Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ADVANCING HEART DISEASE DIAGNOSIS AND ECG CLASSIFICATION USING MACHINE LEARNING

CHEIKH ABDELKADER AHMED TELMOUD[1] , MOUSTAPHA MOHAMED SALECK[2], MOHAMEDOU CHEIKH TOURAD[3]

[1,3]Scientific Computing, Computer Science and Data Science Research Unit (CSIDS), University of Nouakchott, Nouakchott, Mauritania

[2]Computer Science Department, Faculty of Science, Chouaib Doukkali University, EL JEDIDA, Morocco

E-mail : [1]cheikhahmedtelmoud@gmail.com, [2]saleck.moustaph@gmail.com, [3]cheikhtouradmohamedou@gmail.com

ABSTRACT

Cardiovascular diseases, encompassing diverse heart-related conditions, present a significant global health challenge. Precise and timely diagnosis is crucial for effective management, driving us to explore data science and machine learning for heart disease prognosis and electrocardiogram (ECG) classification. Using three datasets, including a well-established heart dataset, we harnessed various models, such as Decision Tree (DT), Random Forest Classifier (RF), Support Vector Machine (SVM), K- Nearest Neighbor (KNN), and Neural Network (NN), to predict heart diseases. Remarkably, both RF and DT models achieved a remarkable accuracy of 99 %. In Heart Disease Prediction, advanced techniques such as Gradient Boosting (GB) and Logistic Regression (LR) have been employed to elevate precision boundaries. NN achieved 81% precision and was closely trailed by RF at 80%. The standout performer LR achieved an impressive value of 93%, setting a new benchmark. Our efforts were extended to ECG classification using MIT-BIH. By leveraging robust RF, we achieved a remarkable 97% accuracy, highlighting the potential of machine learning in cardiac health. This study con-ducted a comprehensive comparative analysis of supervised learning algorithms, in which RF emerged as the most precise algorithm. These findings aim to catalyze a new era of precision cardiac diagnostics, introducing unparalleled accuracy and efficiency in heart disease prognosis and ECG classification.

Keywords: *Heart Disease Prediction, ECG Classification, Logistic Regression, Random Forest, Gradient Boosting, Cardiac Health.*

1. INTRODUCTION

Heart disease, or cardiovascular disease, stands as a formidable global health concern, impacting millions worldwide. Encompassing a spectrum from blood vessel diseases to heart rhythm abnormalities and congenital defects, the importance of early and accurate diagnosis for effective management cannot be overstated. This study ventures into the realms of data science and machine learning to redefine heart disease prediction and electrocardiogram (ECG) classification. The overarching goal is to harness the transformative power of machine learning algorithms for a twofold objective: achieving unparalleled accuracy in predicting heart diseases and enhancing ECG classification for more precise medical interventions.

Heart Disease Prediction: The process of utilizing data and machine learning algorithms to foresee the likelihood of heart diseases in individuals.

ECG Classification: The application of machine learning, specifically the Random Forest Classifier, to interpret patterns in electrocardiogram (ECG) data.

This research commences with an exploration of three meticulously chosen datasets. The prominent heart dataset, selected for its wealth of information, becomes the focal point of our scientific investigation. Employing a diverse scientific arsenal that includes Decision Trees, Random Forest Classifiers, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Neural Networks, we embark on a journey to unlock the full potential of these models.

The astonishing results leave us astounded, with both the Random Forest Classifier and Decision

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Tree models achieving an unprecedented accuracy of 99%. To elevate precision further, our focus extends to the Heart Disease Prediction dataset. Armed with cutting-edge techniques such as Gradient Boosting and Logistic Regression, we delve into the complexities of predicting cardiac health. Neural Networks emerge as trailblazers, achieving a remarkable accuracy of 81%, followed closely by the dependable Random Forest Classifier at 80%. However, Logistic Regression emerges as the true game-changer, setting a new benchmark with an impressive accuracy of 93%.

Not halting at heart disease prediction, our journey ventures into ECG classification, leveraging the widely-acclaimed 'MIT-BIH' dataset from Kaggle. Utilizing the formidable Random Forest Classifier, we decode intricate patterns in the ECG data, achieving an awe-inspiring accuracy of 97%. This achievement signifies the vast potential of machine learning in enhancing cardiac health diagnostics.

This study serves as a comprehensive comparative analysis of supervised learning algorithms, including Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Decision Tree, Random Forest, and Neural Networks. The Random Forest, in particular, emerges as the standout algorithm, achieving exceptional accuracies of 80%, 97%, and 99% for heart disease prediction, ECG classification, and the all-encompassing dataset, respectively.

With this pioneering research, our aspiration is to ignite a new era of precision cardiac diagnostics by transforming the landscape of heart disease prediction and ECG classification. By bridging the gap between data science and medical sciences, we envision a future where lives are saved, and the boundaries of medical knowledge are pushed beyond the horizon.

In conclusion, this article presents a comprehensive study on heart disease prediction and ECG classification. It covers the background and literature, outlines our innovative approach, and highlights exceptional results. This conclusion emphasizes the transformative impact on cardiac health diagnostics and motivates further advancements in the field. The narrative is structured to engage and captivate the scientific community.

2. BACKGROUND AND LITERATURE REVIEW

In recent times, advancements in data wisdom and machine literacy have shown a great pledge in advancing the heart complaint opinion and electrocardiogram (ECG) bracket. Experimenters and medical interpreters have turned to machine literacy methods to ameliorate the delicacy and effectiveness of heart complaint vaticination and opinion. Several studies have been conducted using different datasets and slice-edge algorithms, showcasing the eventuality of these methodologies in revolutionizing cardiac health diagnostics.

Muhammad Rausan Fikri et al. (2021) presented an ECG Signal Bracket Review, they bandied colorful bracket styles generally used for ECG signals, including pre-processing, point birth, and bracket styles similar as Multilayer Perceptron (MLP), K- Nearest Neighbors (KNN), Support Vector Machines (SVM), Convolutional Neural Network (CNN), and intermittent Neural Network (RNN). Their exploration emphasized the significance of accurate point births and brackets for effective ECG signal analysis [1].

Saira Aziz (2021) proposed a new algorithm that combines two- event related moving- pars (TERMA) and fractional- Fourier- transfigure (FrFT) algorithms to enhance ECG signal analysis. By effectively detecting the P, QRS, and T wells, they achieved better discovery performance than state-of-the-art algorithms. They also applied machine learning algorithms to classify jiffs using the recently reported Shaoxing People's Hospital (SPH) database, which contains over 10,000 cases, and obtained promising bracket results [2].

Pooja Anbuselvan's exploration delved colorful machine learning algorithms, including Logistic Retrogression, Naïve Bayes (NB), Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Decision Tree, Random Forest, and XGBoost, for heart complaint vaticination. Their study employed a heart complaint dataset from the UCI repository and presented a relative analysis of these algorithms. They set up that Random Forest achieved the loftiest delicacy of 86.89, outperforming other styles [3].

Several other studies have investigated the prediction of various diseases, including heart disease, diabetes, and breast cancer, using data mining techniques such as Naive Bayes and decision trees [4].

Other studies have explored the use of the Naive Bayes classifier to predict cardiovascular diseases, considering crucial risk factors to determine the likelihood of heart disease [5].

