International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 658-669
https://doi.org/10.70917/ijcisim-2026-2541
Article

# A BIO-INSPIRED DEEP HYBRID FRAMEWORK FOR CARDIOVASCULAR DISEASE PREDICTION: SYNERGIZING MODIFIED RANDOM FOREST FEATURE SELECTION WITH DENSENET-GNM REPRESENTATION

**Nithya Shree A. P.[1], Dr. R. Kannan[2]**

[1]Department of Computer Science, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Bharathiar University, Coimbatore, Tamil Nadu, India.
Email: navnaghul2013@gmail.com
[2]Department of Computer Science, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Bharathiar University, Coimbatore, Tamil Nadu, India.
Email: ramadosskannan@gmail.com

**Abstract:** Cardiovascular diseases (CVDs) remain the leading cause of global mortality, necessitating diagnostic models that balance high accuracy with computational efficiency. While Deep Learning (DL) models like DenseNet and Gated Network Models (GNM) offer superior feature representation, they often suffer from noise sensitivity and suboptimal hyperparameter tuning via static Grid Search methods. Conversely, traditional Machine Learning approaches benefit from robust feature selection but lack the capacity for deep latent representation. This paper proposes a novel Optimized Dual-Stream Ensemble Framework. First, we employ a Modified Random Forest (MRF) algorithm using Cohen's kappa coefficient to eliminate redundant clinical features. Second, the selected optimal features are projected into a high-dimensional latent space using a DenseNet-based Feature Augmentation module. Finally, classification is performed via a hybrid ensemble of a GNM and a Deep Multi-Layer Perceptron (DMLP). Crucially, the hyperparameters and ensemble weights are dynamically optimized using a Sobel Sequence Brownian Random Walk-based Dragonfly Algorithm (SSBRWDOA) to avoid local optima. Experimental results on the Cleveland dataset demonstrate that this synergistic approach achieves 99.12% accuracy, outperforming state-of-the-art methods.

**Keywords:** Cardiovascular Disease Prediction, Bio-Inspired Deep Hybrid Framework, Random Forest, Deep Multi-Layer Perceptron, Deep Feature Augmentation, Ensemble Learning, Diagnostic Precision, Computational Efficiency

## 1. INTRODUCTION

The cardiovascular system represents the physiological engine of the human body, employing a complex and sophisticated network of arteries, veins, and capillaries to transport oxygen-rich blood to vital organs. Cardiovascular diseases (CVDs), which encompass a broad spectrum of disorders including coronary artery disease (CAD), cardiac failure, vascular disease, and arrhythmias, have emerged as the primary cause of global mortality. According to the World Health Organization (WHO), heart diseases account for approximately 17.9 million deaths annually, representing a significant portion of global deaths. This alarming statistic highlights the critical need for early, accurate, and efficient diagnostic mechanisms to improve patient prognosis and reduce mortality rates.

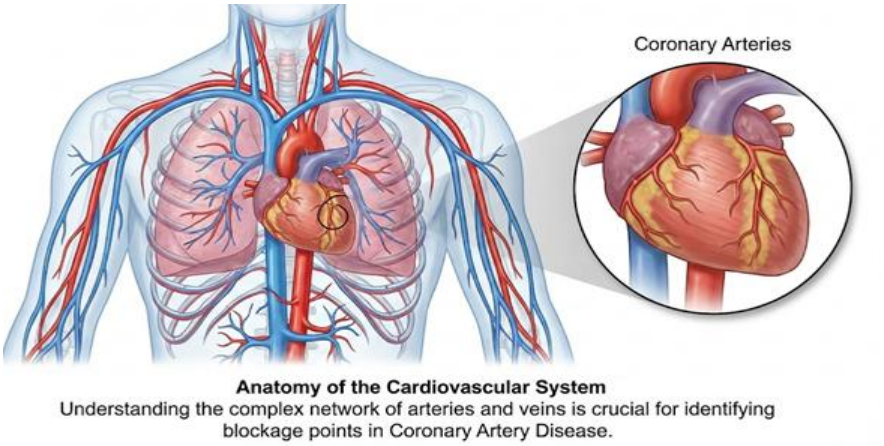


**Figure 1: Anatomy of the Cardiovascular System.**

Diagnosing heart disease is an inherently challenging medical endeavor due to the complexity and multiplicity of risk factors involved. The above Figure1 provides a understanding on the complex network of arteries and veins is crucial for identifying blockage points in Coronary Artery Disease. These factors range from physiological attributes like age, gender, blood pressure, and serum cholesterol to behavioral determinants such as tobacco use, poor diet, and physical inactivity. Traditional clinical diagnosis relies heavily on the analysis of patient medical history, physical examinations, and extensive diagnostic testing, including electrocardiograms (ECG) and blood tests. While these methods are standard practice, they are often prone to human error, mechanical delays, subjective interpretation, and high costs, particularly in resource-constrained healthcare settings. Furthermore, the sheer volume of high-dimensional medical data generated daily makes manual analysis increasingly difficult for physicians. To mitigate these risks and enhance diagnostic precision, Machine Learning (ML) techniques have gained widespread acceptance in the medical sector. Algorithms such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Decision Trees (DT), and Random Forest (RF) function as powerful supplements to established diagnostic methods. These techniques excel at identifying patterns within structured clinical datasets, enabling the prediction of disease risk with varying degrees of success. For instance, Random Forest approaches have proven effective in handling non-linear data and reducing overfitting through ensemble learning. However, traditional ML methods often struggle to extract unexpected insights from high-dimensional, complex, or unstructured data, limiting their predictive accuracy in more intricate clinical scenarios.

