kumar, Kedam, Sharma, Mehta, & Caloiero,

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

2023). For example, studies using XGBoost for heart failure prediction have reported significant gains in precision and recall, particularly when hyperparameter optimization was employed. However, these methods are not without limitations; their computational complexity increases with large datasets and the need for extensive tuning of hyperparameters (Bischl et al., 2023).

Deep learning, a subset of machine learning, has introduced transformative capabilities in handling high-dimensional and unstructured data (Malekloo, Ozer, AlHamaydeh, & Girolami, 2022). Models such as Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs) have gained traction in healthcare applications due to their ability to capture intricate patterns in data. MLPs have been effectively applied to structured datasets, including electronic health records, where they have demonstrated superior performance compared to traditional machine learning models (Xie et al., 2022). CNNs, while originally developed for image recognition, have been adapted for sequential and tabular data, enabling the identification of subtle patterns that may be indicative of heart failure (Petmezas et al., 2024). Despite their advantages, deep learning models often require large training datasets to avoid overfitting, and their "black-box" nature poses challenges for clinical interpretability (Hassija et al., 2023). A growing body of comparative studies has attempted to evaluate the relative performance of ensemble methods and deep learning models in heart disease prediction. Ensemble methods are frequently favored for their interpretability and computational efficiency, making them well-suited for tabular datasets (Amekoe, Azzag, Dagdia, Lebbah, & Jaffre, 2024). Deep learning models, on the other hand, excel in capturing complex, non-linear interactions but often come with increased computational demands and reduced transparency (S. Kumar et al., 2023). Few studies, however, have systematically compared these approaches within the context of heart failure prediction under consistent experimental conditions, highlighting a gap in literature (Ahsan & Siddique, 2022; Shah et al., 2020).

Another critical consideration in prediction of heart failure is the issue of class imbalance, which arises when datasets contain a disproportionate number of positive and negative cases. This imbalance can skew model predictions toward the majority class, reducing sensitivity to the minority class (Brownlee, 2020). Traditional evaluation metrics, such as accuracy, fail to adequately reflect model performance in such scenarios. Techniques like Synthetic Minority Oversampling Technique (SMOTE) and loss functions such as focal loss have been proposed to address this challenge (Alkhawaldeh, Albalkhi, & Naswhan, 2023). While focal loss has shown promise in improving sensitivity by assigning higher weights to hard-to-classify examples, its application in heart failure prediction remains limited. Preprocessing techniques also play a pivotal role in enhancing model performance (Heidari et al., 2020). Effective handling of categorical variables, feature scaling, and imputation of missing values are critical for ensuring the accuracy and reliability of predictive models. Studies that neglect these steps often report suboptimal results, underscoring the importance of a comprehensive preprocessing pipeline (Gupta, Sehgal, & Acken, 2024). Similarly, the choice of evaluation metrics is crucial for meaningful comparisons. Metrics such as precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC) provide a more nuanced assessment of model performance compared to accuracy alone (Imani & Arabnia, 2023).

An emerging area of interest is the integration of advanced optimization techniques, such as neural architecture search (NAS), to automate the design of deep learning models. NAS has demonstrated significant potential in identifying optimal architectures and hyperparameters, reducing the reliance on manual trial-and-error approaches. Despite its success in other domains, the application of NAS in heart failure prediction remains underexplored, offering an avenue for future research. The literature reviewed highlights the need for a systematic comparative analysis of ensemble learning and deep learning approaches for heart failure prediction. While both paradigms have demonstrated individual strengths, their relative performance under consistent experimental conditions remains unclear. This study aims to address this gap by leveraging a comprehensive dataset and employing rigorous preprocessing, advanced loss functions, and consistent evaluation metrics. By systematically comparing these approaches, this research seeks to provide actionable insights into the most effective strategies for heart failure prediction, ultimately contributing to improved clinical decision-making and patient outcomes.

