Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 2, February : 2024

ADVANCES IN PREDICTIVE CARDIOLOGY: A COMPREHENSIVE SYSTEMATIC REVIEW OF AI/ML APPLICATIONS IN CVD PREDICTION

Maheshkumar Bargaje, Research Scholar, Computer/IT Engineering, Gujarat Technological University, Ahmedabad, India maheshsb.it@gmail.com
Dippal Israni, Lecturer, Information Technology Department, R. C. Technical Institute, Ahmedabad, India dippalisrani90@gmail.com

***Abstract*— Our heart is known to be the most significant organ in our body. Recently many young adults in the age group of 10 to 30 are reported with a sudden heart failure. Among the most complex issues in healthcare is the prediction of CVDs. To determine the cause, it requires considerable amount of diligence and time, specifically for medical practitioners and other healthcare providers. Stroke disease is commonly fatal or it may cause severe disability. Data analytics, Machine learning and AI tools are useful for such prediction of CVDs. Recently, many researchers are working on retinal images to predict CVDs. Many reputed datasets are available which are consists of huge amount of patient health related data. Machine learning techniques such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision tree, Naive Bayes, Fuzzy Logic and Artificial Neural Network (ANN) are used to predict the CVDs well in advance. This review article includes an overview of the existing algorithms and an overview of the previous work. It aims to give reader a comprehensive and systematic overview of recent AI/ML applications in CVD forecast.**

***Index Terms*—Cardio Vascular Disease (CVD), Machine learning, Artificial Intelligence, KNN, SVM, ANN**

I. INTRODUCTION

According to a recent World Heart Federation research, CVD accounts for one in every three fatalities [1]. According to WHO figures, more than 2.36 crore individuals might die due to CVD by 2030, largely by heart failure as well as strokes [2]. Timely treatment is crucial during heart related emergencies [3]. Predicting any such fatal situation beforehand, gives us a chance to survive [2]. Thus, proactive initial prevention as well as prompt detection of predictive indicators are crucial [4]. A Sudden rise in the cases of cardiovascular disease among young hearts in India is the headlines of many newspapers recently. Few to mention here (Source: TOI, Dated: June 30th, 2023) [5]:

a) A teenager girl (17 years of age) from Navsari (India) and a young boy (28 years of age) from Rajkot (India) fainted and died at their respective schools and colleges as a result of heart failure.
b) On February 25, a 19-year-old kid died of heart failure while dancing at a wedding in Telangana. Three days ago, a young officer (24-year age), while working out at a gym in Hyderabad, died of cardiac arrest.
c) A young man in Meerut died from a heart attack while out for a walk with friends.
d) Hetal of Bhavnagar (Gujarat) died from a heart failure during her wedding rituals.
e) Shyam Yadav of Telangana died from sudden cardiac arrest while playing badminton.
f) Sushmita Sen, a fit celebrity, recently recovered from a heart attack.

As mentioned in TOI [5], According to the EMRI (Emergency Management and Research Institute) 108 emergency services, they recorded 6,780 cardiac emergencies in youth under 30 years of age. The number of cardiac emergencies in youth has already reached 4,027 or 60% of last year's total. If the current rate continues, the number of heart crises in young people is projected to reach 8,258 by the end of the year, representing a 22% year-on-year increase. The age group of 10-30 years accounts for 14% of all emergency calls.

Above mentioned cases (a to f) and the data of EMRI 108 emergency services clearly shows a sudden rise in CVDs. Not only aged people, but teenagers are also now started having sudden cardiac emergencies. This situation might get worse after COVID like pandemics. We are in an urgent need of an AI/ML based predictive system which can alert us before facing any such CVD related issues.

Table 1 lists types of cardiovascular diseases with a short description of each. In this article, the term CVD is used to refer to these diseases collectively.

Table 1. Types of CVDs [6]:

Type of CVD	Terminology
Arrhythmia	Heartbeat rate appears abnormal.
Cardiac arrest	Sudden disfunction of heart.
Congestive heart failure	Improper pumping of blood by heart, which is a chronic problem.
Congenital heart disease	A malformation of the heart that takes place before the birth.
Coronary artery disease	The primary arteries and veins of the heart can be hurt and illness can emerge in them.
High Blood Pressure	It is a condition in which the blood's force on the arterial walls is excessive.
Peripheral artery disease	A narrowing of arteries that inhibits the circulation of blood to the limbs.
Stroke	Damage to the brain occurs when blood circulation is interrupted.