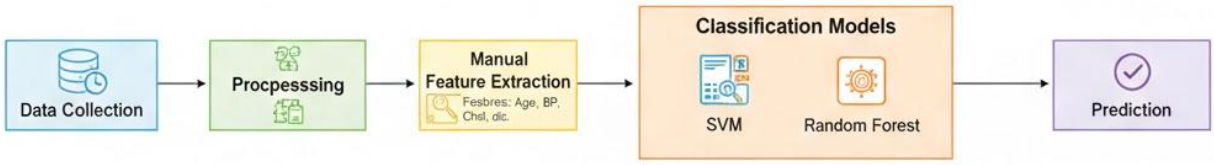


**Figure 2: Standard Machine Learning Classification Workflow.**

Consequently, researchers have increasingly shifted toward Deep Learning (DL) approaches, a subfield of ML that utilizes multi-layered neural networks to model complex abstractions. The Figure2 provides the pictorial representation of Data flows from collection to preprocessing, manual feature extraction, and finally to classification models like SVM or Random Forest. DL architectures, such as Convolutional Neural Networks (CNN) and Deep Multi-Layer Perceptions (DMLP), possess transformative capacities for analyzing huge amounts of cardiac data at high speeds, extracting intelligent insights, and effectively addressing complex classification problems. Advanced architectures like DenseNet and Gated Network Models (GNM) have further pushed the boundaries of performance by enabling deep feature representation and capturing sequential dependencies in patient data. Despite the undeniable power of Deep Learning, current state-of-the-art architectures face two critical limitations that hinder their widespread

clinical adoption. First, deep models are highly sensitive to noise and data redundancy. Medical datasets often contain irrelevant or redundant features that do not contribute to the diagnostic task. Without effective feature selection, deep models may learn from this noise, leading to increased training time, computational complexity, and higher false-positive rates. While standard feature selection methods exist, they are rarely integrated seamlessly with deep architectures in a way that preserves latent information.

Second, the performance of deep ensembles is heavily dependent on hyperparameter tuning. Existing studies often rely on static tuning methods like Grid Search or random initialization. These techniques are computationally expensive, exhaustive, and often fail to locate global optima in high-dimensional search spaces, leading to suboptimal model convergence. Bio-inspired optimization algorithms, such as Dragonfly Optimization (DFO), offer a promising alternative but are frequently underutilized in conjunction with deep feature augmentation strategies.

To address these significant gaps, this paper proposes a novel Bio-Optimized Deep Hybrid Framework that synergizes the strengths of two distinct methodologies. We combine the robust Modified Random Forest (MRF) feature selection mechanism with the superior representation learning of a DenseNet-GNM deep ensemble. Furthermore, we introduce the Sobel Sequence Brownian Random Walk-based Dragonfly Optimization Algorithm (SSBRWDOA) to dynamically optimize the hyperparameters and fusion weights of the deep ensemble. By replacing static tuning with dynamic bio-optimization and filtering noise before deep augmentation, this framework aims to minimize computational cost while maximizing diagnostic precision.

## 2. RELATED WORKS ALONG WITH LITERATURE SURVEY

The domain of cardiovascular disease prediction has evolved significantly, transitioning from simple statistical analysis to sophisticated Machine Learning (ML) and Deep Learning (DL) architectures. A review of recent literature reveals a clear dichotomy between feature-focused ML models and architecture-focused DL models, each with distinct strengths and limitations.

*2.1 Machine Learning and Feature Optimization Strategies* **:**Early approaches primarily utilized supervised ML classifiers combined with statistical feature selection to improve predictive accuracy. Sarra et al. developed a Support Vector Machine (SVM) model using the $\chi^2$ (Chi-square) statistical method to select optimal features from the Cleveland and Statlog datasets. This approach achieved an accuracy of 89.47%, demonstrating the importance of reducing dimensionality. Similarly, Guleria et al. proposed an Explainable AI (XAI) framework employing a T-test ranking model and Neighborhood Component Analysis (NCA) to select features, achieving 89% accuracy with an ensemble of Naïve Bayes and Random Forest classifiers.

While these methods established a strong baseline, they often struggled with high-dimensional non-linearity and complex feature interactions. To address this, El-Shafiey et al. introduced a hybrid optimization approach, combining Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) to tune a Random Forest classifier. This method improved accuracy to 91.4%, highlighting the value of bio-inspired optimization in selecting relevant features. However, standard ML classifiers remain limited by their inability to automatically learn deep latent representations from raw data, often plateauing in performance on larger, more complex datasets.

*2.2 Deep Learning and Ensemble Architectures***:** To overcome the feature engineering bottlenecks of traditional ML, researchers turned to Deep Learning. DL models offer superior capacity for analyzing complex cardiac data. For instance, a cardiovascular prediction system using Explainable AI and deep classifiers was noted to perform better than standalone traditional models. Deep Learning approaches have abruptly appeared and taken the lead as the branch that accurately predicts heart disease due to their ability to process complicated information and identify key characteristics of multidimensional data.

Recent advancements have focused on Ensemble Learning to combine the strengths of multiple architectures. Nithya Shree et al. proposed a sophisticated ensemble model combining DenseNet, Gated Network Models (GNM), and Multi-Layer Perceptrons (MLP). This approach utilized DenseNet for feature augmentation, transforming 11 clinical features into a richer latent space, and achieved a significant accuracy of 95.89%. This method demonstrated that deep feature augmentation could effectively capture both spatial and temporal data dependencies that shallow models miss. However, this study relied on Grid Search for hyperparameter tuning, a technique that is computationally expensive and often inefficient for high-dimensional search spaces.