METHOD

The research methodology adopted in this study involves a structured workflow to evaluate the effectiveness of machine learning and deep learning models for heart disease prediction. It encompasses dataset preprocessing, model development, hyperparameter tuning, and rigorous evaluation metrics, ensuring reproducibility and robustness. The dataset used in this study can be downloaded from (fedesoriano, 2021) and contains 918 observations with 11 features. Each observation is represented as $(D = \{(x_i, y_i)\}_{i=1}^{n})$, where $(x_i \in R^m)$ is the (m)-dimensional feature vector, and $(y_i \in \{0,1\})$ is the binary target variable. Here, $(y_i = 1)$ indicates the presence of heart disease, while $(y_i = 0)$ indicates its absence.

The preprocessing stage begins by encoding categorical variables, including Sex, ChestPainType, RestingECG, ExerciseAngina, and ST_Slope, using label encoding. Each unique category (c) is mapped to an integer $(l_c \in \{0,1,\dots,|C|-1\})$. This transformation ensures that categorical data is compatible with machine learning algorithms. For numerical features such as Age, RestingBP, Cholesterol, MaxHR, and Oldpeak,

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14347
e-ISSN : 2541-2019
p-ISSN : 2541-044X

standardization is applied to achieve a mean of 0 and a standard deviation of 1. The transformation is expressed as $x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$, where (μ_j) and (σ_j) are the mean and standard deviation of feature (j), respectively. This standardization prevents numerical features with larger magnitudes from dominating the model training process.

To address class imbalance in the target variable, Synthetic Minority Oversampling Technique (SMOTE) is employed. SMOTE generates synthetic samples for the minority class by interpolating between two existing minority samples, (x_a) and (x_b). A new synthetic sample (x_{new}) is created using the formula $x_{\text{new}} = x_a + \lambda(x_b - x_a)$, where $(\lambda \sim \text{Uniform}(0,1))$. This process balances the dataset, enhancing the model's ability to generalize across both classes.

The study evaluates a variety of machine learning and deep learning models. Traditional models include Random Forests, Gradient Boosting, Logistic Regression, Support Vector Machines (SVM), and XGBoost. In a Random Forest, predictions are aggregated from an ensemble of decision trees, with the final output determined by majority voting. The ensemble prediction is given as $\hat{y} = \text{mode}(\{y_t(x_i)\}_{t=1}^{T})$, where (T) is the total number of trees. Gradient Boosting iteratively refines predictions by minimizing errors. At iteration (t), the updated prediction is $\widehat{y^{(t)}} = \widehat{y^{(t-1)}} + \eta f^{(t)}(x)$, where (η) is the learning rate, and $(f^{(t)}(x))$ is the weak learner at step (t). Deep learning models include Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs). An MLP consists of multiple layers, where each layer transforms its input using a linear operation followed by a non-linear activation. For layer (l), the transformation is $a^{(l)} = \sigma\left(W^{(l)}a^{(l-1)} + b^{(l)}\right)$, where $(W^{(l)})$ and $(b^{(l)})$ are the weight matrix and bias vector for the layer, and $(\sigma(\cdot))$ is the activation function. CNNs apply convolutional operations to extract spatial hierarchies in the data. For an input feature map (X) and a convolutional filter (F) of size $(k \times k)$, the convolution operation at position $((i,j))$ is $Y_{ij} = \sum_{u=1}^{k}\sum_{v=1}^{k} F_{uv}X_{i+u-1,j+v-1}$.

Loss functions are critical for training both traditional and deep learning models. Binary cross-entropy is the primary loss function used in this study, defined as: $L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$. For imbalanced datasets, focal loss is employed to focus on hard-to-classify samples. It is given by $\mathcal{L}_{focal} = -\alpha(1 - \hat{y}_i)^{\gamma}\log(\hat{y}_i)$, where (α) and (γ) are hyperparameters controlling the weight and focus on challenging samples. Hyperparameter tuning is performed using Neural Architecture Search (NAS) to optimize hyperparameters such as learning rate (η), dropout rate (p), and the number of layers (L). The optimization process is formulated as $\min_{\Theta} \mathcal{L}(\Theta; X, y)$, where (Θ) represents the hyperparameters being tuned. Model evaluation is conducted using stratified (K)-fold cross-validation to ensure class distributions are preserved across folds. For each fold (k), the dataset is split into training and testing subsets: $\{(X_{\text{train}}^{(k)}, y_{\text{train}}^{(k)}), (X_{\text{test}}^{(k)}, y_{\text{test}}^{(k)})\}_{k=1}^{K}$. The models are evaluated using metrics such as precision, recall, F1-score, and ROC-AUC. Precision is defined as $P = \frac{TP}{TP+FP}$, recall as: $R = \frac{TP}{TP+FN}$, and F1-score as: $F1 = \frac{2PR}{P+R}$. Here, (TP), (FP), and (FN) denote true positives, false positives, and false negatives, respectively. The ROC-AUC evaluates the model's ability to distinguish between positive and negative classes, providing a comprehensive performance measure.

