Educational Administration: Theory and Practice
2024, 30(2) 1531 - 1538
ISSN: 2148-2403
https://kuey.net/

Research Article

Performance Evaluation Of AI Models In Early Heart Disease Diagnosis

Thejaswini S[1*], Mamatha K R[2], Rashmi N[3], Dr. Girish H[4]

[1*]Dept. of ETE, B M S Institute of Technology and Management Bengaluru, Karnataka, India.
[2]Dept. of ECE, B M S Institute of Technology and Management Bengaluru, Karnataka, India
[3]Dept. of ECE, B M S Institute of Technology and Management Bengaluru, Karnataka, India.
[4]Professor, Department of ECE, Cambridge Institute of Technology, Bangalore, Karnataka, India.

Citation: Thejaswini S, et.al (2024) Performance Evaluation Of AI Models In Early Heart Disease Diagnosis, *Educational Administration: Theory and Practice*, *30(2)* 1531 - 1538
Doi: 10.53555/kuey.v30i2.7918

ARTICLE INFO

ABSTRACT

Machine learning is increasingly being utilized throughout various stages of healthcare, particularly in predicting conditions like locomotor disorders and heart disease. Treating heart disease through therapies and surgeries in hospitals and clinics can be prohibitively expensive. As a result, early detection becomes crucial, allowing individuals to take preventive measures before more severe health complications occur. Heart disease is a widespread concern today, often linked to factors such as excessive alcohol use, smoking, and lack of physical exercise. Over time, machine learning has shown great potential in decision-making and prediction by leveraging large datasets from the healthcare sector. In this project, we developed models to predict heart disease by analyzing various heart-related factors of affected individuals, employing machine learning techniques. Early diagnosis can help high-risk patients adopt lifestyle changes, potentially preventing complications and representing a major step forward in medical care.

Keywords— Heart Disease, Coronary Sickness, Machine Learning

Introduction

Cardiovascular disease (CVD), a prevalent form of heart disease, remains one of the leading global causes of mortality, contributing to over 30% of deaths worldwide. Without intervention, global fatalities are expected to rise to 22 million by 2030. Plaque buildup in arteries can impede blood flow, leading to heart attacks or strokes. Key risk factors for heart disease include physical inactivity, poor diet, and excessive alcohol or tobacco use. These risks can be mitigated through lifestyle changes, such as reducing salt intake, eating more fruits and vegetables, engaging in regular physical activity, and avoiding alcohol and tobacco. One proposed solution to these issues involves utilizing patient records from various healthcare facilities for analysis. A decision support system (DSS) can help physicians make informed decisions, reduce unnecessary testing, and ultimately save time and money. Hospital management systems have increasingly been used to handle patient data, generating large datasets. A DSS leveraging the Naïve Bayes (NB) algorithm was created to predict heart disease, using key features extracted from historical databases like the Cleveland dataset. A web-based application facilitates this prediction process [1].

In the early stages of heart failure disease (HFD), neurohormonal regulatory mechanisms are triggered, which, although initially compensatory, can lead to worsened ventricular dysfunction, shortness of breath, fluid retention, and changes in the heart's structure. Treatment options for HFD range from lifestyle modifications to medical devices like defibrillators and pacemakers [2] . However, frequent hospitalizations due to acute HFD decompensation remain a leading cause of healthcare costs. Studies highlight that heart diseases, particularly HFD, are a significant public health concern. Early detection is key to effective treatment and management.

HFD is emerging as a condition associated with hypertension, insomnia, and other heart diseases. ECG can detect HFD by identifying variations in heartbeats, while magnetocardiography (MCG) shows promise as a noninvasive tool for detecting ischemic heart disease (IHD). Though MCG offers higher signal quality than

ECG, it requires extensive interpretation and has limited clinical application. An autonomous system for early ischemia detection could greatly aid clinicians [3].