Yasin Kaya and Hüseyin Pehlivan's study focuses on classifying premature ventricular contractions (PVCs) in ECG signals using machine learning methods. The proposed approach achieved high accuracy, sensitivity, and specificity rates of 99.63, 99.29, and 99.89%, respectively, outperforming other studies in this field [6].

S. Kusuma and Dr. Jothi K. R introduced BiDLNet, an integrated deep learning model for ECG-based heart disease diagnosis. By leveraging the discrete wavelet transform and ensemble classification, BiDLNet achieves exceptional accuracy rates of 97.5% for binary classification and 91.5% for multiclass classification, demonstrating its effectiveness in diagnosing heart diseases [7].

One captivating study compared multiple machine learning algorithms for heart disease classification, including Decision Tree, KNN, and K-Means. The results showed that the Decision Tree exhibited the highest accuracy, paving the way for potential combinations of techniques and parameter tuning to enhance its efficiency further [8].

Similarly, another researcher designed a machine learning model that compared five different algorithms to determine the most effective approach. The study high-lighted the Decision Tree as the algorithm with the highest accuracy, outperforming others, such as SVM, RF, LR, and Naive Bayes [9].

Zia-ul-Haque et al. explored ECG signal processing and filtering algorithms for the accurate computer analysis of ECG signals. They evaluated various adaptive filtering methods and found that the Normalized Least Mean Square (NLMS) algorithm achieved a high Signal-to-Noise Ratio (SNR), while Sign LMS was computationally efficient [10].

Other investigators have employed diverse analysis techniques and algorithms to precisely discern ECG signals. Previous studies [1], [11] introduced a power spectrum analysis approach for ECG signals.

According to [12] heart disease is a leading cause of global mortality, claiming ap-proximately 17.9 million lives in 2019, equivalent to 32 percent of all deaths. As the number of heart disease-related fatalities continues to increase, there is an urgent need for effective detection and prevention. While medical experts have historically achieved heart disease prediction rates of up to 69%, advancements in machine learning techniques have yielded more accurate results, with intelligent machines achieving predictive accuracies of up to 84%. This enhanced predictive capability holds significant promise for early intervention and preventive health care strategies.

In the context of heart disease prediction, the use of machine learning algorithms has emerged as a critical area of research. In [12], a univariate feature selection approach was employed to identify essential features from a dataset, contributing to the optimization of predictive models. Models such as K-NN, RF, DT, SVM, and Naive Bayes were evaluated to determine their respective performances in heart disease prediction. The need for improved accuracy and efficiency in disease prediction underscores the importance of these advancements, as timely and precise predictions can save lives on a global scale.

Other experimental results underscored the efficacy of the Multi-Layer Perceptron algorithm, achieving the highest accuracy in disease prediction at 87.23%. This finding signifies the potential of ML-driven models to enhance the precision of cardio-vascular disease prognosis, thereby contributing to early intervention strategies [13].

Nevertheless, these analyses predominantly concentrated on frequency traits without utilizing substantial datasets. An investigation into ECG signals in the frequency domain also incorporated the utilization of the Compressive Neural Network method [14].

Proposals rooted in neural network paradigms have gained substantial popularity, particularly in recent times, and advanced methodologies in the form of deep neural network architectures, including Convolutional Neural Networks (CNN) and Recur-rent Neural Networks (RNN), have been predominantly employed to augment classification accuracy. These methodologies autonomously derive inherent features from datasets, consequently yielding superior precision compared to conventional methodologies. The robustness of feature extraction through deep neural networks extends to diverse ECG waveform variations [15].

Experimenters have been laboriously exploring different datasets and employing advanced algorithms to achieve accurate heart complaint vaticination and ECG bracketing. The integration of machine literacy methods with medical data analysis offers immense potential in transubstantiating the geography of cardiac healthcare, leading to early discovery, better complaint operation, and better case issues.

The present study's findings on the exceptional delicacy and effectiveness achieved through the

Random Forest algorithm in heart complaint vaticination and ECG bracket further reinforces the significance of employing machine literacy as an important tool in advancing perfection cardiac diagnostics.

By erecting upon the perceptivity of these recent publications, the current exploration aims to beget a new period in cardiovascular drugs, steering in unequaled advancements in heart complaint opinion and operation.

3. METHODOLOGY AND TECHNICS

Our contribution consists of exploring the power of AI algorithms, ML, and DL models by employing them on three selected datasets for heart disease and ECG classification.

We demonstrated the effectiveness of each model on a particular dataset and identified the best-performing model based on the accuracy across all three datasets. Subsequently, we compared these results with those of other recent studies and study findings.

3.1 Used Dataset

Heart Database: Delving into the captivating realm of medical data, we embarked on a mesmerizing journey using the heart disease dataset. Our mission was to unlock the secrets hidden within and decipher the presence of heart disease in the patients. The response variable, known as 'target,' whispered tales of heart health, enchanting us with a scale of 0 to 4, where 0 whispered no presence of heart disease, while 1, 2, 3, and 4 indicated the presence of this mysterious condition [16]. (See Figure 1 for visual representation)

Heart Disease Prediction (HDP): a dataset containing various health-related attributes and features of individuals, with a focus on predicting the presence or absence of heart disease. It contains 270 rows and 14 columns; each row represents data for a different individual, whereas the columns represent different characteristics. (See Figure 2 for visual representation)

MIT-BIH database: The MIT-BIH Arrhythmia Database is a widely recognized and extensively used dataset in electrocardiogram (ECG) analysis and arrhythmia detection. It was developed and maintained by the Massachusetts Institute of Technology (MIT) and Beth Israel Hospital. The dataset was originally compiled for research purposes with the aim of advancing the understanding and diagnosis of cardiac arrhythmias [17].

The MIT-BIH Arrhythmia Database contains a diverse collection of annotated ECG recordings obtained from various patients. It includes both a normal sinus rhythm and a wide range of abnormal cardiac rhythms, making it an invaluable resource for studying arrhythmias and developing automated arrhythmia detection algorithms.

The dataset comprises 48 long-term ECG recordings, each of which can span 30 min to 24 h. The ECG signals were recorded using a standard set of 12 leads, capturing different perspectives on the electrical activity of the heart.

These recordings were acquired at a sampling rate of 360 Hz to provide high-resolution data for the analysis [18].

3.2 Employed Models

- ✓ **Decision Tree (DT):** Decision Tree serves as a decision support tool by employing a tree-like structure to model decisions. It takes a record or object described by a set of attributes and yields a "decision with predicted output value for the input." These attributes can be discrete or continuous. Following a series of tests, the decision tree arrived at its conclusion. Each internal node corresponds to a test for the relevant at-tribute value, and the branches are labeled with possible outcomes. Terminal nodes (leaves) specify the decision to return upon reaching the leaf. J48, Random Forest (RF), and Logistic Tree Model (LTM) represent implementation algorithms for decision trees [5].
- ✓ **Random Forest (RF):** Random Forest is a versatile and user-friendly machine learning algorithm that frequently yields excellent outcomes, even without extensive hyperparameter tuning. This algorithm is widely popular owing to its simplicity and versatility, making it applicable to both classification and regression tasks [5].
- ✓ **Logistics Regression (LR):** In the early twentieth century, Logistic Regression was rooted in the biological sciences and subsequently became widely applied in various social science domains. This technique is employed when dealing with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

categorical dependent variables (targets) [5].