RESULT

The evaluation of multiple machine learning and deep learning architectures for heart disease prediction reveals distinct differences in their performance across precision, recall, F1-score, and ROC-AUC metrics as presented in the table 1 and figure 1-4. Each model's performance highlights its strengths and weaknesses in addressing the predictive task. MLP with Focal Loss achieved a precision of 0.8697, recall of 0.8860, F1-score of 0.8763, and ROC-AUC of 0.9221. This model demonstrates a strong balance between sensitivity and specificity, indicating its reliability in identifying positive cases. MLP with Binary Crossentropy slightly outperformed its focal loss counterpart, with a precision of 0.8756, recall of 0.8782, F1-score of 0.8758, and an ROC-AUC of 0.9232. The slight improvement suggests its effectiveness in optimizing the classification boundaries during training.

Table 1. Algorithm Performance

Model	Precision	Recall	F1 Score	ROC-AUC
MLP with Focal Loss	0.8697	0.886	0.8763	0.9221
MLP with Binary Crossentropy	0.8756	0.8782	0.8758	0.9232
CNN with Focal Loss	0.8632	0.8939	0.877	0.9174
CNN with Binary Crossentropy	0.8543	0.8644	0.8575	0.9051

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14347
e-ISSN : 2541-2019
p-ISSN : 2541-044X

TabNet	0.4817	0.1046	0.1684	0.5408
Random Forest	0.868	0.9016	0.8832	0.9313
Gradient Boosting	0.8722	0.8996	0.8851	0.9279
SVM	0.8664	0.9076	0.8855	0.9264
Logistic Regression	0.8603	0.8742	0.8658	0.9117
XGBoost	0.865	0.8997	0.8813	0.9228

CNN with Focal Loss recorded competitive results, with a precision of 0.8632, recall of 0.8939, F1-score of 0.8770, and an ROC-AUC of 0.9174. While its recall value is high, indicating good sensitivity, its overall performance remained slightly behind the MLP models. CNN with Binary Crossentropy demonstrated lower performance, achieving an ROC-AUC of 0.9051 and an F1-score of 0.8575, suggesting that its ability to generalize on tabular data was somewhat limited. On the other hand, TabNet significantly underperformed in comparison to other models, achieving a precision of 0.4817, recall of 0.1046, F1-score of 0.1684, and ROC-AUC of 0.5408. The low values across all metrics indicate that TabNet struggled to effectively capture the underlying relationships in the dataset, highlighting its unsuitability for this task.

Random Forest emerged as the top-performing traditional machine learning model, achieving a precision of 0.8680, recall of 0.9016, F1-score of 0.8832, and an ROC-AUC of 0.9313. Its robust performance underscores its ability to capture complex feature interactions and reduce overfitting through ensemble learning. Gradient Boosting followed closely, with a precision of 0.8722, recall of 0.8996, F1-score of 0.8851, and ROC-AUC of 0.9279. Next, Support Vector Machines (SVM) delivered competitive results, achieving a precision of 0.8664, recall of 0.9076, F1-score of 0.8855, and ROC-AUC of 0.9264. Logistic Regression, while simpler in design, performed well with a precision of 0.8603, recall of 0.8742, F1-score of 0.8658, and ROC-AUC of 0.9117. XGBoost, as a variant of Gradient Boosting, achieved strong metrics with a precision of 0.8650, recall of 0.8997, F1-score of 0.8813, and an ROC-AUC of 0.9228.