Predictive models for heart disease can help identify high-risk individuals, improve diagnosis, and guide treatment decisions, potentially reducing mortality. Clinical decision support systems (CDSS) are increasingly used to assess heart disease risk and recommend appropriate interventions, improving decision-making and preventive care. Coronary artery disease (CAD), or ischemic heart disease (IHD), is the leading cause of death among adults over 35 in many countries, including China. CAD occurs when coronary artery stenosis reduces blood flow to the heart, leading to myocardial damage, which can result in serious complications such as ventricular arrhythmia or sudden cardiac death [4].

As per the World Health Organization, heart disease causes 12 million deaths globally each year. The burden of cardiovascular disease has been rapidly increasing worldwide in recent years. Numerous studies have been conducted to identify the most significant factors contributing to heart disease and to accurately predict overall risk. Early diagnosis of heart disease is crucial in guiding lifestyle changes for high-risk patients, which helps reduce complications. The Random Forest algorithm has shown a prediction accuracy of 88.52%, performing competitively against other methods [5].

This paper aims to predict heart disease by analyzing patient data to classify whether individuals have the condition. The architecture of the model is illustrated in Figure 1. The process begins with collecting relevant patient details, followed by data preprocessing and feature selection. After these steps, machine learning algorithms are applied by splitting the data into training and testing sets. The model then makes predictions and provides the accuracy scores for the algorithms used.