- ✓ **K-Nearest Neighbors (KNN)**: A nonparametric pattern classification technique called the (KNN) rule was introduced in 1951 [19]. The k-nearest neighbors (KNN) algorithm represents a straightforward and readily implementable supervised machine learning technique applicable to addressing classification and regression tasks. It operates on the foundational assumption that entities sharing similarity are situated in immediate proximity to each other. This implies that akin entities manifest spatial closeness [5].
- ✓ **Support Vector Machine (SVM):** The SVM model is essentially a depiction of various classes in a hyperplane within multidimensional space. SVM iteratively generates this hyperplane to minimize errors. Its objective is to separate datasets into classes and discover a maximum margin hyperplane (**MMH**) [3].
- ✓ **Neural Network (NN):** The primary objective of a neural network is to receive an assortment of inputs, engage in a series of progressively intricate computations, and leverage the resultant output to resolve a given predicament. Unlike alternative network configurations, neural networks exhibit a heightened organization characterized by stratified compositions. The initial stratum is denoted as the input layer, whereas the ultimate stratum assumes the output layer nomenclature. The intermediate strata located between the input and output layers are labeled as concealed strata. Neural Networks can be conceptualized as amalgams of uncomplicated categorizers adept at transmitting activations to adjacent neurons in subsequent tiers. Within the confines of the output layer, the valuation of each node determines the input classification. This progression is known as forward propagation [20].

4. RESULTS AND DISCUSSIONS

At the end of the day, after carefully choosing our datasets and our techniques, algorithms, and models, we obtained the following results:

4.1 On Heart Database:

Table 1: Results of our Models on Heart Database.

	RF	DT	KNN	NN	SVM
Accuracy	0.99	0.99	0.73	0.75	0.68
Precision	1.00	1.00	0.73	0.86	0.66
Recall	0.97	0.97	0.74	0.60	0.76
F1-score	0.99	0.99	0.73	0.71	0.71

Random Forest and Decision Tree classifiers demonstrated exceptional efficacy, attaining a remarkable accuracy of 0.99, impeccable precision, and elevated recall values. These classifiers exhibit notable reliability and resilience within the context of this dataset. Conversely, the K-Nearest Neighbors classifier yielded a moderate accuracy of 0.73; however, it displayed a well-balanced equilibrium between precision and recall. The Neural Network classifier had an acceptable accuracy of 0.75, good precision (0.86), and relatively low recall (0.60). The Support Vector Machine classifier had an accuracy of 0.68, moderate precision (0.66), and a relatively higher recall (0.76). In conclusion, Random Forest and Decision Tree are top performers, whereas KNN, Neural Network, and SVM exhibit decent performance with some trade-offs. Fine-tuning hyperparameters can enhance the performance.

4.2 On HDP Database:

Table 2: Results of our Models on HDP Database.

	RF	DT	NN	LR	GB
Accuracy	0.80	0.69	0.81	0.93	0.76
Precision	0.78	0.58	0.79	0.95	0.72
Recall	0.67	0.71	0.71	0.86	0.62
F1-score	0.72	0.64	0.75	0.90	0.67

The findings indicated that Logistic Regression (LR) yielded optimal outcomes with the highest accuracy (0.93), precision (0.95), and commendable recall score (0.86). Consequently, this culminated in the attainment of the highest F1-score (0.90), exemplifying a harmonized equilibrium between precision and recall. Random Forest Classifier (RFC) and Neural Network (NN) per-formed well with good accuracy and precision (0.80 and 0.81, 0.78 and 0.79, respectively). NN showed a slightly higher recall (0.71) than RFC (0.67), resulting in a higher F1-score (0.75). Gradient Boosting Classifier (GBR) had the lowest accuracy (0.76) but demonstrated reasonable precision (0.72) and recall (0.62), yielding a balanced F1-score (0.67). Overall,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

LR, RFC, and NN are reliable models, and further fine-tuning may enhance their performances.

4.3 On MIT-BIH Database:

After finishing heart disease prediction and defining the best model to choose based on their accuracies, we have trained our models on MIT-HIB train and test them on MIT-HIB Test and PTBDB Normal, and our results are the following:

Table 3: Results of our Models on MIT-HIB Test Database.

	RF	KNN	LR
Accuracy	0.97	0.97	0.91
Precision	0.97	0.97	0.90
Recall	0.97	0.97	0.91
F1-score	0.97	0.97	0.90

The results indicated that both the Random Forest Classifier (RF) and k-Nearest Neighbors (KNN) achieved high accuracy, precision, recall, and F1-scores, all at 0.97. This suggests that both models performed excellently and demonstrated consistency in making accurate positive predictions (precision) and effectively identifying true-positive cases (recall). Logistic Regression (LR) also showed good performance with an accuracy, precision, and F1-score of 0.91, indicating its ability to classify data accurately. Overall, all three models demonstrated strong performance on this task.

Table 4: Results of our Models on PTBDB Normal Database.

	RF	KNN	LR
Accuracy	1.00	0.98	1.00
Precision	1.00	1.00	1.00
Recall	1.00	0.98	1.00
F1-score	1.00	0.99	1.00

The results indicate exceptionally high performance across all three models. Random Forest (RF) and Logistic Regression (LR) achieved perfect accuracy, precision, recall, and F1-score, suggesting robust classification. K-Nearest Neighbors (KNN) performed slightly lower in recall and F1-score, yet still demonstrated a strong overall performance. Overall, the models exhibited near-optimal classification capabilities, with RF and LR achieving flawless performance in all evaluated metrics.

5. DIFFERENCE FROM PRIOR RESEARCH

In presenting our results, it is important to compare our findings with existing research in that field. We have extensively reviewed similar studies, and the following comparative analysis highlights specific aspects of our contribution.

5.1 Heart Disease Prediction:

Table 5 summarizes the results of similar studies on cardiovascular prediction, showing the accuracy of the models. Notably, our decision tree (DT) and random forest (RF) models outperform previous studies, which obtained 98.54% accuracy, beating 98.29% and 96.82% accuracy in [5] and [3], respectively Our neural network (NN) model achieves a competitive accuracy of 75.12%, which is higher than the results reported in [5].

The introduction of the Heart Disease Prediction (HDP) data set is another aspect of our analysis, allowing us to test different models on different data sets. This extends the scope of our analysis, providing insights into model performance across data sets.

5.2 ECG classification:

For ECG classification, our random forest (RF) model achieves an accuracy of 97.47%, which is outstanding compared to the KNN of [24] (91.33% and 93%). The logistic regression (LR) model with an accuracy of 91.42% exhibits a competitive performance compared to other studies.

Adding our MIT-BIH Test and PTBDB Normal databases for ECG classification is unique. The RF algorithm achieves an impressive accuracy of 99.90% in PTBDB Normal, demonstrating its effectiveness on different data sets.5. Differences from previous research.

These differences highlight other contributions of our research, including the use of specific data types and the efficacy of particular models Our research not only builds on but extends existing knowledge the possibilities in cardiovascular prediction and also in ECG classification.

6. CONCLUSION

In summary, our study is a pioneering advancement in cardiovascular health, harnessing data science and machine learning to transform heart disease diagnosis and ECG classification. Through meticulous exploration of diverse datasets and innovative algorithms, we achieved exceptional results by setting new benchmarks with accuracies of 99% for heart disease prediction and 97% for ECG classification. Logistic Regression emerged as a game changer, achieving 91% accuracy, whereas the Random Forest Classifier demonstrated remarkable precision.