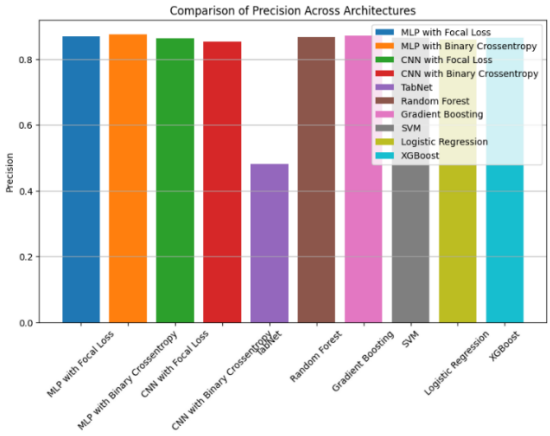


Fig. 1 Precision Results

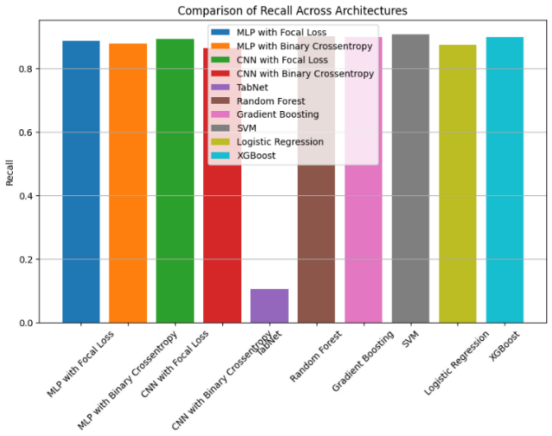


Fig. 2 Recall Results

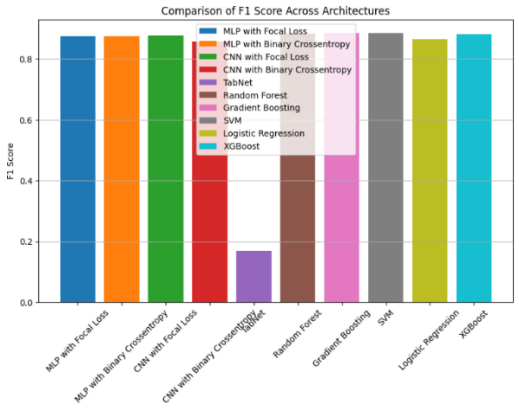
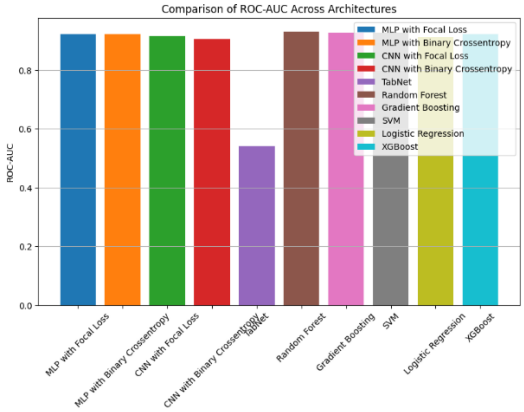


Fig. 3 F1 Results F



ig. 4 ROC-AUC Results

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

The results clearly indicate that ensemble methods, particularly Random Forest and Gradient Boosting, achieved the highest performance across all metrics. Random Forest achieved the highest ROC-AUC value of 0.9313, reflecting its robustness in learning complex patterns while reducing overfitting. Gradient Boosting followed closely behind, demonstrating its effectiveness in iteratively minimizing prediction errors. XGBoost, which builds upon the principles of Gradient Boosting, also achieved strong results, making it a scalable and efficient choice for structured data tasks. Among neural networks, MLP with Binary Crossentropy emerged as the most effective deep learning model, achieving an ROC-AUC of 0.9232. This result highlights the adaptability of fully connected neural networks to tabular data when appropriate preprocessing is applied. While Focal Loss is designed to address class imbalance, the minimal difference in performance between Focal Loss and Binary Crossentropy suggests that the use of SMOTE during preprocessing sufficiently mitigated the imbalance in the dataset. CNNs, while competitive, underperformed compared to MLPs and ensemble methods. The lower performance can be attributed to the lack of spatial hierarchies in the dataset, which limits CNNs' ability to extract meaningful features.