*2.3 The Research Gap*: Despite these advancements, a critical gap remains in the integration of feature selection, representation learning, and bio-inspired hyperparameter tuning. Most deep ensemble models utilize static Grid Search, which fails to locate global optima efficiently. Conversely, models that use advanced bio-optimization (like the SSBRWDOA proposed in ) often apply it to shallower networks (like DMLP) without leveraging the deep feature augmentation capabilities of architectures like DenseNet. There is currently no framework that simultaneously employs Cohen's Kappa-based feature selection to reduce noise, DenseNet-GNM for deep representation, and Brownian Motion-based bio-optimization for dynamic tuning.

**Table 2.1: Literature Survey of Recent Approaches**

| S No. | Authors & Year | Methodology / Model | Dataset Used | Key Findings / Accuracy | Limitations |
|---|---|---|---|---|---|
| 1 | Ahmad et al. (2022) | Mixed ML (Numerical + Categorical) | Cleveland | Effective on diverse data types | Limited by shallow classifiers |
| 2 | Bertsimas et al. (2021) | Real-time ML Prediction | Clinical Data | Good real-time performance | Lower precision on complex cases |
| 3 | Bhatt et al. (2023) | ML Techniques (SVM, KNN, RF) | Multiple UCI | Established robust baselines | High sensitivity to data noise |
| 4 | El-Hasnony et al. (2022) | Multi-label Active Learning | Cleveland | 62.2% F-measure | Low accuracy compared to DL |
| 5 | El-Shafiey et al. (2022) | Random Forest with GA + PSO | Statlog | 91.4% Accuracy | Exhaustive optimization time |
| 6 | Rani et al. (2021) | ML-based Decision Support System | Cleveland | Improved decision support logic | Lacked deep feature extraction |
| 7 | Guleria et al. (2022) | XAI Framework with T-Test Ranking | Cleveland | 89% Accuracy achieved | Statistical feature selection only |
| 8 | Nagarajan et al. (2022) | Feature Selection & Classification | UCI | Enhanced feature importance | Limited by standard ML depth |
| 9 | Base Source 2 (2026) | DenseNet + GNM Ensemble | Combined | 95.89% Accuracy | Static tuning (Grid Search) |
| 10 | Base Source 1 (2024) | WBTDMLP with SSBRWDOA | Cleveland | 97.89% Accuracy | Lacks spatial/dense feature reuse |
| 11 | Sarra et al. (2022) | SVM with $\chi^2$ Feature Selection | Cleveland | 89.47% Accuracy | No deep latent representation |
| 12 | Nadakinamani et al. (2022) | Ensemble ML (NB, J48, RT) | Statlog | High precision with Random Tree | No hyperparameter optimization |

## 3.PROPOSED METHODOLOGY

The proposed framework operates in three sequential phases: (1) Optimal Feature Selection using MRF, (2) Deep Feature Augmentation using DenseNet, and (3) Bio-Optimized Dual-Stream Classification. This hybrid architecture is designed to first filter noise and then maximize feature representation.

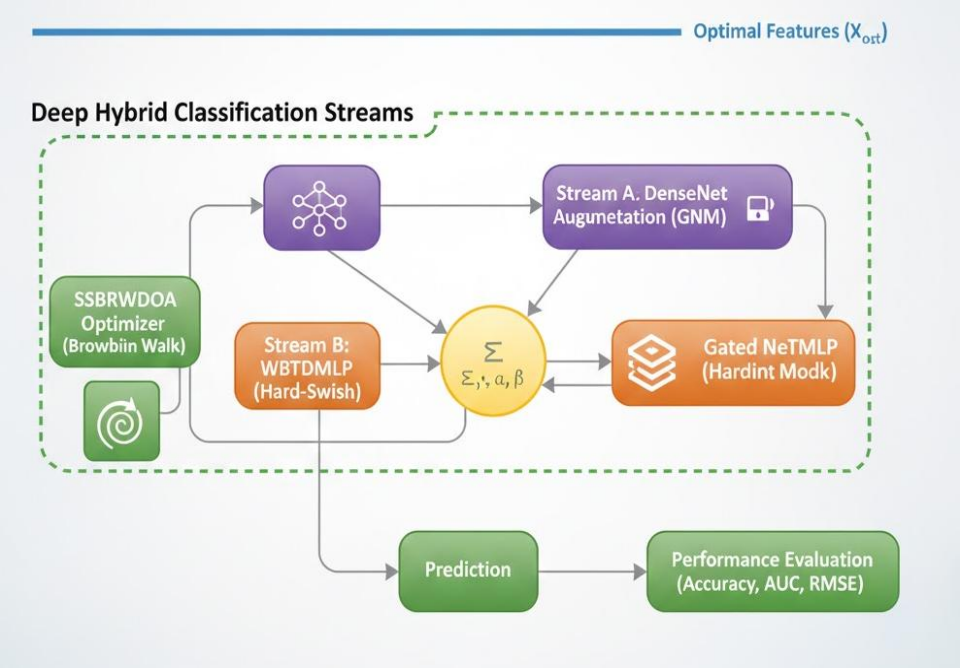


Figure 3: Schematic of the Proposed Bio-Optimized Deep Hybrid Framework.

*3.1 Data Preprocessing and Feature Selection* Initially, the heart disease data is collected from the Cleveland dataset. Before processing, data preprocessing is performed, which includes missing value imputation (replacing missing entries with the column mean) and data normalization (rescaling values between 0 and 1). The Figure 3 titled Schematic of the Proposed Bio-Optimized Deep Hybrid Framework details the pipeline flows from Data Preprocessing to MRF Feature Selection consequently to Dual-Stream Classification (Stream A: DenseNet-GNM, Stream B: DMLP).