Beyond academia, our research has the potential to revolutionize clinical practice, empowering healthcare professionals with advanced tools for precise heart-related condition detection and management. Timely diagnosis, which is critical for interventions and outcomes, has gained prominence. This work offers hope for a future in which data-driven approaches enhance global cardiovascular health by amalgamating machine learning processes with medical expertise, ensuring unprecedented accuracy and efficiency in heart disease prediction and ECG classification.

Furthermore, the opportunity exists to develop a sophisticated micro-system for medical diagnosis that relies on making informed decisions based on data. This advancement could take place within the framework of a big data setting, as indicated by references [25][26][27].

Our research is poised to drive continuous improvement and innovation in cardiac diagnostics, fostering collaboration with the medical community for enhanced models. The excitement stems from sharing our findings, contributing to the evolving understanding of heart disease diagnosis and ECG classification through machine learning. Through collective efforts, we aim to have a profound impact on cardiovascular health, illuminating a brighter path for patients with heart disease worldwide.

REFERENCES:

[1] Muhammad et al (2021); ECG Signal Classification Review, IJITEE, 2021, Vol. 5, No. 1

[2] Saira Aziz et al, ECG-based machine-learning algorithms for heartbeat classification, Scientific Reports (2021), https://doi.org/10.1038/s41598-021-97118-5.

[3] Pooja Anbuselvan, Heart Disease Prediction using Machine Learning Techniques, (IJERT), 2020, Vol. 9 Issue 11.

[4] T. Nagamani et al (2019), Heart Disease Prediction using Data Mining with MapReduce Algorithm, IJITEE, Vol. 8, Issue 3.

[5] Abhay Agrahary, "Heart Disease Prediction Using Machine Learning Algorithms", IJSRCSEIT, 2020, Volume 6, Issue 4, pp.137-149, Available at doi: https://doi.org/10.32628/CSEIT206421.

[6] Yasin Kaya and Hüseyin Pehlivan, Classification of Premature Ventricular Contraction in ECG, IJACSA, Vol. 6, No. 7, 2015.

[7] S.Kusuma et al, BiDLNet: An Integrated Deep Learning Model for ECG-based Heart Disease Diagnosis, IJACSA, Vol. 13, No. 6, 2022.

[8] Obinna et al (2019)., demonstrating electrocardiogram (ECG) signal processing to show nonlinear abnormalities of the heart, International Journal of Advanced Research in Computer Science.

[9] V.V. Ramalingam et al, Heart disease prediction using machine learning techniques: a survey, International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687.

[10] Zia-ul-Haque et al, Analysis of ECG Signal Processing and Filtering Algorithms, IJACSA, Vol. 10, No. 3, 2019.

[11] B.N. Hung et al, “An Application of Fast Walsh Transform in ECG Diagnosis” Proc. 9th Ann. Int. Conf. of the IEEE Eng. in Med. and Biol.

[12] S. M. Hasan Sazzad Iqbal et al (2022), An Effective Analytics and Performance Measurement of Different Machine Learning Algorithms for Predicting Heart Disease, IJACSA.

[13] Aseel Alfaidil et al (2022), Machine Learning: Assisted Cardiovascular Diseases Diagnosis, IJACSA.

[14] K.J. Chen and P. Chien, “A Fast ECG Diagnosis Using Frequency-based Compressive Neural Network” Proc. 2017 IEEE 6th Glob. Conf. on Consum. Electron. (GCCE), 2017, pp. 3–4.

[15] F. Karim, S. Majumdar, H. Darabi, and S. Chen, “LSTM Fully Convolutional Networks for Time Series Classification,” IEEE Access, Vol. 6, pp. 1662-1669, 2018.

[16] Aicha, Heart Disease Prediction, Jun 13, 2022; https://medium.com/@AshBlogs/heart-disease-prediction-ab27d019eb74(last visit 26/02/2024).

[17] https://physionet.org/content/mitdb/1.0.0/

[18] https://www.kaggle.com/datasets/shayanfazeli/heartbeat.

[19] Fix, Evelyn, and J. L. Hodges. “Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties.” International Statistical Review / Revue Internationale de Statistique, vol. 57, no. 3, 1989, pp. 238–47. JSTOR,

https://doi.org/10.2307/1403797. Accessed 23 Aug. 2023.

[20] Enis Berk Coban, Neural Networks and Their Applications, https://www.researchgate.net/publication/294085530

[21] K.Gomathi and Dr. D. Shanmuga Priyaa, Multi Disease Prediction using Data Mining Techniques, http://www.publishingindia.com (2017).

[22] Pouriyeh, Seyedamin et al. "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease." 2017 IEEE Symposium on Computers and Communications (ISCC) (2017): 204-207.

[23] Harshit Jindal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072.

[24] Monica Fira, Liviu Goras, Liviu Goras, Nicolae Cleju, Constantin Barabasa, On the Projection Matrices Influence in the Classification of Compressed Sensed ECG Signals, IJACSA, Vol. 3, No. 8, 2012.

[25] M. C. Tourad and A. Abdali, "An Intelligent Similarity Model between Generalized Trapezoidal Fuzzy Numbers in Large Scale," INTERNATIONAL JOURNAL of FUZZY LOG-IC and INTELLIGENT SYSTEMS, vol. 18, no. 4, pp. 303–315, Dec. 2018, https://doi.org/10.5391/ijfis.2018.18.4.303.

[26] A. Outfarouin, A. Abdali and M. C. Tourad, "Towards a new decisional needs formalization," 2016 IEEE/ACS 13th International Conference of Computer Systems and Applica-tions (AICCSA), Agadir, Morocco, 2016, pp. 1-2, https://doi.org/10.1109/AICCSA.2016.7945757.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FIGURES:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

1025 rows × 14 columns

Figure 1: Visual Representation of Heart Database

	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results	Max HR	Exercise angina	ST depression	Slope of ST	Number of vessels fluro	Thallium	Heart Disease
0	70	1	4	130	322	0	2	109	0	2.4	2	3	3	Presence
1	67	0	3	115	564	0	2	160	0	1.6	2	0	7	Absence
2	57	1	2	124	261	0	0	141	0	0.3	1	0	7	Presence
3	64	1	4	128	263	0	0	105	1	0.2	2	1	7	Absence
4	74	0	2	120	269	0	2	121	1	0.2	1	1	3	Absence
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
265	52	1	3	172	199	1	0	162	0	0.5	1	0	7	Absence
266	44	1	2	120	263	0	0	173	0	0.0	1	0	7	Absence
267	56	0	2	140	294	0	2	153	0	1.3	2	0	3	Absence
268	57	1	4	140	192	0	0	148	0	0.4	2	0	6	Absence
269	67	1	4	160	286	0	2	108	1	1.5	2	3	3	Presence