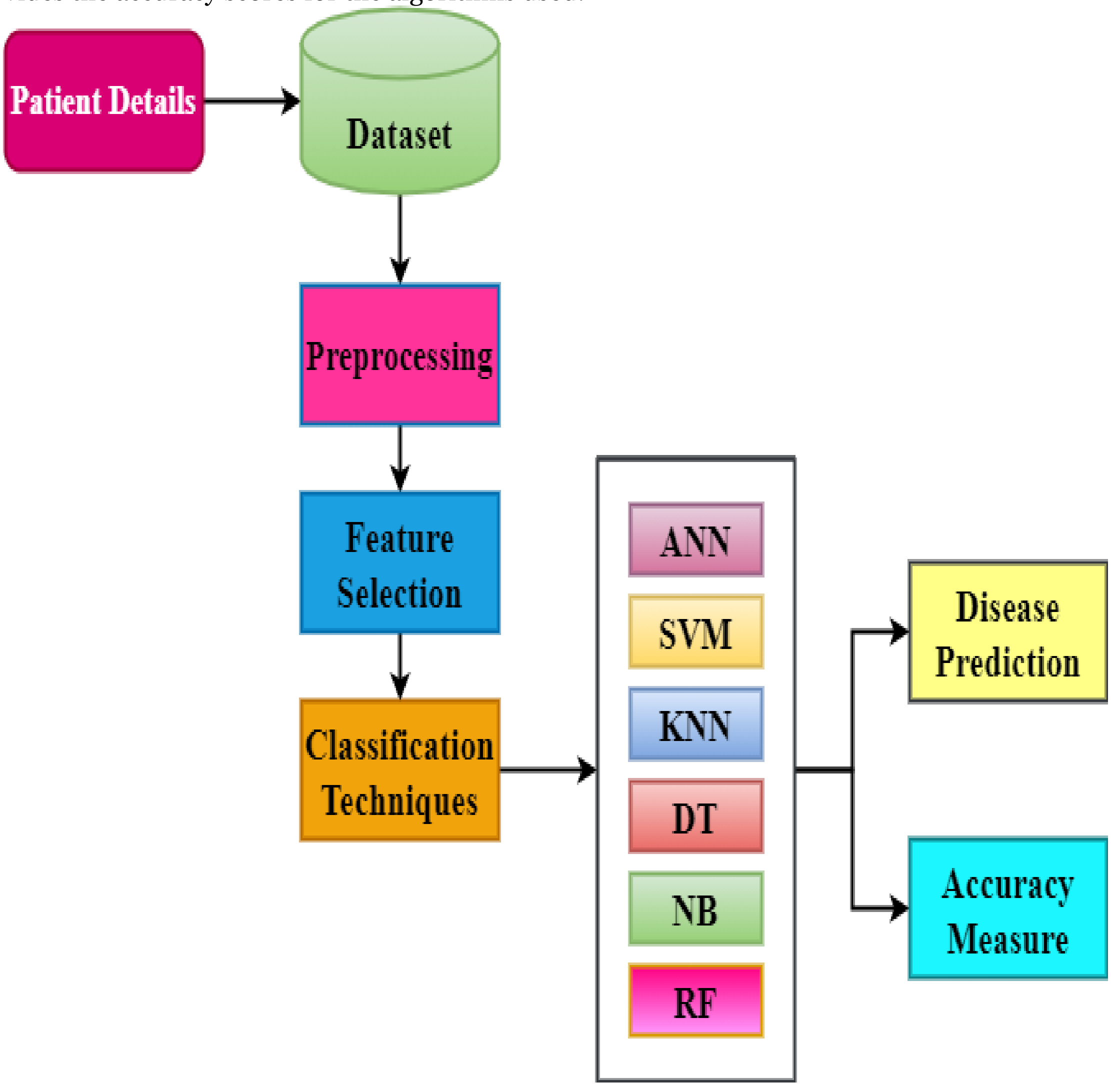

Figure 1: Architecture Diagram

METHODOLOGY

The proposed system methodology is implemented at different stages as shown in the following figure:

