IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Heart Disease Prediction Using Deep Learning Techniques

[1] D Nagabhushanam, [2] Varshitha K R, [3]Vinitha V, [4]Venkatesh P, [5]Srivenkat S
[1]Assistant Professor, [2,3,4,5]Student
Department of Computer Science and Engineering
Annamacharya Institute of Technology & Sciences, Tirupati, India.

ABSTRACT

Given its prominence as a leading cause of deaths worldwide, careful detection of this condition is still very important and current heart disease diagnosis methods are heavily dependent on experts, who however, need time to process patients and show the potential possibilities of human error. The research project is aiming to develop a deep learning-based system for heart disease prediction in order to make it efficient and accurate. To do such complex patterns extraction the system uses Convolutional Neural Networks and Long Short-Term Memory (LSTM) networks, which work on the blood pressure readings (t) together with patient's age (l) and cholesterol levels (b) as well as electrocardiogram (ECG) result. Feature selection methods are used to find acceptable information characteristics to reveal important predictive variables, improving the system understanding and the system performance at operation. Using a comprehensive dataset, the model completes training and its accuracy works to the optimal levels for it to be able to identify the patients within the appropriate risk groups. Robustness and reliability of the performance of the model has been established through the use of precision, recall, F1- score and ROC-AUC evaluation metrics. The system is designed to be working as a diagnosis tool that helps healthcare personnel to detect the heart diseases in a timely and convenient way. The proposed solution tries to minimize diagnosis delays and improve patient well-being and also assist in preventive healthcare operations through automation of predictions. A system for implementation within clinical operations could do this and change the direction of heart disease strategies by providing quick and reliable early diagnosis options to healthcare professionals.

KEYWORDS

heart disease prediction, deep learning, CNN, LSTM, neural networks, medical diagnosis, feature selection, healthcare analytics, patient risk assessment, ROC-AUC, precision, recall.

1. Introduction

At the moment, this ailment that leads to the death of the disease has achieved standing as one of the primary killers among both fatal diseases and deaths worldwide since it affects millions of people yearly. Cardiovascular WHO reports that CVDs trigger 17.9 million annual fatalities that correspond to 31% of world-wide fatalities [1]. Improving health outcomes and making healthcare resources more effective for patients and health care systems demands both timely diagnosis and urgent medical treatment for heart disease. Multiple diagnostic procedures include human mistakes and delays together with needing skilled clinical personnel who conduct electrocardiogram (ECGs) echocardiogram and stress tests. Healthcare diagnostic needs to become more productive with simplified procedures that deliver precise clinical assessment and workplace support for medical personnel in heart disease diagnosis and treatment. Deep learning systems which form part of ML technology achieve high performance when it comes to examining hard medical data and identifying hidden patterns and precise diagnostic evaluations.[2] Many studies conduct research to determine how combined AI technology with deep learning methods screen for heart diseases. Multiple predictive models utilized the Framingham Heart Study dataset together with Cleveland Heart disease dataset to identify high risk patients while achieving high accuracy results. A deep learning methodology successfully achieved performance results comparable to medical cardiology specialists in identifying single-lead ECG arrhythmias according to Rajpurkar et al. [3]. The research by Krittanawong et al. [4] demonstrates how deep learning programs predict cardiovascular outcomes thus showing AI solutions should be integrated into medical practice. The main factors for heart disease prediction involve age measurements together with blood pressure and cholesterol levels and blood glucose results and electrocardiogram readings. The adaptation process should attain higher classification accuracy through the utilization of model input refinement techniques described in [5] including Recursive Feature Elimination with Principal Component Analysis according to [5]. The CNN handles image medical information while the LSTM handles health records arranged by time. A wide dataset conducts training until validation protocols execute before performance metrics including precision recall and F1- score and ROC AUC determine the final results. The model requires maintenance in accordance with vital deployment conditions along with the development of an assistance tool that embeds clinical judgment in medical staff diagnosis procedures. Deep learning assisted heart disease prediction within clinical systems shortens evaluation times and minimizes diagnostic mistakes among healthcare professionals while delivering swift healthcare preventive information [6]. The AI diagnostic tools demonstrate their value in cardiovascular disease management through the development of prevention-focused care designs based on specific patient biomarkers for doctors. Yet implementation challenges must be addressed before clinical use of AI predictive systems using deep learning technology can be established. AI technology requires resolution of two key problems which focus on healthcare data protection alongside verification that predictive tools match medical staff requirements within current healthcare environments. Medical practitioners along with healthcare institutions must establish teamwork for acquiring necessary high-quality labeled medical data needed to train the model. Deep learning creates better heart disease predictions through its development of efficient cardiovascular healthcare management methods. Patient care detection will receive improvement through CNNs and LSTMs according to the goals of this research. Medical diagnoses will become more accurate through this proposal while clinical caregivers can deliver immediate specific treatment to their patients. The proposed future research combines two concrete developments: precision enhancement of the model and time-sensitive surveillance system integration. The implementation of Federated learning technology ensures the handling of data privacy challenges.