Several diagnostic procedures are necessary for predicting cardiac disease. Healthcare personnel's incompetency could end up in incorrect forecasts [7]. Around fifty percent of individuals who receive diagnoses of CVD suffer deaths within the first one to two years of diagnosis [8]. Properly trained on sufficient data, algorithms for ML are potentially useful in diagnosing illnesses [9]. P. Israni utilized machine learning techniques to diagnose breast cancer [10]. Islam J. et al employed DL (Deep Learning) methods to detect Alzheimer's disease [11]. Not only in medical field, the ML methods are also useful in many other disciplines such as multimedia. For e.g., Machine Learning techniques can be used for camera autofocusing [12]. It can also predict the emotions from audio (music) data [13]. Nevertheless, many CVD-related disorders have been found to be totally treatable if detected earlier [14].

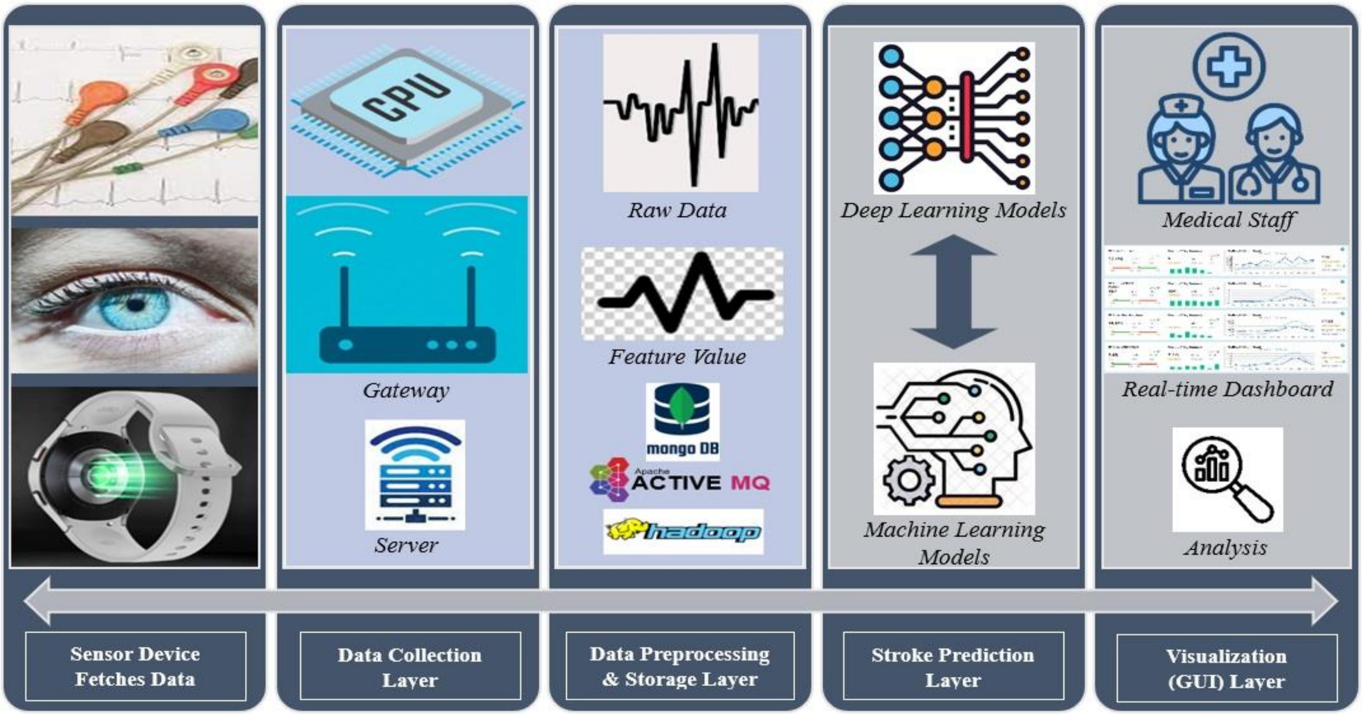


Figure 1. A CVD prediction system based on AI that uses Electrocardiogram and other biological data

Figure 1 illustrates an example of an AI-based real-time cardiovascular disease prediction system that uses electrocardiogram and other biological data. Sensor devices such as optical heart rate sensor (smart watch, smart ring), ECG, PPG, Holter monitor, retinal scan etc. are used to fetch raw data such as heart rate, resting blood pressure, blood sugar level, status of major vessels and many more. This data is collected and stored on servers in the data collection layer. Next, during data preprocessing phase, the raw data is processed and actual feature value is extracted and stored. These feature values are then fed into the pre-trained ML/DL models for testing. ML/DL model performs prediction and results are fed into the visualization layer. This layer shows the results of the model in a human understandable format using graphs, charts etc. Medical professionals and experts analyze these results and take prompt decision.