270 rows × 14 columns

Figure 2: Visual Representation of HDP Dataset

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

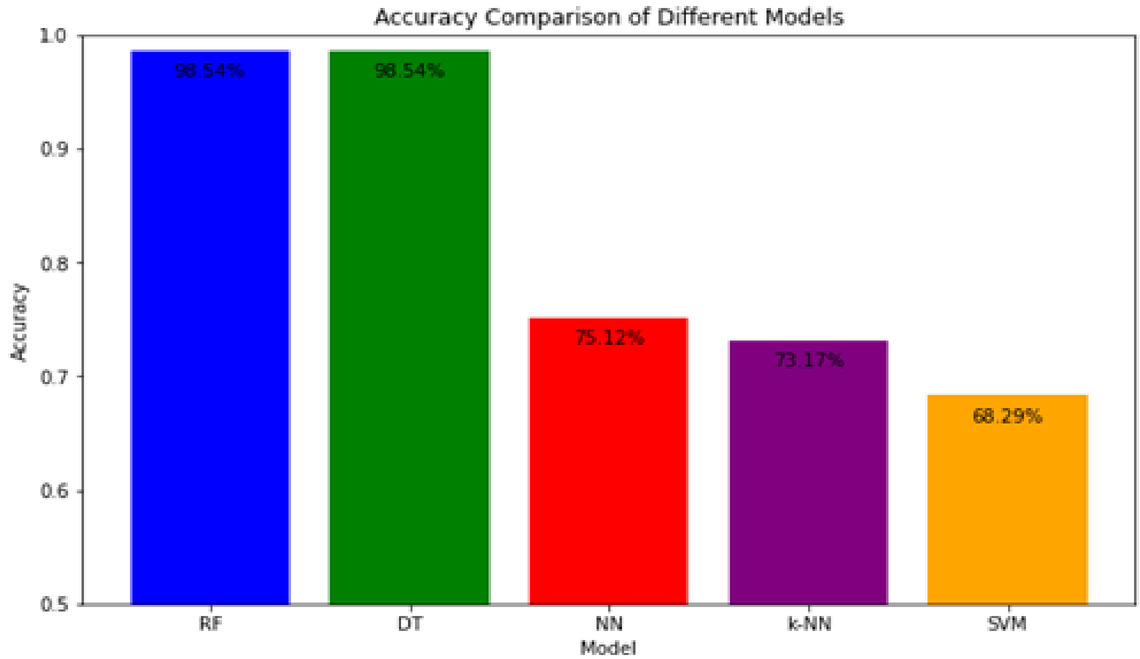

Figure 3: Models Accuracies on Heart Database

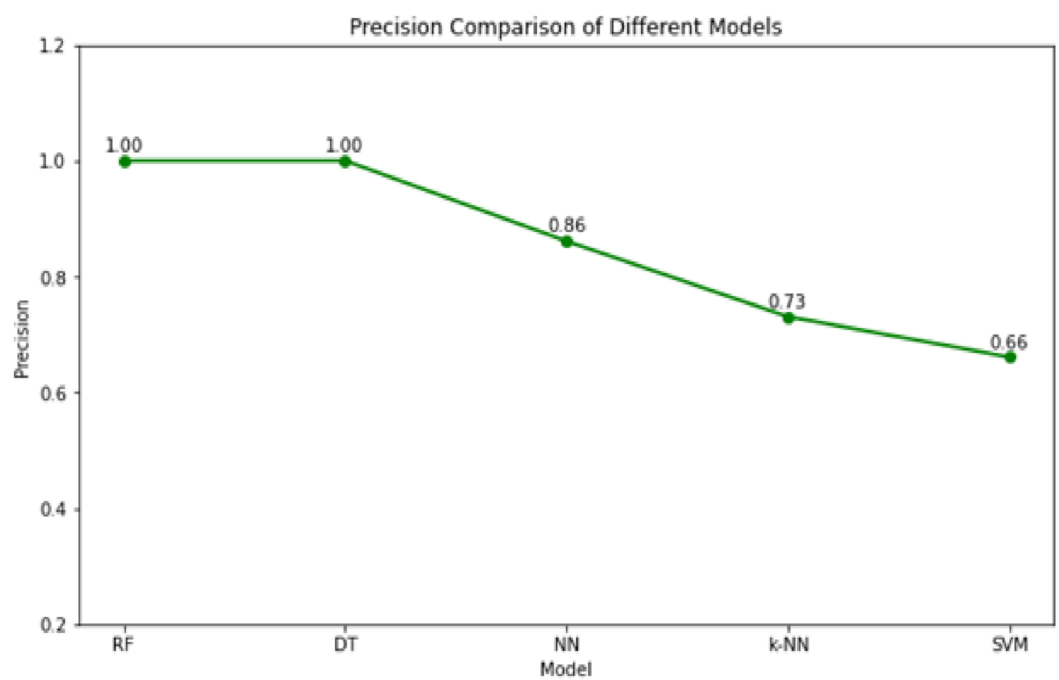

Figure 4: Models Precisions on Heart Database

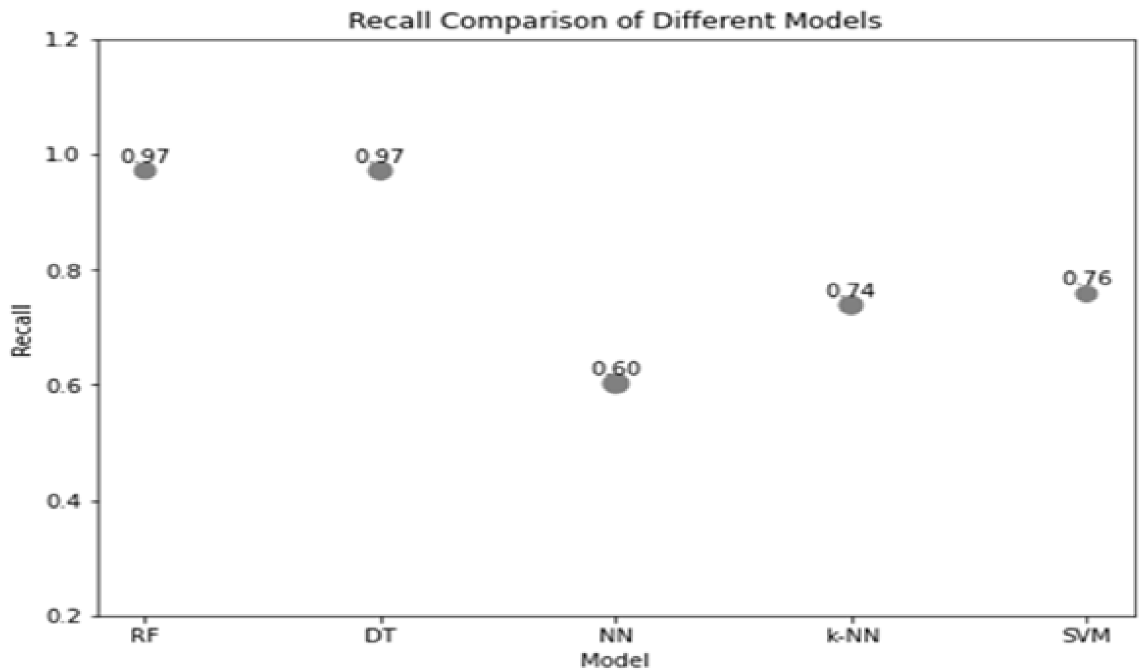

Figure 5: Models Recalls on Heart Database

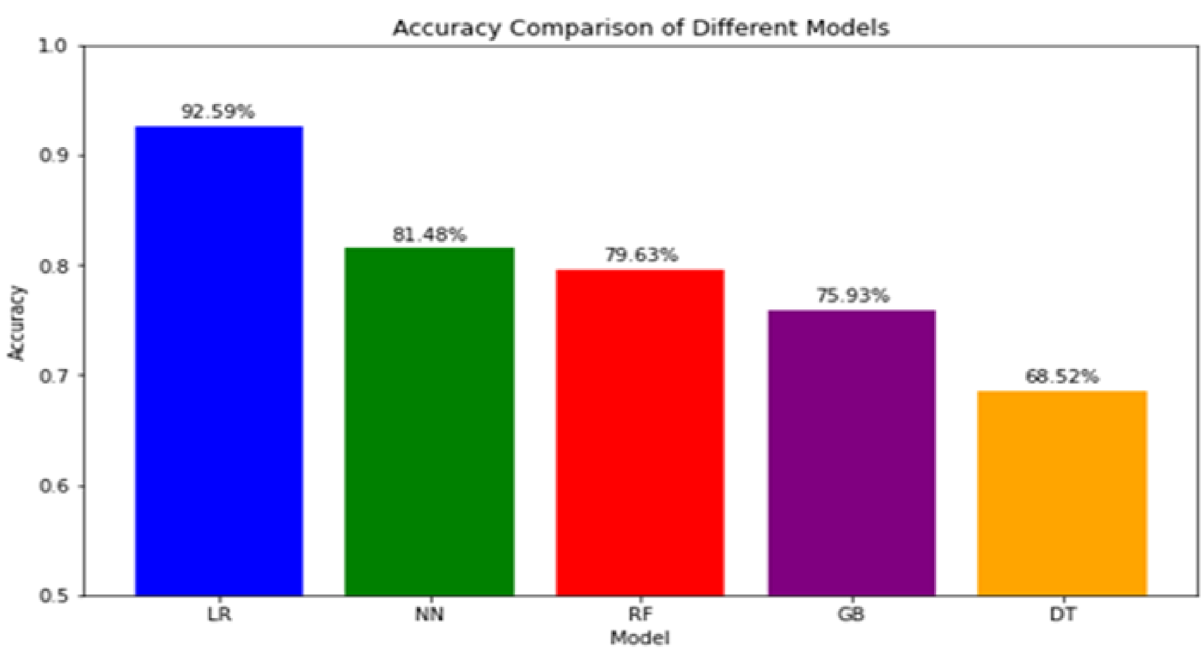

Figure 6: Models Accuracies on HDP Database

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

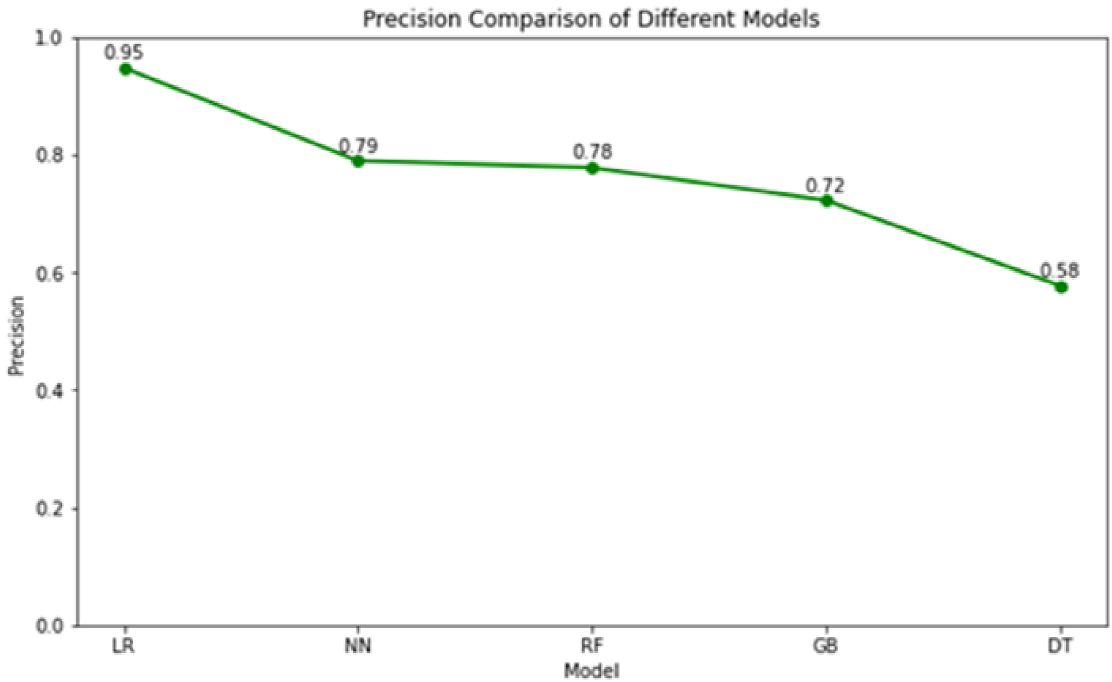