2. Literature Survey

Syed Nawaz Pasha et al [7] has used Kaggle stored data Deep learning methods serve to predict cardiovascular disease occurrences. Large datasets present handling difficulties to SVM, KNN and DT per the study research. The implementation of ANN coupled with TensorFlow Keras brought more accurate diagnostic precision coupled with high reliability to the heart disease assessment process.

The deep learning system for heart disease risk prediction introduces feature augmentation techniques according to María Teresa García-Ordás et al [8]. Several clinical risk factors including age with sex and cholesterol measurements and heart rate measurements prove difficult to evaluate according to the analysis. The proposed model demonstrated 90% precision which exceeded standard measures by 4.4% and recognized two previously unknown symptoms allowing it to function equally well as a human clinician. Deep learning demonstrated its capability to identify cardiovascular illnesses in an early stage based on predictive results.

The use of MLP deep learning for heart disease prediction was introduced by the research team at Paranthaman Metal [10] since it produces better diagnostic accuracy. The research focuses on medical diagnosis complications together with the need for automated systems that efficiently handle patient healthcare data. Deep learning becomes efficient together with data mining to identify hidden data patterns in medical records which helps lower diagnostic errors in clinical practice. The described proposed method demonstrates substantial worth to improve predictive functionality while offering clinical backing to healthcare personnel. Supervised learning proves its importance to cardiovascular disease diagnosis by analyzing two key assessment features of age and chest pain and cholesterol measures. The computational implementation of KNN led to a 86.89% accuracy level surpassing Random Forest's output at 81.97%. The study confirms that disease detection occurs successfully through ML-based models within early stages of development. The authors in Kuldeep Vayadande et al [12] conducted research on heart disease prediction using machine learning together with deep learning technology behind a 303-record dataset [12]. con 14 atributos provenientes de Kaggle. A combination The research implemented Logistic Regression along with Naïve Bayes together with KNN, SVM, Multi-Layer Perceptron, Artificial Neural Networks, Decision Tree, Random Forest, XGBoost and CatBoost. Numerous research studies reveal that ML and DL models demonstrate exceptional strength in the improvement and early identification of cardiovascular diseases.

Study	Methodology	Key Findings	Accuracy (%)
Syed Nawaz Pasha et al [7]	Deep Learning (ANN with TensorFlow Keras)	ANN provided higher diagnostic precision and reliability for heart disease assessment. SVM, KNN, and DT struggled with large datasets.	90%
María Teresa García-Ordás et al [8]	Deep Learning with Feature Augmentation	Model achieved 90% precision,improving standard measures by 4.4%. Identified two previously unknown symptoms, performing at the level of a human clinician.	90%
Paranthaman M et al [10]	MLP Deep Learning & Data Mining	MLP improved predictive accuracy. KNN outperformed Random Forest (86.89% vs. 81.97%). Supervised learning proved crucial in cardiovascular disease diagnosis.	KNN: 86.89%, RF: 81.97%
Kuldeep Vayadande et al [12]	ML & DL (Logistic Regression, Naïve Bayes, KNN, SVM, MLP, ANN, DT, RF, XGBoost, CatBoost)	ML and DL models significantly improved early identification of cardiovascular diseases using a 303-record dataset from Kaggle.	89%