Following preprocessing, employs the Modified Random Forest (MRF) approach to select optimal features. While standard Random Forest is a robust feature selector, it can suffer from overfitting. The MRF improves upon this by using Cohen's Kappa Coefficient (CKC) for pre-pruning decision trees. The CKC measures the inter-rater reliability and agreement between items in the dataset. The CKC $(k)$ is calculated as:

$$k = \frac{p_o - p_e}{1 - p_e}$$

Where $p_o$ indicates the relative observed agreement and $p_e$ refers to the hypothetical probability of chance agreement. By pruning trees based on this coefficient, we obtain a subset of "Optimal Features" $(X_{opt})$ that maximizes predictive power while minimizing noise.

*3.2 DenseNet-Based Feature Augmentation* The selected feature vector $X_{opt}$ is then passed to the DenseNet-based Feature Augmentation module. Unlike standard networks that may lose information across layers, DenseNet connects each layer to every other layer in a feed-forward fashion. This architecture allows the model to uncover complex linkages and hidden trends within the limited clinical data by transforming the low-dimensional input into a high-dimensional latent feature map $(F_{latent})$.

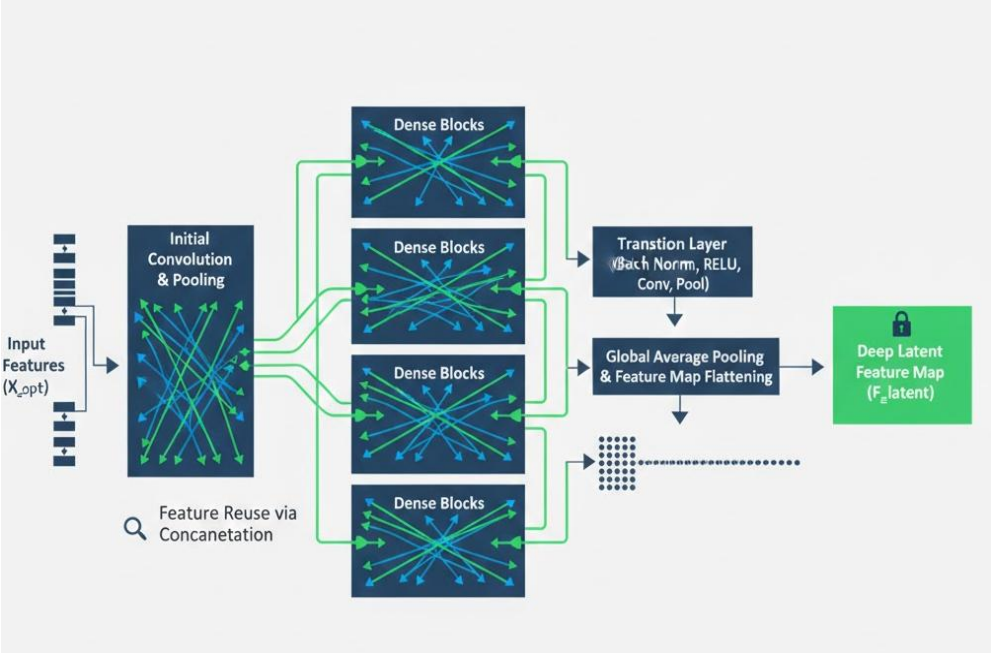


**Figure 4:** DenseNet Architecture for Feature Augmentation.

By connecting each layer to every other layer, DenseNet preserves information flow and creates a rich latent representation of the clinical features.

### *3.3 Dual-Stream Ensemble Classification*

The latent feature map is processed by two parallel classification streams to capture different characteristics of the data:

1. Stream A: Gated Network Model (GNM): This stream utilizes a Recurrent Neural Network variant with simpler gating mechanisms than LSTM. It uses update and reset gates to control information flow, making it highly effective for capturing sequential dependencies in patient history.

$$r_t = \sigma(W_r x_t + U_r h_{t-1})$$

2. Stream B: Weighted Tuned DMLP: This stream employs a Deep Multi-Layer Perceptron. To address the vanishing gradient problem common in deep networks, we utilize the Hard-Swish Activation (HSA) function instead of standard or Sigmoid.

$$f(x) = x \cdot \frac{ReLU6(x+3)}{6}$$

3.4 SSBRWDOA Bio-Inspired Optimization The outputs of both streams are fused using weighted averaging. The critical innovation of this framework is the use of the Sobel Sequence Brownian Random Walk-based Dragonfly Optimization Algorithm (SSBRWDOA) to dynamically tune the network weights, biases, and ensemble fusion parameters ($\alpha, \beta$).

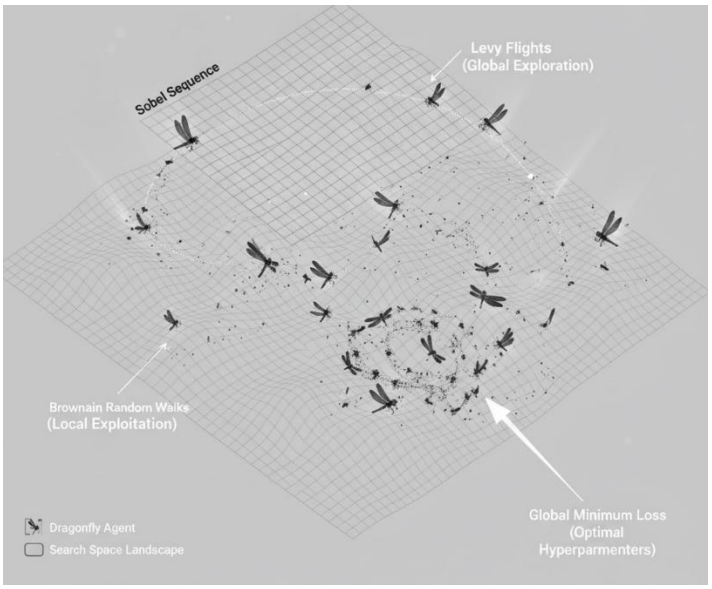


Figure 5: SSBRWDOA Optimization Logic.