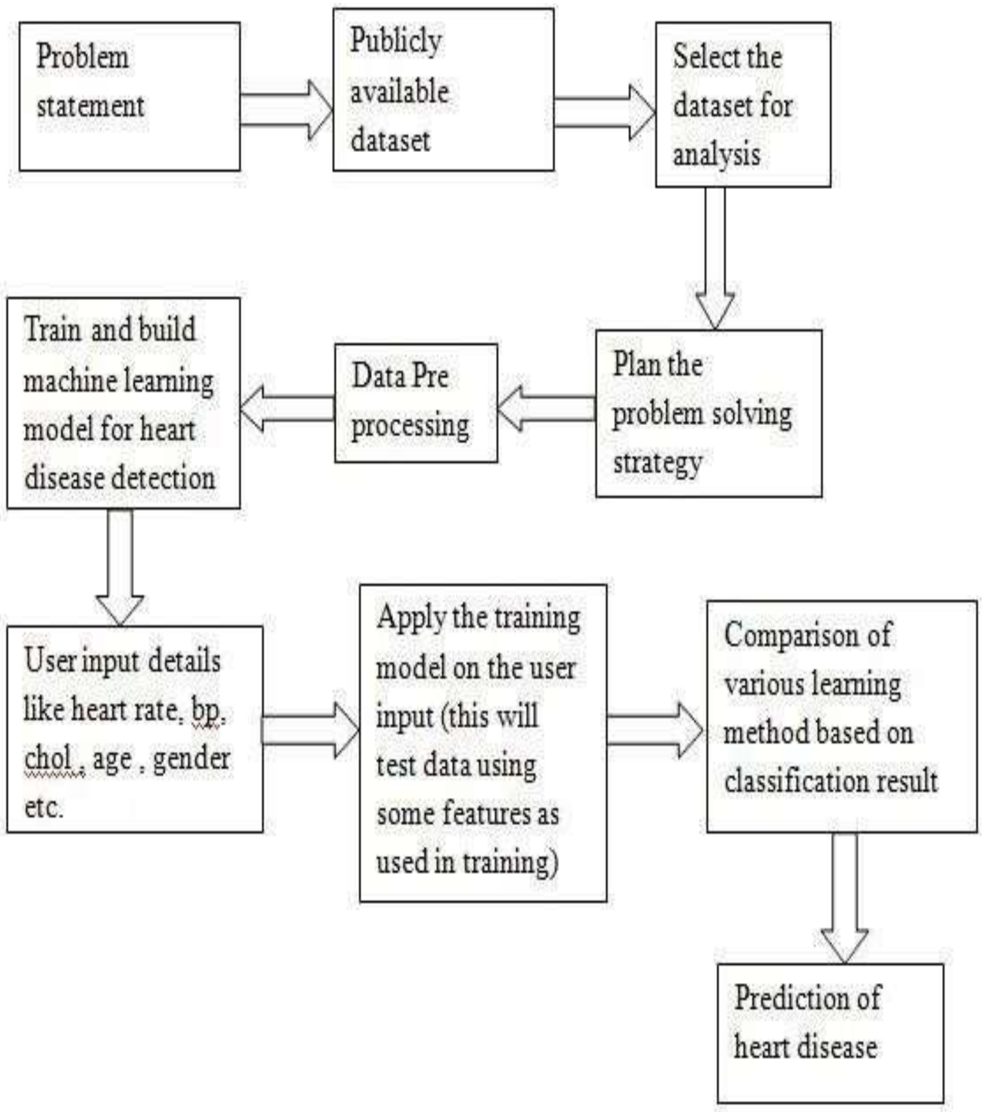

Figure 2: Methodology Diagram

Figure 3: Flow Diagram

The dataset is initially imported from a publicly available open-source platform. It then undergoes a preprocessing phase, during which missing values are addressed and duplicate entries are removed. Following this, the dataset is analyzed to prepare it for machine learning algorithms. Once preprocessing is complete, classification and prediction algorithms are applied. Finally, the model is validated, and accuracy metrics are generated.

DATA PREPROCESSING

Data preprocessing, also known as data cleaning, involves reviewing and modifying raw data by removing irrelevant information, correcting incomplete data, and converting data types as needed. This project focuses on predicting the presence or absence of heart disease in patients based on 14 health examination indicators and one target variable.

A. Loading the dataset

In machine learning projects, properly loading data is crucial. The most commonly used format is Comma Separated Values (CSV). The standard method for loading CSV files in Python is through its built-in libraries, particularly the csv module and the reader() function.

B. Cleaning the data

Data cleaning, or data cleansing, is a critical step in model building that follows data collection. It enhances data quality, leading to improved overall productivity. By cleaning the data, outdated or incorrect entries are removed, leaving only high-quality data for analysis.

C. Identifying null values

Understanding why data is missing is crucial for handling the remaining dataset effectively. If the missing values are random, the sample may still be representative. However, if the missing values follow a pattern, the analysis may become biased.

D. Exploratory Data Analysis (EDA)

EDA involves summarizing the main features of a dataset, often through visual plots. This step is crucial before modeling, as it provides insights into the data that inform the application of machine learning techniques.

E. Training and testing data

The processed data is split into 80% training data and 20% testing data. These subsets are then passed to the ANN and SVM models for fitting, prediction, and scoring..

MACHINE LEARNING ALGORITHMS

A. Artificial Neural Network (ANN)

Artificial Neural Networks are computational models designed to perform tasks more efficiently than traditional methods. ANN mimics the structure of biological neurons, similar to those in the human brain, where nerve cells are interconnected. This model consists of a network of linked elements (neurons) working together to complete tasks.

B. Support Vector Machine (SVM)

Support Vector Machine is a supervised machine learning algorithm. It works by using labeled training data to create classifiers in 1D, 2D, 3D, or even 4D space, which divide the data into distinct classes. These spatial classifiers are represented as points, lines (in 2D), planes (in 3D), or hyperplanes (in 4D).

C. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a simple algorithm used for classification tasks. The value of 'K' represents the number of neighbors considered. KNN is a type of classification where the function is approximated locally, and computation is delayed until the function is evaluated. Since the algorithm relies on distance, normalizing the training data, especially if features have different units or scales, can significantly enhance its accuracy.