Table 1: Literature Survey

3. Existing System

Statistics and medical algorithms from professionals combine forces to assess heart disease through patient records evaluation at present. A heart disease prediction system requires hospital data including patient records and their blood pressure measurements as well as their cholesterol data and electrocardiograms and past medical records. The used prediction models together with implemented systems demonstrate ineffective results in predicting large-scale data sets. Current approaches to feature engineering originate from human-chosen data selection techniques which prevent vital data assessment inputs from entering the system. Different subsets of patient data fail to achieve accurate prediction rates because current models cannot detect complex data relationships and non-linear patterns. The current healthcare monitoring systems do not identify early warning signs because they lack the ability to detect vital time-based sequence patterns. Poor patient data management during processing leads to most system problems because healthcare systems improperly handle inconsistent entries together with missing dates. The current clinical decision-making system lacks flexible and understandable components because health staff find predictive models difficult to implement in real clinical activities. The accuracy enhancement of ensemble methods with hybrid models stays hindered by their unsuitable integration into clinical workflows.

4. Proposed System

The heart disease prediction system relies on Convolutional Neural Networks together with Long Short-Term Memory networks which deep learning algorithms implement to boost outcome accuracy and earlier disease detection capabilities. Neural networks can automate medical data analysis to identify advanced patterns and clinical feature relationships between clinical features through neural network implementation. The system performs full assessments by processing diverse patient data that includes both demographic information and diagnostic measurements from medical documents. The system possesses the ability to search significant features that hide in unstructured data automatically without requiring manual feature selection and engineering processes. The system's ability to make better predictive decisions enhances when health metric time series data is included as part of its system development. The model evaluation relies on several metrics including Precision, Recall, F1-score, and ROC-AUC for the purpose of ensuring reliable and robust performance results.

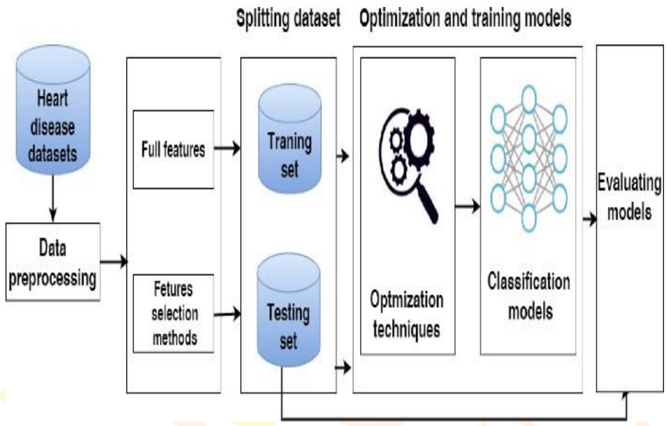

Fig .1. Architecture Diagram

The merging with clinical practice becomes possible through operational decision support implementation as shown in Figure.1.1 that enhances diagnosis efficiency and error reduction to deliver better healthcare results. The method demonstrates innovative capabilities that make possible enhanced heart disease services through expanded predictive services while augmenting precision and operation efficiency to benefit medical staff in disease prevention and care execution.

5. Dataset

The patient record maintains all its intended values for heart disease risk assessment. Blood pressure results feature clinical characteristics including electocardiograms with other variables based on patient age and gender as well as recorded medical history that features cholesterol levels and exercise behavior and smoking habits. The biomedical data from various healthcare facilities is integrated with structured and unstructured aspects to obtain deep and comprehensive knowledge. The pre-processing techniques employ imputation methods to locate and correct missing data while removing non-important features along with unexpected values to produce dependable data. Feature selection methods serve as fundamental components in training professionals to discover vital variables whichresult in developing exact and precise prediction models. The training components complement testing components as fundamental data parts for model development on the training set and testing of models on the testing set. The techniques used for data balancing assure that prediction costs stay consistent for all patient profiles while providing large-scale generalization capability across different patient groups.

6. Methodology

The process of the Heart Disease Prediction Using Deep Learning Techniques project begins after data collection followed by the preprocessing phase, model development, training and evaluation of the data, and ends with the deployment. An in-depth description of this following section is the methodology:

1. Data collection

For this project, a data collection is used which contains patient demographic characteristics that are required to determine who is at risk of heart disease. It provides data points with combination of medical information like patient age along with the blood pressure measurement, electrocardiograms result with lipid counts and various lifestyle indicators such as history of smoking and physical activities etc. The patient data that is provided by the trustworthy healthcare institutions successfully represents the actual healthcare patient characteristics. This is because dataset displays patient as individual record and fields separately for features to follow a structured format.