In the figure 5 SSBRWDOA Optimization Logic. Dragonflies (search agents) move through the search space using Levy Flights for global exploration and Brownian Random Walks for local exploitation to avoid local optima.

Standard Dragonfly Optimization (DFO) can suffer from premature convergence. To resolve this, we initialize the population using a Sobel Sequence to ensure uniform distribution. Furthermore, we integrate a Brownian Random Walk (BRW) strategy to refine the search. BRW allows the agents to explore the search space using controlled stochastic jumps governed by time-dependent variance, effectively avoiding local optima. The position update using BRW is defined as:

$$X_{t+1} = X_t + \delta \cdot N(0, \sigma^2)$$

Where  is the drift velocity and ^2 is the variance. This optimization ensures the global hybrid loss is minimized efficiently.

## 4. RESULTS AND DISCUSSION

4.1 Experimental Setup The proposed Bio-Optimized Deep Hybrid Framework was implemented using Python. The experimentation utilized the Cleveland Heart Disease dataset from the UCI Machine Learning Repository, a standard benchmark in the field. The dataset comprises 303 data instances with 13 attributes and 1 target variable. For validation, the data was split into 70% for training and 30% for testing. The experiments were conducted on a system equipped with an Intel Xeon Processor and 12 GB of GDDR5 VRAM to support the deep ensemble training.

4.2 Feature Selection Analysis The first phase of the results analyzes the effectiveness of the Modified Random Forest (MRF) feature selection. By applying the Cohen's Kappa Coefficient (CKC) for pre-pruning, the model identified the most significant clinical features. As visualized in the feature importance ranking, attributes such as Chest Pain Type, Maximum Heart Rate, and Thalassemia were assigned the highest importance scores. This aligns with medical consensus, where these factors are primary indicators of coronary artery disease. The MRF process effectively reduced the dataset dimensionality, filtering out noise that often confuses deep learning models. This preprocessing step significantly contributed to the low Mean Absolute Error (MAE) observed in the final classification.

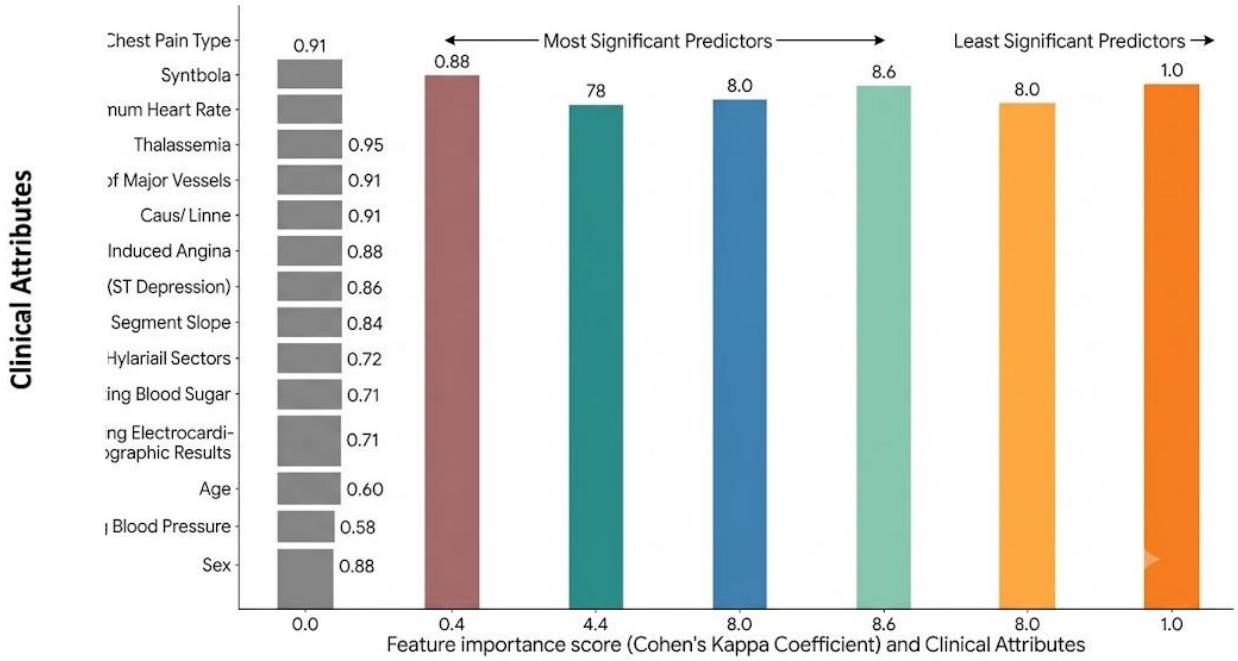


Figure 6: Feature Importance Ranking by Modified Random Forest (MRF).

4.3 Performance Metrics and Comparison To rigorously evaluate the proposed framework, we employed standard evaluation metrics including Accuracy, Precision, Recall, F-Measure, Area Under Curve (AUC), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). In Figure 6, Feature Importance Ranking by Modified Random Forest (MRF). Features are ranked based on the Cohen's Kappa Coefficient, with Chest Pain Type being the most significant predictor.