D. Decision Tree (DT)

A Decision Tree is a supervised learning algorithm that can be used for regression problems. It consists of decision nodes and leaf nodes. The decision nodes split the tree into branches, while the leaf nodes represent the outcomes of these decisions. The tree visually represents all possible outcomes based on different conditions of the dataset.

E. Naïve Bayes (NB)

Naïve Bayes is a set of supervised learning rules based on Bayes' Theorem, often used for classification tasks. It is particularly useful for text-based datasets, like those found in spam detection. Naïve Bayes is known for its simplicity and efficiency in generating quick predictions, making it one of the fastest classification algorithms.

F. Random Forest (RF)

Random Forest is a widely used supervised machine learning algorithm suitable for both classification and regression tasks. It is based on the concept of ensemble learning, which combines multiple classifiers to enhance model performance. The more trees (classifiers) in the forest, the higher the accuracy and the lower the risk of overfitting.

RESULT

The results of the machine learning algorithms used to predict heart disease are summarized in terms of accuracy, as follows Artificial Neural Network (ANN) achieved an accuracy of 83.61%. ANN models are effective in handling complex patterns and large datasets due to their neuron-like structure, but they require significant computational resources. Support Vector Machine (SVM) achieved an accuracy of 81.97%. SVM is known for its ability to classify data into distinct groups by creating hyperplanes, but its performance can be sensitive to the choice of kernel and parameters. K-Nearest Neighbor (KNN) achieved a lower accuracy of 72.13%. KNN is a simple algorithm that classifies data points based on the closest 'K' neighbors, but its performance decreases when dealing with larger or more complex datasets, which may explain its lower accuracy here. Decision Tree (DT) also reached an accuracy of 81.97%. Decision Trees are intuitive and easy to interpret but can be prone to overfitting, particularly when the model becomes too complex. Naïve Bayes (NB) achieved a higher accuracy of 85.25%. Despite its simplicity, Naïve Bayes performs well, especially for problems with clear probabilistic relationships, making it efficient for heart disease prediction. Random Forest (RF) recorded the highest accuracy at 88.52%. Random Forest, which is an ensemble method combining multiple decision trees, is highly robust and less prone to overfitting, which explains its superior performance compared to the other algorithms.

Table 2: Accuracy score table

Sl. No.	**Methods Used**	**Accuracy**
1	Artificial Neural Network (ANN)	83.61%
2	Support Vector Machine (SVM)	81.97%
3	K-Nearest Neighbor (KNN)	72.13%
4	Decision Tree (DT)	81.97%
5	Naïve Bayes (NB)	85.25%
6	Random Forest (RF)	88.52%