II. RELATED WORK

Numerous studies have been conducted in the medical domains on CVD prediction systems using different machine learning algorithms and data mining techniques.
Jaehak yu et al, [15] In their study, they present a system that uses many biological signals, such as PPG and ECG, collected while walking in everyday activities of the aged to enable semantic interpretation of disorders in the elderly. The dataset used here is composed of bio signals gathered through 287 aged individuals suffering from stroke and 287 healthy individuals. The suggested approach may identify and forecast determinants of prognosis of stroke sickness among aged people by capturing multiple PPG (Photoplethysmography) and ECG (Electrocardiogram) biosignals in the real-time. They have verified that they can successfully predict 91.56% C4.5, 97.51% and 99.15% with Decision Tree, Random Forest and CNN-LSTM models respectively for DL (Deep-Learning) by dividing stroke and aged-people into 10-folder CV (cardiovascular) datasets.

Table 2. An analysis of different algorithms in a review of the available literature.

Year	Journal	Title	Technique	Acc %	Gap (Issues)
2019	Elsevier-Telematics and Informatics [Q1]	Identification of significant features and data mining techniques in predicting heart disease	Used 7 ML models & feature selection	87.41	-Low accuracy, -Too many models were used (High computation time)
2020	IEEE Access, 8,pp.210318-210327 [Q1]	Identifying Stroke Indicators Using Rough Sets	Feature selection & RST	88.10	-Only binary features can be used
2021	IEEE Access, 9,pp.135210-135223 [Q1]	An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets	Adapt with highly biased subset of data	89	-No feature selection -Low accuracy
2022	IEEE Access, 10,pp.51079-51092 [Q1]	Quickly Convert Photoplethysmography to Electrocardiogram Signals by a Banded Kernel Ensemble Learning Method for Heart Diseases Detection	PPG to ECG prediction	81	- Computation and stability issue
2022	IEEE Access, 10,pp.43623-43638 [Q1]	AI-Based Stroke Disease Prediction System Using ECG and PPG Bio-Signals	CNN-LSTM	88.3	-Wearing Bio-signal sensors for real time monitoring
2022	Sensors [Q1]	Cardiovascular Disease Diagnosis from DXA Scan and Retinal Images Using Deep Learning	Retinal image and dual-energy X-ray Absorptiometry (DXA).	78.3	-QBB Dataset (Consists of Qatar people data only) -Low accuracy
2023	Translational Vision Science and Technology [Q1]	A Systematic Review and Meta-Analysis of Applying Deep Learning in the Prediction of the Risk of Cardiovascular Diseases from Retinal Images	Retinal Images were used with Deep Learning for prediction	68 to 81	-Low accuracy -Usability issues -High heterogeneity -Lack of quality assessment tools

Wen-hsien ho et al, [16] points out that PPG signals of poor-quality signals were challenging to use for sickness detection and were primarily utilized to record SpO2 readings and calculate heartbeats. By advancing these widely available, low-cost devices to a practical level appropriate throughout routine medical execution, this work promotes cardiac applications. Here they have used PhysioNet.com's freely available MIMIC III database (multi-parametric) which comprises 25,328 ICU entries. Their method, which is based on cardiology's hemodynamics and electro-fluid mechanics, achieves a notable level of reconstruction performance accuracy, and produced ECGs accurately reproduce the waveform properties of the actual ECG signals. Their method helps a large number of high-risk, previously fit individuals visit doctors' offices before their cardiac condition irrevocably deteriorates by detecting suspected cardiovascular signs from their residence. According to their research, PPG to the ECG signal translation is reliable and accurate. Once ECG waveforms have been