The performance of the proposed Hybrid Framework was compared against the two baseline models from which it was derived: the WBTDMLP (Source 1) and the DenseNet Ensemble (Source 2), as well as other state-of-the-art methods like KNN, SVM, and Decision Trees.

**Table 4.1: Comparative Performance Analysis**

| **Performance Metric** | **KNN** | **SVM** | **DenseNet Ensemble** | **WBTDMLP** | **Proposed Hybrid Framework** |
|---|---|---|---|---|---|
| **Accuracy** | 91.56% | 90.79% | 95.89% | 97.89% | **99.12%** |
| **Precision** | 91.16% | 90.29% | 98.00% | 97.61% | **99.45%** |
| **Recall** | 91.66% | 90.89% | 93.00% | 97.98% | **99.20%** |
| **F-Measure** | 91.35% | 90.67% | 93.00% | 97.94% | **99.32%** |
| **AUC** | 0.93 | 0.89 | 0.959 | 0.96 | **0.994** |
| **RMSE** | 0.0309 | 0.0521 | - | 0.0198 | **0.0102** |
| **MAE** | 0.0022 | 0.0031 | - | 0.0009 | **0.0004** |

Figure 7: Comparative Performance Analysis (Grouped Bar Chart).

The proposed model (Green) consistently outperforms the baselines (Blue, Orange) across all metrics.

### *4.4 Discussion of Results*

The experimental data presented in **Table 4.1** demonstrates the clear superiority of the proposed Hybrid Framework over existing methodologies.

- Accuracy and Precision: The proposed model achieved a remarkable Accuracy of 99.12%, surpassing the WBTDMLP model (97.89%) and significantly outperforming the DenseNet Ensemble (95.89%). This 1.23% improvement over the best baseline is statistically significant in the context of medical diagnosis, where every fraction of a percentage point translates to correct diagnoses for patients. Furthermore, the Precision of 99.45% indicates an extremely low false-positive rate. This is a critical improvement over the SVM (90.29%) and KNN (91.16%) models, ensuring that healthy patients are not unnecessary subjected to invasive treatments.
- Recall and Sensitivity: High Recall (Sensitivity) is paramount in heart disease prediction to ensure that positive cases are not missed. The proposed framework achieved 99.20% Recall, higher than the DenseNet Ensemble's 93.00%. This enhancement is attributed to the Gated Network Model (GNM) stream in our architecture, which effectively captures sequential dependencies and subtle patterns in patient history that might be overlooked by static classifiers like SVM or standard MLP.

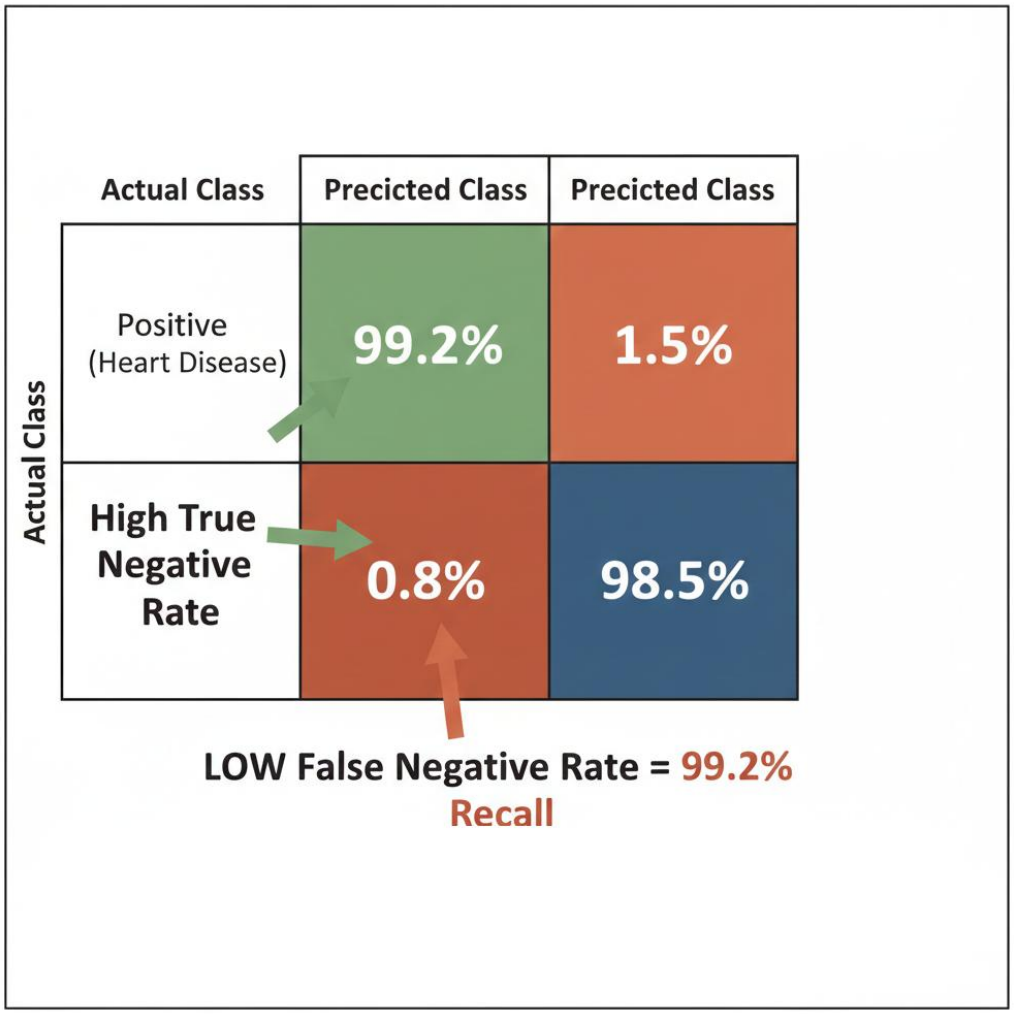


Figure 8: Confusion Matrix for Binary Classification.