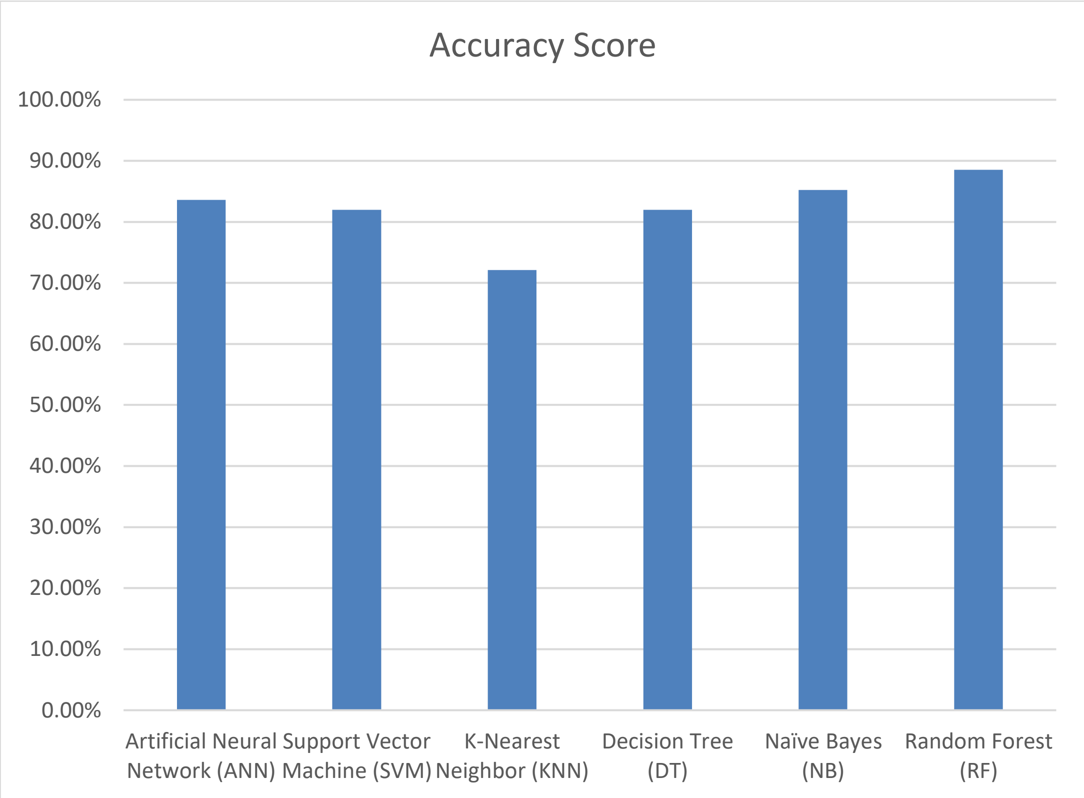

Figure 5: Comparison table

Conclusion

Heart disease is a significant concern in today's rapidly developing world, highlighting the need for an automated system to predict it at earlier stages. Such a system would assist doctors in diagnosing patients more efficiently, and benefit individuals by allowing them to monitor their health using the automated tool. Feature selection and prediction are crucial for the success of any automated system. By selecting features effectively, we can achieve better results in predicting heart disease. We have outlined several useful algorithms for feature selection, including ANN, SVM, KNN, DT, NB, and RF. Among these, the Random Forest algorithm achieved the highest accuracy. In this paper, we propose a method for heart disease prediction using some of these machine learning techniques, which demonstrated excellent accuracy for better estimation results.

REFERENCES

[1] Predicting Heart Disease at Early Stages using Machine Learning: A Survey(ICESC 2020) Rahul Katarya, Polipireddy Srinivas, Department of Computer Science and Engineering, Delhi Technological University, Delhi, India.

[2] Heart Disease Prediction Using Machine Learning Techniques (IRJET 2020)Galla Siva Sai Bindhika, Munaga Meghana, Manchuri Sathvika Reddy, Rajalakshmi Department of Computer Science, R.M.D. Engineering College.

[3] Predicting Heart Disease using Machine Learning Algorithm (IRJET 2019)Anagha Sridhar1, Anagha S Kapardhi Dept. of Information Science and Engineering, NIE, Mysore, India

[4] Heart Disease Detection Project Report. Member: Yangguang He, Xinlong Li, Ruixian Song IoT based home automation using Blynk cloud, J. Chetan Naik, B. S. Deeksha, H. Girish, Gurusharan Kashetty, T. G. Shivapanchakshari, Manohar Sai Polakam, Journal: AIP Conference Proceedings, AIP Conf. Proc. 2965, 020006 (2024), https://doi.org/10.1063/5.0215427, Published: July 2024

[5] IoT based fire alarm system using Blynk app, T. Bhargava, H. Kishor Kumar, Pramod Mahalingappa Godi, H. Girish, J. Chetan Naik, Journal: AIP Conference Proceedings, AIP Conf. Proc. 2965, 020007 (2024), https://doi.org/10.1063/5.0212325, Published: July 2024

[6] Dr. GIRISH H et. al., "Design and Analysis of a Machine learning based Agriculture bot", Tuijin Jishu/Journal of Propulsion Technology ISSN: 1001-4055 Vol. 44 No. 6 (2023)

[7] Dr. GIRISH H et. al., "Integrating Disaster Data Clustering with Neural Networks for Comprehensive Analysis", Tuijin Jishu/Journal of Propulsion Technology ISSN: 1001-4055 Vol. 45 No.1 (2024).