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 2, February : 2024
obtained from PPGs, many disorders, including severe atrial infarction and the arrhythmia, can be prescreened in regular monitoring cameras.
Tsatsral amarbayasgalan et al, [17] in their study, a method to predict the risk of CHD using 2 DNN models was presented and applied to the KNHANES dataset. Suggested strategy tackled with the issue of creating an effective training dataset using PCA as well as VAE models by identifying and refining a heavily biassed set which influences the model's effectiveness. Two DNN models learned via separated training by utilizing the PCA model enhanced the effectiveness of a CHD risk predictor derived from a single Deep neural network (DNN) (0.826 of specificity, 0.873 of accuracy, 0.899 of recall, 0.899 of f-measure, 0.90 of precision and 0.862 of AUC).
Muhammad salman pathan et al, [18] proposes that RST is a feature selection approach that is optimized for detecting and ranking the most essential characteristics in a big dataset. Its primary idea is to portray defective or ambiguous knowledge as closely as possible with known knowledge. They suggested an effective feature selection strategy based on rough set theory in this work. The suggested approach may discover crucial stroke signs from a huge dataset in order to create excellent stroke prediction models. On a dataset of 29,072 records for patients with ten shared features, they applied the proposed methodology. When in comparison to other characteristics, they found that the four most important factors: age, heart disease, hypertension, and average blood sugar are the ones that most strongly predict the diagnosis of stroke.
Pronab ghosh et al, [19] According to their study, the Relief feature selection method might yield a feature set with strong correlations, which could subsequently be utilized in conjunction with different ML (machine learning) algorithms. Here they employed five datasets combinedly - Cleveland, Long Beach VA, Switzerland, Hungarian and Stat log – along with feature selection. Furthermore, the study discovered that RFBM has much greater accuracy than previous studies and works particularly well high impact features (as determined by a feature selection algo.). Using ten features, accuracy of 99.05% was achieved.
Evaluation Parameters:
Prediction using ML is based on evaluation parameters. Below listed four parameters are the most common and effective during model testing phase.

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1) \qquad Precision = \frac{tp}{tp+fp} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$Recall = \frac{tp}{tp+fn} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3) \qquad F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

tp, fp, tn, and fn stands for true positives, false positives, true negatives and false negatives respectively.
Features:
After thorough literature analysis, the list of features mentioned below are found to be most prominent and frequently used to detect various CVDs. Features: 1) Age, 2) Heart-rate, 3) Gender: Female/Male/Other, 4) Blood-Pressure, 5) Blood-Sugar level, 6) Health of Blood vessels of retina, 7) Cholesterol levels and 8) Electrocardiographic data (ECG/PPG)
Dataset:
Extensive literature review reveals that there is sufficient availability of number of different publicly available datasets. Few most popular and frequently used public datasets are listed in Table 3.

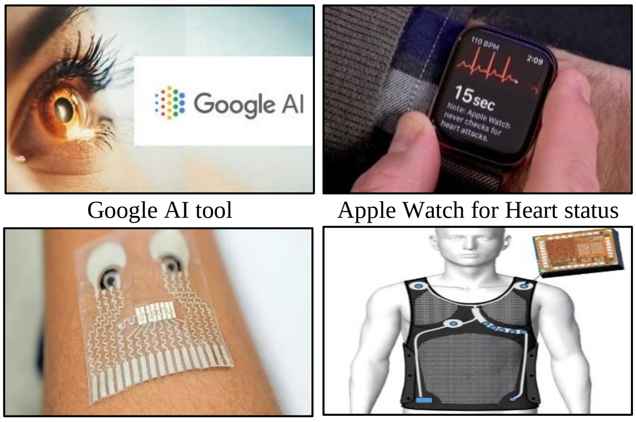
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 2, February : 2024

Table 3. Publicly available CVD datasets:

Dataset	Description
cardio_train [20]	This CVD dataset contains seventy thousand patient data records, with Eleven features, and a target.
heart_statlog_ cleveland_hungary_ final [21]	UCI machine learning cardiovascular disease dataset of Cleveland and Hungarian heart disease.
NHIS National Cardiovascular [22]	National Health Interview Survey (NHIS) - National Cardiovascular Disease Surveillance Data

III. RESEARCH GAPS (ISSUES):

1. ArdiaMobile and Apple Watch have exact market positioning overlap [23]. Other laboratory-grade gadgets, such as a one-time-use patch with a seven-day life [24] and a fabric-based IC in apparel [25],



Google AI tool
Apple Watch for Heart status
One-time use Patch
IC in Apparel

most likely still need to work through some usability issues at this time.