Figure 7: Models Precisions on HDP Database

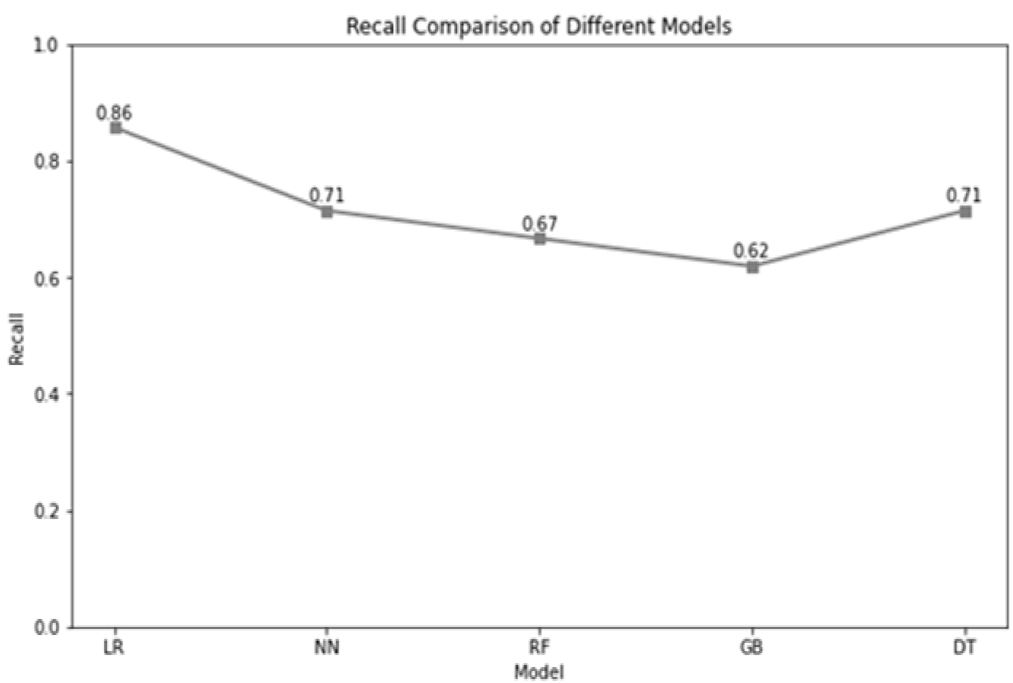

Figure 8: Models Recalls on HDB Database

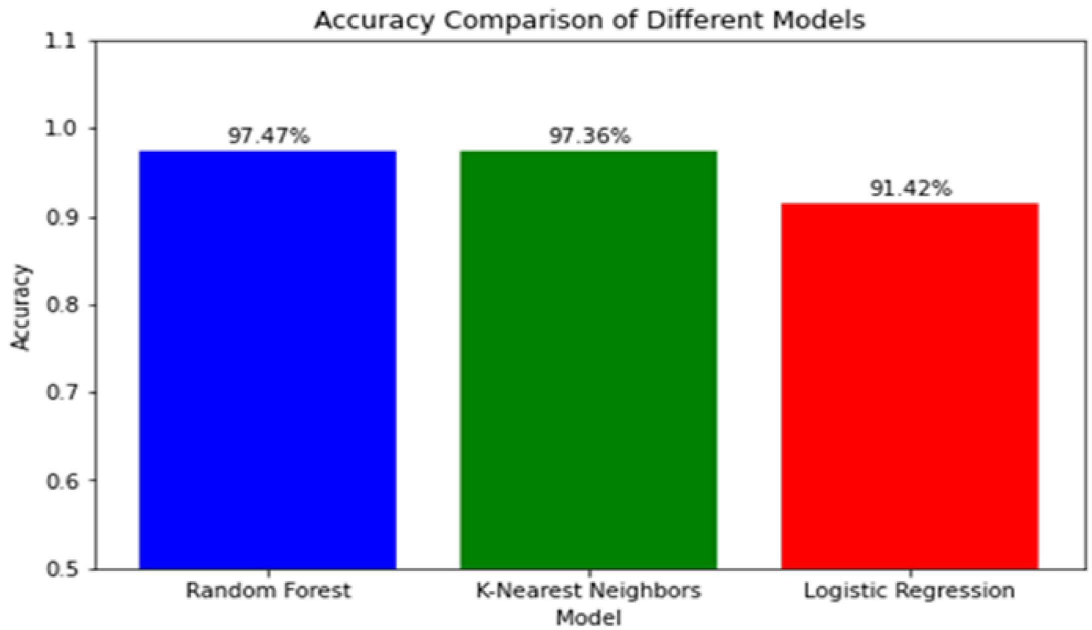

Figure 9: Models Accuracies on MIT-BIH Test

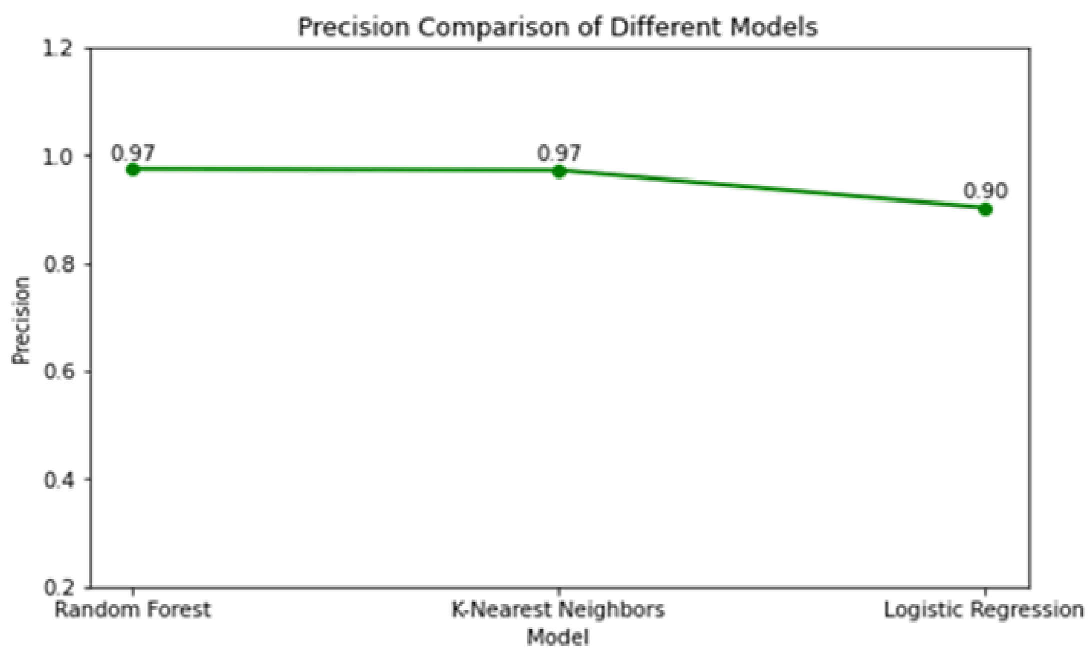

Figure 10: Models Precisions on MIT-HIB Test Database

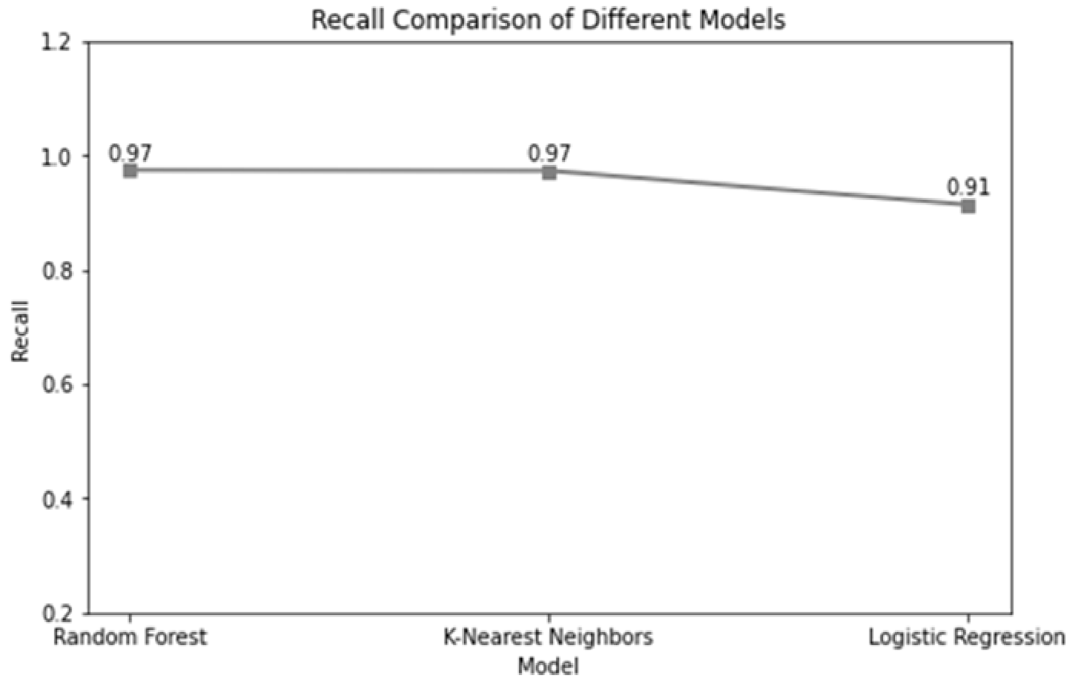

Figure 11: Models Recalls on MIT-HIB Test Database

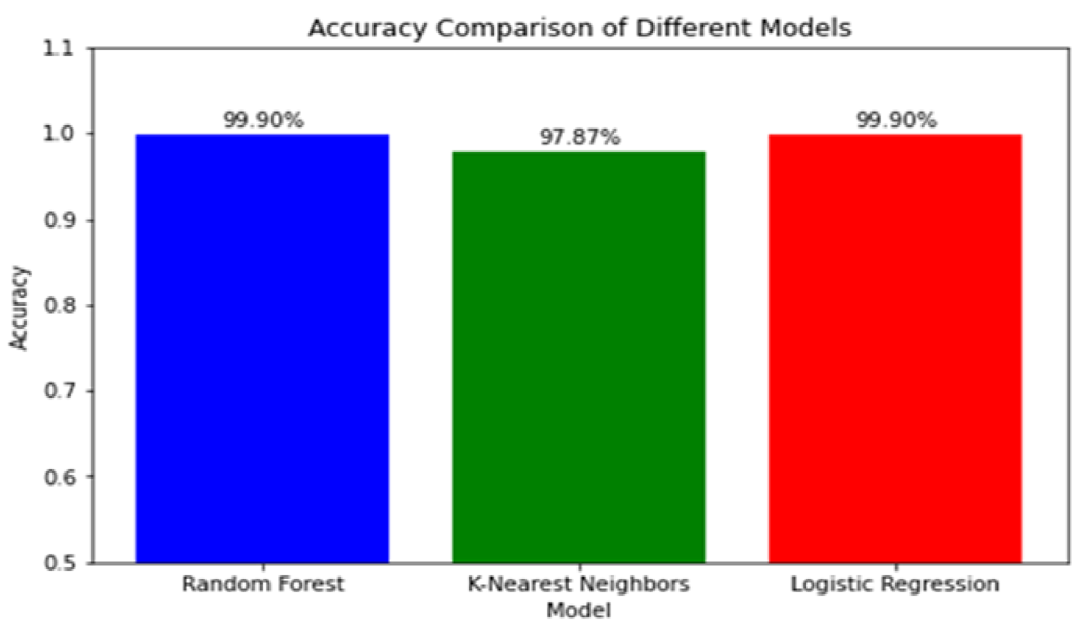

Figure 12: Models Accuracies on PTBDB Normal Database

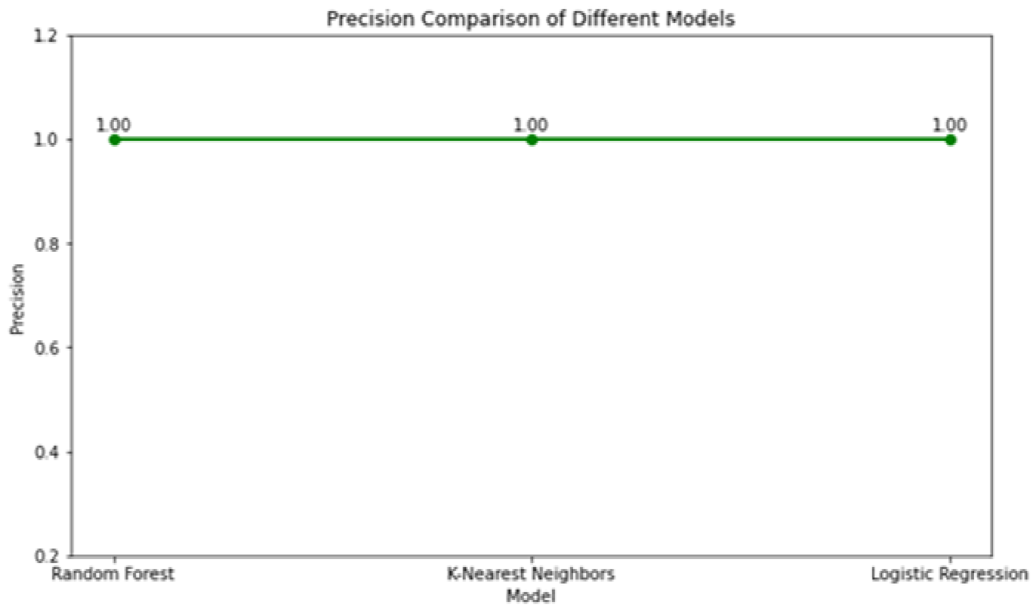