[8] Dr. GIRISH H et. al., "ROBOTIC INNOVATION IN E-DELIVERY SYSTEMS: A DESIGN AND MODELING APPROACH", The Seybold report ISSN 1533-9211, 2023.

[9] Dr. GIRISH H et. al., "Machine Learning based Agriculture Bot", International Conference on Smart Technologies, Communication and Robotics [ICSCR-2023] in association with IJSRSET Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com) doi : https://doi.org/10.32628/IJSRSET

[10] Dr. GIRISH H et.al., "A GSM-Based System for Vehicle Collision Detection and Alert". International Conference on Smart Technologies, Communication and Robotics [ICSCR-2023] in association with IJSRSET Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com) doi : https://doi.org/10.32628/IJSRSET

[11] H. Girish, T. G. Manjunat and A. C. Vikramathithan, "Detection and Alerting Animals in Forest using Artificial Intelligence and IoT," 2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC), 2022, pp. 1-5, doi: 10.1109/ICAECC54045.2022.9716679.

[12] A. Devipriya, H. Girish, V. Srinivas, N. Adi Reddy and D. Rajkiran, "RTL Design and Logic Synthesis of Traffic Light Controller for 45nm Technology," 2022 3rd International Conference for Emerging Technology (INCET), 2022, pp. 1-5, doi: 10.1109/INCET54531.2022.9824833.

[13] S. Vaddadi, V. Srinivas, N. A. Reddy, G. H, R. D and A. Devipriya, "Factory Inventory Automation using Industry 4.0 Technologies," 2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET), 2022, pp. 734-738, doi: 10.1109/GlobConET53749.2022.9872416.

[14] T G Manjunath, A C Vikramathithan, H Girish, "Analysis of Total Harmonic Distortion and implementation of Inverter Fault Diagnosis using Artificial Neural Network", Journal of Physics: Conference Series, Volume 2161, 1st International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS 2021) 28-30 October 2021, Manipal, India. https://iopscience.iop.org/issue/1742-6596/2161/1

[15] GIRISH H , "Internet of Things Based Heart Beat Rate Monitoring System", © September 2022 | IJIRT | Volume 6 Issue 4 | ISSN: 2349-6002 IJIRT 156592 INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY 227

[16] Dr Girish H, Dr. Mangala Gowri , Dr. Keshava Murthy , Chetan Naik J, Autonomous Car Using Deep Learning and Open CV, PAGE NO: 107-115, VOLUME 8 ISSUE 10 2022, Gradiva review journal, ISSN NO : 0363-8057, DOI:10.37897.GRJ.2022.V8I8.22.50332

[17] Girish H. "Intelligent Traffic Tracking System Using Wi-Fi ." International Journal for Scientific Research and Development 8.12 (2021): 86-90.

[18] Girish H, Shashikumar D R, "Emission monitoring system using IOT", Dogo Rangsang Research Journal, UGC Care Group I Journal, ISSN: 347-7180, Vol – 8 Issue-14, No.6, 2020.

[19] Girish H, Shashikumar D R, "Display and misson computer software loading", International journal of engineering research & Technology (IJERT), ISSN: 2278-0181, Vol – 10 Issue-08, No.13, August 2020.