Figure 2. Usability issues in current products

2. The features found in the patient's medical records are diverse and highly variable [26]. Any disease's diagnosis relies on connecting its symptoms, which is dependent on how quickly diagnosis can be made in the moment. As a result, any diagnostic system must operate with extreme precision and speed [27].

3. Minor currents of electricity on the skin's surface can be detected with comparatively costly equipment that needs to be handled by a professional. As a result, it is challenging to keep track of cardiac activity in healthy individuals while placing the wiring in place for a long time in an everyday environment.

4. According to recent research, the prediction must have at least 14 characteristics in order to be accurate and trustworthy [28]. In order to precisely forecast heart disease, present-day researchers are having difficulty combining these features with the right machine learning methods.

5. Overfitting issues could arise in the machine algorithms for learning used for classification [29].

6. Accurately predicting heart disease may still be hampered by a number of problems, including a lack of thorough analysis, feature selection, limited medical datasets, and the use of machine learning algorithms.

IV. CONCLUSION

The recent researches that employ ML and optical data to forecast cardiovascular disease (CVD) outcomes are assessed qualitatively in this systematic review. Although there is little information available on the long-term prediction of incident CVD, a wide range of expected outcomes were investigated. To validate and enhance the algorithms, more research is required, especially in large-scale longitudinal cohorts. Further research is also necessary to show how the technology can be used in practical settings. The model needs to be utilized in the future with multiple feature selection algorithms; using a random forest classifier is an additional option. Also, we saw an urgency of a real-time system which alerts user when any heart related abnormalities are detected without disturbing the day-to-day activities. This review's main objective is to identify areas for improvement in the body of work already done and to render the model usable as well as user-friendly in real-life circumstances.

REFERENCES

[1] Thanga Selvi, R., and I. Muthulakshmi. "RETRACTED ARTICLE: An optimal artificial neural network based big data application for heart disease diagnosis and classification model." *Journal of Ambient Intelligence and Humanized Computing* 12, no. 6 (2021): 6129-6139.
[2] Bazoukis, George, Stavros Stavrakis, Jiandong Zhou, Sandeep Chandra Bollepalli, Gary Tse, Qingpeng Zhang, Jagmeet P. Singh, and Antonis A. Armoundas. "Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review." *Heart failure reviews* 26 (2021): 23-34.
[3] Mourao-Miranda, Janaina, Arun LW Bokde, Christine Born, Harald Hampel, and Martin Stetter. "Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data." *NeuroImage* 28, no. 4 (2005): 980-995.
[4] Abdellatif, Abdallah, Hamdan Abdellatef, Jeevan Kanesan, Chee-Onn Chow, Joon Huang Chuah, and Hassan Muwafaq Gheni. "An effective heart disease detection and severity level classification model using machine learning and hyperparameter optimization methods." *Ieee access* 10 (2022): 79974-79985.
[5] https://timesofindia.indiatimes.com/city/ ahmedabad/emri-108- logs-14-of-heart- emergencies-in-10-30- age-grp/ articleshow /101377301. cms (Dated: 30th June, 2023, Last accessed on: 5th December, 2023)
[6] Marimuthu, M., M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra. "A review on heart disease prediction using machine learning and data analytics approach." *International Journal of Computer Applications* 181, no. 18 (2018): 20-25.
[7] Pouriyeh, Seyedamin, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, and Juan Gutierrez. "A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease." In *2017 IEEE symposium on computers and communications (ISCC)*, pp. 204-207. IEEE, 2017.
[8] Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." *Mobile information systems* 2018 (2018): 1-21.
[9] Shamrat, FM Javed Mehedi, Md Abu Raihan, AKM Sazzadur Rahman, Imran Mahmud, and Rozina Akter. "An analysis on breast disease prediction using machine learning approaches." *International Journal of Scientific & Technology Research* 9, no. 02 (2020): 2450-2455.
[10] Israni, Priyanka. "Breast cancer diagnosis (BCD) model using machine learning." International Journal of Innovative Technology and Exploring Engineering 8, no. 10 (2019): 4456-4463.
[11] Islam, Jyoti, and Yanqing Zhang. "A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data." In *Brain Informatics: International Conference*, BI 2017, Beijing, China, November 16-18, 2017, Proceedings, pp. 213-222. Springer International Publishing, 2017.
[12] Israni, Dippal, Patel Sandip, and Shah Arpita. "Comparison of different techniques of camera autofocusing." In *Proceedings of first international conference on information and communication Technology for Intelligent Systems:* Volume 1, pp. 125-135. Springer International Publishing, 2016.
[13] Bargaje, Mahesh. "Emotion recognition and emotion based classification of audio using genetic algorithm-an optimized approach." In *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, pp. 562-567. IEEE, 2015.
[14] Makhlouf, Amina, Isma Boudouane, Nadia Saadia, and Amar Ramdane Cherif. "Ambient assistance service for fall and heart problem detection." *Journal of Ambient Intelligence and*