TabNet, despite its innovative approach to sequential feature selection, struggled significantly with this dataset. Its poor performance across precision, recall, and ROC-AUC indicates that it may not be suitable for tabular datasets without temporal or sequential dependencies. This result highlights the importance of selecting models that align with the dataset's structure and characteristics. Support Vector Machines demonstrated strong performance, achieving an ROC-AUC of 0.9264. SVM's ability to construct optimal decision boundaries using kernel methods enables it to perform well on structured data, particularly when nonlinear relationships exist. However, its computational complexity may limit its scalability for larger datasets. Logistic Regression, while less accurate compared to ensemble methods and SVM, remains an interpretable and efficient model for initial explorations or real-world clinical applications where transparency is critical.

The comparison between Focal Loss and Binary Crossentropy in neural networks further emphasizes that loss function selection plays a limited role when preprocessing steps, such as SMOTE, address class imbalance. While Focal Loss provided a slight improvement in recall, Binary Crossentropy achieved comparable results, making it a simpler and effective alternative. The practical implications of these findings suggest that ensemble methods, particularly Random Forest and Gradient Boosting, are the most reliable choices for real-world deployment due to their superior accuracy and robustness. Neural networks, particularly MLPs, provide effective alternatives for integrating into deep learning pipelines. Logistic Regression remains a viable option for applications prioritizing simplicity and interpretability. The choice of the most suitable model ultimately depends on the specific requirements of the application, such as the need for interpretability, accuracy, or computational efficiency.

CONCLUSION

This study highlights the comparative strengths of machine learning and deep learning models for heart disease prediction. Ensemble methods, particularly Random Forest and Gradient Boosting, emerged as the most accurate and robust approaches, demonstrating their ability to capture complex feature interactions while maintaining reliability. Among neural networks, Multi-Layer Perceptron (MLP) with Binary Crossentropy showed strong performance, proving its adaptability to structured data when combined with appropriate preprocessing. While Focal Loss improved recall slightly, its contribution was less pronounced due to effective preprocessing using SMOTE. The findings emphasize the importance of model selection based on the structure and characteristics of the dataset. Ensemble methods are ideal for achieving high accuracy and robustness, while simpler models like Logistic Regression provide interpretable and computationally efficient alternatives. Neural networks remain promising for more advanced pipelines, particularly when scalability and integration into larger systems are required. Future work can focus on expanding the dataset and incorporating advanced hybrid models to further improve predictive performance. Additionally, exploring domain-specific feature engineering and optimization techniques could enhance the applicability of these models in clinical settings. The insights from this study provide a foundation for selecting suitable machine learning approaches, balancing accuracy, interpretability, and complexity to meet real-world requirements for heart disease prediction.

REFERENCES

Abdollahi, J., Nouri-Moghaddam, B., & Ghazanfari, M. (2021). Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases). *ArXiv Preprint ArXiv:2103.08182*.

Ahmad, T., Madonski, R., Zhang, D., Huang, C., & Mujeeb, A. (2022). Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renewable and Sustainable Energy Reviews*, *160*, 112128.

Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, *128*, 102289.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14347
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Alkhawaldeh, I. M., Albalkhi, I., & Naswhan, A. J. (2023). Challenges and limitations of synthetic minority oversampling techniques in machine learning. *World Journal of Methodology*, *13*(5), 373.

Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., & Consortium, P. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*, 1–9.

Amekoe, K. M., Azzag, H., Dagdia, Z. C., Lebbah, M., & Jaffre, G. (2024). Exploring accuracy and interpretability trade-off in tabular learning with novel attention-based models. *Neural Computing and Applications*, *36*(30), 18583–18611.

Anitha, S., Varshini, E. K., Mahalakshmi, N. H., & Jishnu, S. (2024). Optimizing Multi-Class Text Classification Models for Imbalanced News Data. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–6.

Aria, M., Cuccurullo, C., & Gnasso, A. (2021). A comparison among interpretative proposals for Random Forests. *Machine Learning with Applications*, *6*, 100094.

Asif, S., Wenhui, Y., ur-Rehman, S.-, ul-ain, Q.-, Amjad, K., Yueyang, Y., … Awais, M. (2024). Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision. *Archives of Computational Methods in Engineering*, 1–31.