Figure 13: Models Precisions on PTBDB Normal Database

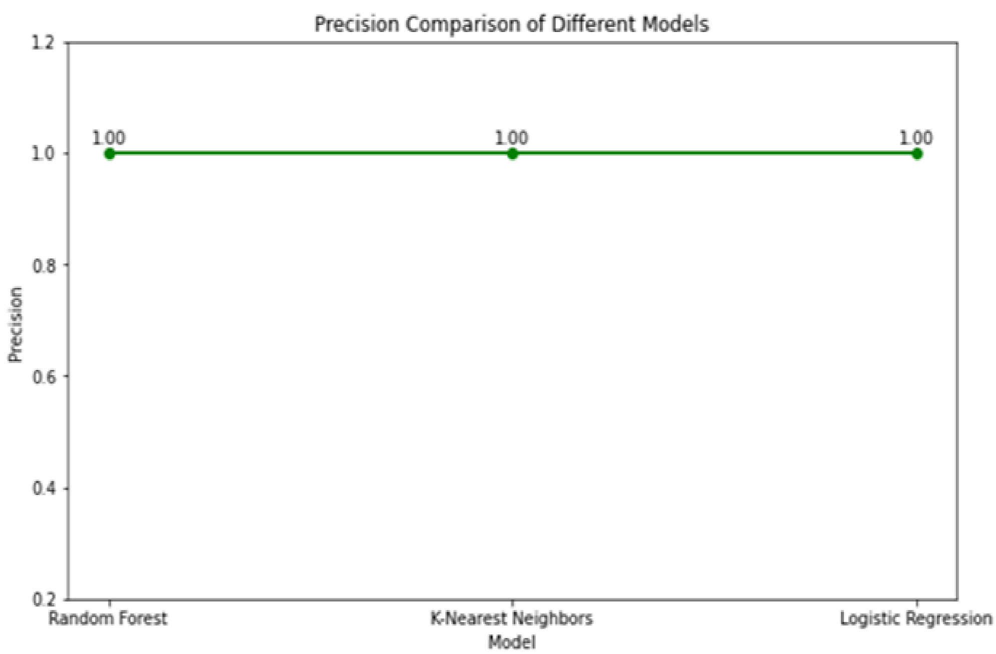

Figure 14: Models Recalls on PTBDB Normal Database

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Tables:

Table 5: Similar Studies Results.

Mission	***Model***	***Accuracy***	***Reference***
Heart Disease Prediction	*DT*	*98.29%, 77.05%*	*[5], [3]*
	RF	*96.82%, 86.89%*	*[5], [3]*
	KNN	*86.09%, 57.83%, 83.16%, 88.50%*	*[5], [3], [22], [23]*
	SVM	*88.04%, 73.77%*	*[5], [3]*
	BN	*79%*	*[21]*
	J48	*77.33%*	*[21],*
ECG Classification	*KNN*	*91.33% and 93%*	*[24]*
	BiDLNet	*97.5% and 91.5%*	*[7]*
	TERMA	*82.2% and 84.2%*	*[2]*

Table 6: Our Results.

Mission and Models		***Accuracy and Database***	
Mission	***Model***	***Heart DB***	***HDP DB***
Heart Disease Prediction	*DT*	*98.54%*	*68.52%,*
	RF	*98.54%*	*79.73%*
	KNN	*73.17%*	*------*
	SVM	*68.29%*	*------*
	NN	*75.12%*	*81.48%*
	LR	*------*	*92.59%*
Mission and Models		***Accuracy and Database***	
Mission	***Model***	***MIT-BIH Test***	***PTBDB Normal***
ECG Classification	*KNN*	*97.36%*	*97.87%*
	RF	*97.47%*	*99.90%*
	LR	*91.42%*	*99.90%*