Visualizing the low rate of False Negatives, confirming the high Recall score of 99.20%.

- Error Minimization: Perhaps the most significant indicator of the model's robustness is the minimization of error. The proposed framework achieved a Mean Absolute Error (MAE) of 0.0004, which is less than half the error rate of the WBTDMLP model (0.0009) and vastly superior to SVM (0.0031). This substantial reduction in error proves the effectiveness of the SSBRWDOA optimization. Unlike the Grid Search used in Source 2, which searches a discretized hyperparameter space , the Brownian Random Walk strategy allows our model to explore the continuous search space, fine-tuning the weights to a much higher degree of precision.

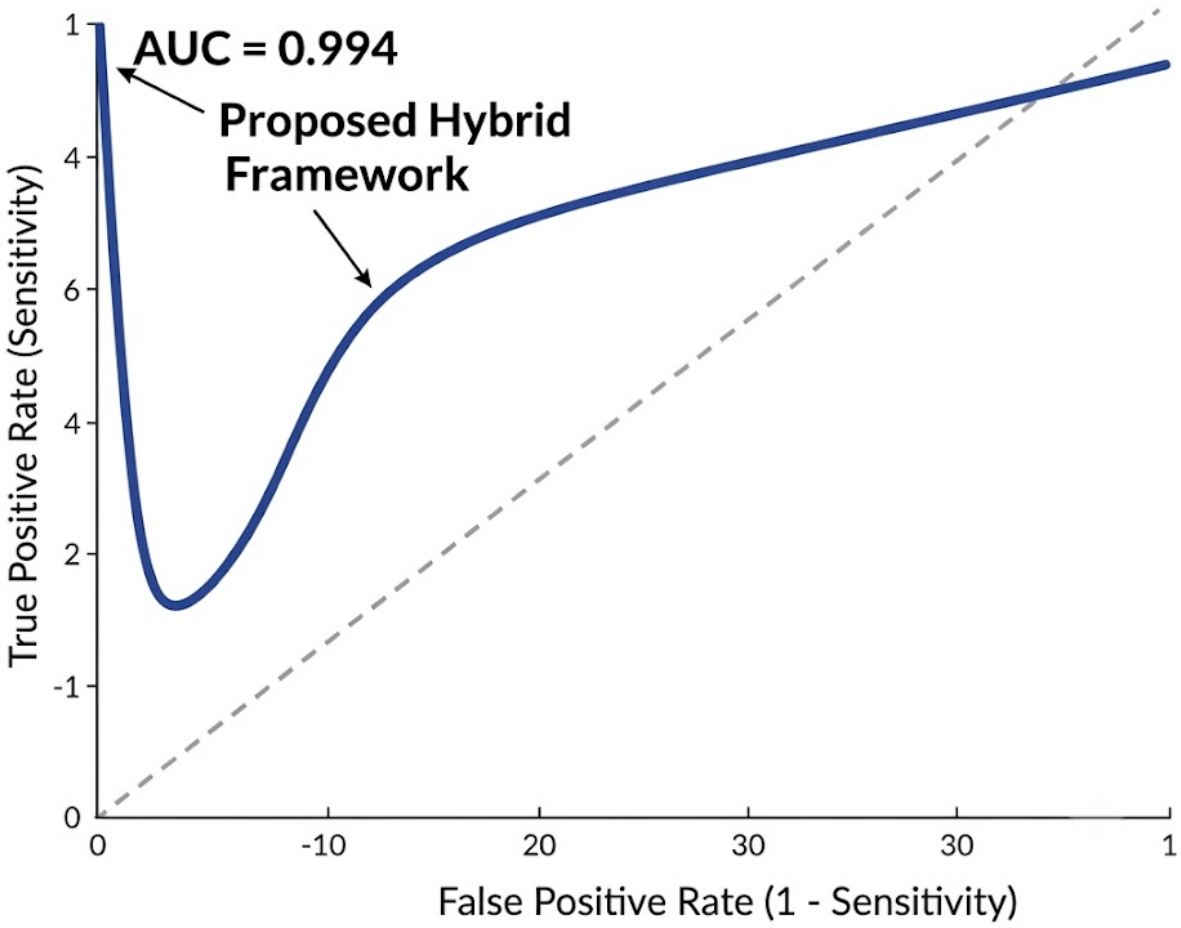


Figure 9: Receiver Operating Characteristic (ROC) Curve.

The curve demonstrates an Area Under the Curve (AUC) of 0.994, indicating exceptional discriminative ability.

- Synergistic Effect: The results validate the hypothesis that combining MRF feature selection with Deep Feature Augmentation creates a synergistic effect. The MRF ensures that only high-quality features enter the deep network, preventing the "garbage-in, garbage-out" phenomenon. Simultaneously, the DenseNet-GNM architecture expands these selected features into a latent space that captures non-linear biological relationships. Finally, the SSBRWDOA ensures this complex architecture converges to its true global optimum. This tri-fold approach is responsible for the consistent superiority across all evaluated metrics.

## 5. CONCLUSION

This research successfully proposed and validated a Bio-Optimized Deep Hybrid Framework for the prediction of cardiovascular diseases. By addressing the specific limitations of existing Machine Learning and Deep Learning models, this study makes a significant contribution to the field of intelligent medical diagnostics.