[20] Girish H, Shashikumar D R, "A Novel Optimization Framework for Controlling Stabilization Issue in Design Principle of FinFET based SRAM", International Journal of Electrical and Computer Engineering (IJECE) Vol. 9, No. 5 October 2019, pp. 4027~4034. ISSN: 2088-8708, DOI: 10.11591/ijece.v9i5.pp.4027-4034

[21] Girish H, Shashikumar D R, "PAOD: a predictive approach for optimization of design in FinFET/SRAM", International Journal of Electrical and Computer Engineering (IJECE) Vol. 9, No. 2, April 2019, pp. 960~966. ISSN: 2088-8708, DOI: 10.11591/ijece.v9i2.pp.960-966

[22] Girish H, Shashikumar D R, “SOPA: Search Optimization Based Predictive Approach for Design Optimization in FinFET/SRAM”, © Springer International Publishing AG, part of Springer Nature 2019 Silhavy (Ed.): CSOC 2018, AISC 764, pp. 21–29, 2019. https://doi.org/10.1007/978-3-319-91189-2_3.

[23] Girish H, Shashikumar D R, “Cost-Effective Computational Modelling of Fault Tolerant Optimization of FinFET-based SRAM Cells”, © Springer International Publishing AG 2017 R. Silhavy et al. (eds.), Cybernetics and Mathematics Applications in Intelligent Systems, Advances in Intelligent Systems and Computing 574, DOI 10.1007/978-3-319-57264-2_1.

[24] Girish H, Shashikumar D R, “A Survey on the Performance Analysis of FinFET SRAM Cells for Different Technologies”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-4 Issue-6, 2016.

[25] Girish H, Shashikumar D R, “Insights of Performance Enhancement Techniques on FinFET-based SRAM Cells”, Communications on Applied Electronics (CAE) – ISSN: 2394-4714, Foundation of Computer Science FCS, New York, USA .Volume 5 – No.6, July 2016 – www.caeaccess.org

[26] Girish H, Shashikumar D R, “DESIGN OF FINFET”, International Journal of Engineering Research ISSN: 2319-6890) (online), 2347-5013(print) Volume No.5 Issue: Special 5, pp: 992-1128, doi: 10.17950/ijer/v5i5/013 2016.

[27] Dr. MANGALA GOWRI S G, Dr. SHASHIDHAR T M, Dr. SUNITHA R, SHYLAJA V, Dr. GIRISH H, “DESIGN OF 3-BIT CMOS WALLACE MULTIPLIER” , Seybold report, ISSN 1533-9211, Volume 18, Page No: 1260-1271 DOI: 10.5281/zenodo.8300481

[28] Dr. Mangala Gowri S G, Dr. Girish H, Ramesh N, Dr. Nataraj Vijaypur, “IOT based plant monitoring system and smart irrigation using new features” ResMilitaris, vol.13, n°2, January Issue 2023 https://resmilitaris.net/menu-script/index.php/resmilitaris/article/view/3312/2608

[29] Girish H.2023, Smart Theft Securityvehicular System Using Iot. Int J Recent Sci Res. 14(02), pp. 2881-2884. DOI: http://dx.doi.org/10.24327/ijrsr.2023.1402.0591

[30] Dr. Mangala Gowri S G, Dr. Girish H, Dr. Santosh Dattatray Bhopale, Weed Detection based on Neural Networks, International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211 Volume 11, Issue 3, March-2023, Impact Factor: 7.429, Available online at: www.ijaresm.com DOI: https://doi.org/11.56025/IJARESM.2023.11323351

[31] Shashidhara, K.S., Girish, H., Parameshwara, M.C., Rai, B.K., Dakulagi, V. (2023). A Novel Approach for Identification of Healthy and Unhealthy Leaves Using Scale Invariant Feature Transform and Shading Histogram-PCA Techniques. In: Shetty, N.R., Patnaik, L.M., Prasad, N.H. (eds) Emerging Research in Computing, Information, Communication and Applications. Lecture Notes in Electrical Engineering, vol 928. Springer, Singapore. https://doi.org/10.1007/978-981-19-5482-5_47

[32] Mamatha K R, Hariprasad S A, Thippeswamy G, & Girish H. (2023). Real-time Animal Detection and Alert System using IoT and Deep Learning. Educational Administration: Theory and Practice, 29(4), 1050–1054. https://doi.org/10.53555/kuey.v29i4.5982