Humanized Computing 10 (2019): 1527-1546.
[15] Yu, Jaehak, Sejin Park, Soon-Hyun Kwon, Kang-Hee Cho, and Hansung Lee. "AI-based stroke disease prediction system using ECG and PPG bio-signals." *IEEE Access* 10 (2022): 43623-43638.
[16] Ho, Wen-Hsien, Chia-Te Liao, Yenming J. Chen, Kao-Shing Hwang, and Yanyun Tao. "Quickly Convert Photoplethysmography to Electrocardiogram Signals by a Banded Kernel Ensemble Learning Method for Heart Diseases Detection." *IEEE Access* 10 (2022): 51079-51092.
[17] Amarbayasgalan, Tsatsral, Van-Huy Pham, Nipon Theera-Umpon, Yongjun Piao, and Keun Ho Ryu. "An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets." *IEEE Access* 9 (2021): 135210-135223.
[18] Pathan, Muhammad Salman, Zhang Jianbiao, Deepu John, Avishek Nag, and Soumyabrata Dev. "Identifying stroke indicators using rough sets." *IEEE Access* 8 (2020): 210318-210327.
[19] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. De Boer. "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques." *IEEE Access* (2021):19304–19326.
[20]Svetlana Ulianova. (2018). cardio_train, Version 1. Retrieved from:
https://www.kaggle.com/datasets/sulianova/cardiovascular -disease-dataset.
[21]MANU SIDDHARTHA. (2020). heart_statlog_cleveland_hungary_final, Version 1. Retrieved from:
https://www.kaggle.com/code/sid321axn/stacked-ensemble-for- heart-disease-classification/data
[22]NHIS. (2016). National_Health_Interview_Survey_ NHIS_National_Cardiovascular, Version 1. Retrieved from: https://data.world/cdc/nhis-national-cardiovascular
[23] Ho, Wen-Hsien, Chia-Te Liao, Yenming J. Chen, Kao-Shing Hwang, and Yanyun Tao. "Quickly Convert Photoplethysmography to Electrocardiogram Signals by a Banded Kernel Ensemble Learning Method for Heart Diseases Detection." *IEEE Access* 10 (2022): 51079-51092.
[24] Bhagat, Yusuf A., Patrick Vcrdon, Sai Avuthu, Daniel Parsons, Mark Sussman, Girish Wable, and Ralph Hugeneck. "Like Kleenex for Wearables: A soft, strong and disposable ECG monitoring system." In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1-1. IEEE, 2018.
[25] Nawawi, Muhammad Muizz Mohd, Khairul Azami Sidek, Alaa KY Dafhalla, and Amelia Wong Azman. "Review on data acquisition of electrocardiogram biometric recognition in wearable smart textile shirts." In *Journal of Physics: Conference Series*, vol. 1900, no. 1, p. 012019. IOP Publishing, 2021.
[26] Roth, Gregory A., George A. Mensah, Catherine O. Johnson, Giovanni Addolorato, Enrico Ammirati, Larry M. Baddour, Noël C. Barengo et al. "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study." *Journal of the American College of Cardiology* 76, no. 25 (2020): 2982-3021.
[27] Gupta, Ankur, Rahul Kumar, Harkirat Singh Arora, and Balasubramanian Raman. "MIFH: A machine intelligence framework for heart disease diagnosis." *IEEE access* 8 (2019): 14659-14674.
[28] Singh, Dilbag, and Jasjit Singh Samagh. "A comprehensive review of heart disease prediction using Machine Learning." *Journal of Critical Reviews* 7, no. 12 (2020): 281-285.
[29] Fitriyani, Norma Latif, Muhammad Syafrudin, Ganjar Alfian, and Jongtae Rhee. "HDPM: an effective heart disease prediction model for a clinical decision support system." *IEEE Access* 8 (2020): 133034-133050.