Bachheti, R. K., Worku, L. A., Gonfa, Y. H., Zebeaman, M., Deepti, Pandey, D. P., & Bachheti, A. (2022). [Retracted] Prevention and Treatment of Cardiovascular Diseases with Plant Phytochemicals: A Review. *Evidence-Based Complementary and Alternative Medicine*, *2022*(1), 5741198.

Bhavsar, K. A., Abugabah, A., Singla, J., AlZubi, A. A., Bashir, A. K., & others. (2021). A comprehensive review on medical diagnosis using machine learning. *Computers, Materials and Continua*, *67*(2), 1997.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., … others. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(2), e1484.

Brownlee, J. (2020). *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. Machine Learning Mastery.

fedesoriano. (2021). *Heart Failure Prediction Dataset*. Retrieved from https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data

Gangwal, A., Ansari, A., Ahmad, I., Azad, A. K., & Sulaiman, W. M. A. W. (2024). Current strategies to address data scarcity in artificial intelligence-based drug discovery: A comprehensive review. *Computers in Biology and Medicine*, *179*, 108734.

Goyal, M., & Mahmoud, Q. H. (2024). A systematic review of synthetic data generation techniques using generative AI. *Electronics*, *13*(17), 3509.

Gupta, P., Sehgal, N. K., & Acken, J. M. (2024). Practical Aspects in Machine Learning. In *Introduction to Machine Learning with Security: Theory and Practice Using Python in the Cloud* (pp. 281–330). Springer.

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., … Hussain, A. (2023). Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, 1–30.

Heidari, M., Mirniaharikandehei, S., Khuzani, A. Z., Danala, G., Qiu, Y., & Zheng, B. (2020). Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. *International Journal of Medical Informatics*, *144*, 104284.

Hossain, M. S., Betts, J. M., & Paplinski, A. P. (2021). Dual focal loss to address class imbalance in semantic segmentation. *Neurocomputing*, *462*, 69–87.

Hussain, A., Li, S., Hussain, T., Lin, X., Ali, F., & AlZubi, A. A. (2024). Computing Challenges of UAV Networks: A Comprehensive Survey. *Computers, Materials \& Continua*, *81*(2).

Imani, M., & Arabnia, H. R. (2023). Hyperparameter optimization and combined data sampling techniques in machine learning for customer churn prediction: a comparative analysis. *Technologies*, *11*(6), 167.

Islam, M. A., Majumder, M. Z. H., Miah, M. S., & Jannaty, S. (2024). Precision healthcare: A deep dive into machine learning algorithms and feature selection strategies for accurate heart disease prediction. *Computers in Biology and Medicine*, *176*, 108432.

Jan, B., Dar, M. I., Choudhary, B., Basist, P., Khan, R., & Alhalmi, A. (2024). Cardiovascular diseases among Indian older adults: A comprehensive review. *Cardiovascular Therapeutics*, *2024*(1), 6894693.

Khan, Z., Anwar, S., & Sikandar, G. (2023). Heart Disease Prediction Using Hybrid Random Forest and Linear Model. *International Journal of Emerging Engineering and Technology*, *2*(1), 6–12.

Kumar, S., Guruparan, D., Aaron, P., Telajan, P., Mahadevan, K., Davagandhi, D., & Yue, O. X. (2023). Deep learning in computational biology: Advancements, challenges, and future outlook. *ArXiv Preprint ArXiv:2310.03086*.

Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., & Caloiero, T. (2023). Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. *Water*, *15*(14), 2572.

*name of corresponding author

La Cava, W. G., Lee, P. C., Ajmal, I., Ding, X., Solanki, P., Cohen, J. B., … Herman, D. S. (2023). A flexible symbolic regression method for constructing interpretable clinical prediction models. *NPJ Digital Medicine*, *6*(1), 107.

Leeuwenberg, A. M., van Smeden, M., Langendijk, J. A., van der Schaaf, A., Mauer, M. E., Moons, K. G. M., … Schuit, E. (2022). Performance of binary prediction models in high-correlation low-dimensional settings: a comparison of methods. *Diagnostic and Prognostic Research*, *6*(1), 1.