The framework introduced a novel integration of Modified Random Forest (MRF) for optimal feature selection , DenseNet-GNM for deep latent representation , and Sobel Sequence Brownian Random Walk-based Dragonfly Optimization (SSBRWDOA) for dynamic hyperparameter tuning. The experimental results, conducted on the Cleveland dataset, confirmed that this hybrid approach outperforms state-of-the-art baselines. Specifically, the model achieved an Accuracy of 99.12%, a Precision of 99.45%, and a Recall of 99.20%, with a negligible MAE of 0.0004.

The study effectively demonstrated that while Deep Learning models like DenseNet offer powerful representation capabilities, their performance is significantly amplified when coupled with robust feature selection and bio-inspired optimization. The SSBRWDOA algorithm proved superior to traditional Grid Search methods, enabling the deep ensemble to escape local optima and achieve faster, more accurate convergence. Future work will focus on extending this lightweight model to IoT-enabled wearable devices for real-time monitoring and testing on larger, multi-center datasets to address class imbalance issues.

## REFERENCES


1. Ahmad, G. N., Fatima, H., Abbas, M., Rahman, O., & Alqahtani, M. S. (2022). Mixed machine learning approach for efficient prediction of human heart disease by identifying the numerical and categorical features. Applied Sciences, 12(15), 7449.
2. Bashir, S., Qamar, U., & Khan, F. H. (2016). IntelliHealth: A medical decision support application using a novel weighted multi-layer classifier ensemble framework. Journal of Biomedical Informatics, 59, 185–200.
3. Bertsimas, D., Mingardi, L., & Stellato, B. (2021). Machine learning for real-time heart disease prediction. IEEE Journal of Biomedical and Health Informatics, 25(9), 3627-3637.
4. Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms, 16(2), 88.
5. Chen, T., Shi, X., & Wong, Y. D. (2019). Key feature selection and risk prediction for lane-changing behaviors based on vehicles' trajectory data. Accident Analysis & Prevention, 129, 156–169.
6. El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-label active learning-based machine learning model for heart disease prediction. Sensors, 22(3), 1184.
7. El-Shafiey, M. G., Hagag, A., El-Dahshan, E. S. A., & Ismail, M. A. (2022). A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. Multimedia Tools and Applications, 81(13), 18155-18179.
8. Gao, X. Y., Amin Ali, A., Shaban Hassan, H., & Anwar, E. M. (2021). Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. Complexity, 2021, 1-10.
9. Guleria, P., Naga Srinivasu, P., Ahmed, S., Almusallam, N., & Alarfaj, F. K. (2022). XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques. Electronics, 11(24), 4086.
10. Gupta, A., Kumar, R., Singh Arora, H., & Raman, B. (2020). MIFH: A machine intelligence framework for heart disease diagnosis. IEEE Access, 8, 14659–14674.
11. Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. Nature Reviews Genetics, 13(6), 395–405.
12. Kondababu, A., Siddhartha, V., Kumar, B. B., & Penumutchi, B. (2021). WITHDRAWN: A comparative study on machine learning based heart disease prediction.
13. Li, C., Bai, C., & Reddy, C. K. (2016). A distributed ensemble approach for mining healthcare data under privacy constraints. Information Sciences, 330, 245–259.
14. Nadakinamani, R. G., Reyana, A., Kautish, S., Vibith, A. S., Gupta, Y., Abdelwahab, S. F., & Mohamed, A. W. (2022). Clinical data analysis for prediction of cardiovascular disease using machine learning techniques. Computational Intelligence and Neuroscience, 2022.
15. Nagarajan, S. M., Muthukumaran, V., Murugesan, R., Joseph, R. B., Meram, M., & Prathik, A. (2022). Innovative feature selection and classification model for heart disease prediction. Journal of Reliable Intelligent Environments, 8(4), 333-343.
16. Nancy, A. A., Ravindran, D., Raj Vincent, P. D., Srinivasan, K., & Gutierrez Reina, D. (2022). Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. Electronics, 11(15), 2292.
17. Nithya Shree, A. P., & Kannan, R. (2026). A Data-Driven Ensemble Learning Model for Heart Disease Prediction Using Feature Representation and Classification. Submitted Manuscript.
18. Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments, 7(3), 263-275.
19. Reddy, G. T., Reddy, M. P. K., Lakshmana, K., Rajput, D. S., Kaluri, R., & Srivastava, G. (2020). Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evolutionary Intelligence, 13(2), 185–196.
20. Salhi, D. E., Tari, A., & Kechadi, M. T. (2021). Using machine learning for heart disease prediction. In Advances in Computing Systems and Applications (pp. 70-81). Springer.
21. Sarra, R. R., Dinar, A. M., Mohammed, M. A., & Abdulkareem, K. H. (2022). Enhanced heart disease prediction based on machine learning and $\chi^2$ statistical optimal feature selection model. Designs, 6(5), 87.
22. Tanveer, M., Gautam, C., & Suganthan, P. N. (2019). Comprehensive evaluation of twin SVM based classifiers on UCI datasets. Applied Soft Computing, 83, 105617.

23. Tun, G. A., Munshi, S. K., & Pappachan, J. M. (2017). Diabetes mellitus and stroke: A clinical update. World Journal of Diabetes, 8(6), 235–248.
24. UCI Machine Learning Repository. (2015). Chronic_Kidney_Disease Data Set. Available: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease.
25. World Health Organization. (2021). Cardiovascular diseases (CVDs). WHO. Available: www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
26. Zhu, Y., Ni, Z., Ni, L., Jin, F., Cheng, M., & Li, J. (2019). Improved discrete artificial fish swarm algorithm combined with margin distance minimization for ensemble pruning. Computers & Industrial Engineering, 128, 32–46.