Lourida, K. G., & Louridas, G. E. (2022). Clinical Phenotypes of Cardiovascular and Heart Failure Diseases Can Be Reversed? The Holistic Principle of Systems Biology in Multifaceted Heart Diseases. *Cardiogenetics*, *12*(2), 142–169.

Malekloo, A., Ozer, E., AlHamaydeh, M., & Girolami, M. (2022). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*, *21*(4), 1906–1955.

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, *10*, 99129–99149.

Miller, A., Panneerselvam, J., & Liu, L. (2022). A review of regression and classification techniques for analysis of common and rare variants and gene-environmental factors. *Neurocomputing*, *489*, 466–485.

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In *Multivariate statistical machine learning methods for genomic prediction* (pp. 109–139). Springer.

Mumuni, A., & Mumuni, F. (2024). Automated data processing and feature engineering for deep learning and big data applications: a survey. *Journal of Information and Intelligence*.

Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics*, *14*(2), 144.

Olsen, C. R., Mentz, R. J., Anstrom, K. J., Page, D., & Patel, P. A. (2020). Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *American Heart Journal*, *229*, 1–17.

Osaba, E., Villar-Rodriguez, E., Del Ser, J., Nebro, A. J., Molina, D., LaTorre, A., … Herrera, F. (2021). A tutorial on the design, experimentation and application of metaheuristic algorithms to real-world optimization problems. *Swarm and Evolutionary Computation*, *64*, 100888.

Petmezas, G., Papageorgiou, V. E., Vassilikos, V., Pagourelias, E., Tsaklidis, G., Katsaggelos, A. K., & Maglaveras, N. (2024). Recent advancements and applications of deep learning in heart failure: A systematic review. *Computers in Biology and Medicine*, 108557.

Rane, N., Choudhary, S., & Rane, J. (2024). Ensemble Deep Learning and Machine Learning: Applications, Opportunities, Challenges, and Future Directions. *Opportunities, Challenges, and Future Directions (May 31, 2024)*.

Ritchey, M. D., Wall, H. K., George, M. G., & Wright, J. S. (2020). US trends in premature heart disease mortality over the past 50 years: Where do we go from here? *Trends in Cardiovascular Medicine*, *30*(6), 364–374.

Sapna, F. N. U., Raveena, F. N. U., Chandio, M., Bai, K., Sayyar, M., Varrassi, G., … Mohamad, T. (2023). Advancements in heart failure management: a comprehensive narrative review of emerging therapies. *Cureus*, *15*(10).

Shah, S. J., Borlaug, B. A., Kitzman, D. W., McCulloch, A. D., Blaxall, B. C., Agarwal, R., … others. (2020). Research priorities for heart failure with preserved ejection fraction: national heart, lung, and blood institute working group summary. *Circulation*, *141*(12), 1001–1026.

Shehadeh, A., Alshboul, O., Al Mamlook, R. E., & Hamedat, O. (2021). Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. *Automation in Construction*, *129*, 103827.

Soltani, A., & Lee, C. L. (2024). The non-linear dynamics of South Australian regional housing markets: A machine learning approach. *Applied Geography*, *166*, 103248.

Tang, J., Yang, Y., Wei, W., Shi, L., Su, L., Cheng, S., … Huang, C. (2024). Graphgpt: Graph instruction tuning for large language models. *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 491–500.

Wang, Z. (2023). *Predictive Learning from Real-World Medical Data: Overcoming Quality Challenges*.

Xie, F., Yuan, H., Ning, Y., Ong, M. E. H., Feng, M., Hsu, W., … Liu, N. (2022). Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *Journal of Biomedical Informatics*, *126*, 103980.

Yahya, T., Jilani, M. H., Khan, S. U., Mszar, R., Hassan, S. Z., Blaha, M. J., … others. (2020). Stroke in young adults: Current trends, opportunities for prevention and pathways forward. *American Journal of Preventive Cardiology*, *3*, 100085.

Zhou, W., Yan, Z., & Zhang, L. (2024). A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction.

*name of corresponding author

Scientific Reports, *14*(1), 5905.

*name of corresponding author