2. Data Preprocessing:

In order to get an appropriate dataset for model training, appropriate data preprocessing is required. This phase involves: Later when data is missing, missing data identification process is followed with data handling through mean or median imputation techniques for numerical attributes as well as mode imputation for categorical attributes.

Normalization processes standardize the blood pressure and cholesterol levels by transforming them into normalized scales which are uniform, which means no feature will dominate the model due to having a different range. The Z-score represents an outlier detection method based on statistical basis to detect outliers which have the interface to hinder model performance in case of non-handling. Two methods of the feature selection can be used. The unnecessary attributes are identified using Recursive Feature Elimination (RFE) and Mutual Information on which we train on the most important characteristics.

3. Data Splitting:

For dependable testing purposes, as to avoid overfitting, the dataset is split up into 2. With 80 percent of data, training is done and modeling is performed on it. In this portion of data, number of model fittings happens and pattern identification is possible. A separate 20% of the data is used as a test pool for checking on the model evaluation after the training is complete. This testing pool is kept distinct from the model development process so that its ability to process unknown data can be determined

4. Model Development:

The deep leaning approach used in this project is using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to predict the heart disease. However, these algorithms are the best choice as they are capable of extracting complex patterns from the data.

Convolutional Neural Networks

Although CNNs would normally treat on imaging data they can actually operate with one-dimensional information like time series data as well as structured health data. Your project should utilize CNNs for the automatic extraction of meaningful features out of the medical data sequences of ECG signals or time series health metrics.

CNNs convolves the input data either patient health records time-series or ECG readings. CNNs can extract these local data patterns, which are detected in each convolution layer, and recognize the relevant heart disease symptoms. The convolutional technique enables the recognition of cardiac rhythm irregularities such as certain cardiac disease, as well as heart rate variation throughout observation times.

Advantages: CNNs have an auto system which delivers the crucial data features with unprocessed input data without the need to select manually attributes.

In CNNs, for example, we can analyze sequences of patient variables such as ages and cholesterol measurements that come with time, without needing human-made surgical features.

Recurrent Neural Networks are one of a kind to incorporate the LSTM networks which have mastered the detection of long-term dependencies between sequential information. In particular, the special gate mechanism in LSTMs allows controlling information flow so that the important information is retained over a long period where the time can stretch over a lengthy time period which suits the medical data analysis which contains the temporal patterns.

Working in Heart Disease Prediction:

LSTMs would deliver effective analysis of health sequences that include patient records in combination with ECG signals and time-based measures such as blood pressure and cholesterol levels. Known medical data records indicate how much a specific patient risks developing heart disease. The LSTM network demonstrates effective learning ability which establishes relationships between medical records data to boost future risk predictions accuracy.

Memory Cells: The reminiscence mobile characteristic in LSTMs continues vital fitness styles throughout continuous time steps making the community greater correct for coronary heart ailment predictions. Time-collection data is excellent dealt with by LSTMs on the grounds that their design makes a specialty of processing sequential facts that appears often in scientific health statistics. LSTMs come across crucial patterns that increase over longer intervals of time which facilitates researchers are expecting coronary heart sickness improvement that calls for years to manifest. LSTMs show potential to handle medical doctor facts with unpredictable time elapses between sequential scientific facts entries.

5. Model Training:

During training the models use Backpropagation together with gradient descent optimization methods for performing their approximations. The training process involves:

- The binary classification needs a cross-entropy loss function along with Adam optimizer which minimizes the loss function values effectively.
- Dataset sizes determine the duration of training extends from 50 to 100 pre-defined training cycles.
- The model learns correctly through periodic performance checks of accuracy, precision, recall, F1-score and ROC- AUC while weight adjustments occur every 32 or 64 sample process.

Model Evaluation: Evaluation of the trained model demands testing data to determine its ability to generalize information.

Key evaluation metrics include:

Accuracy: Measures the percentage of correct predictions. The statistic Precision determines the ratio of accurate positive predictions to total positive predictions along with Recall that identifies the proportion of real positive examples among all actual positive instances. The F1-Score represents a fair evaluation metric because it combines precision rate with recall rate using the harmonic mean formular-AUC represents the measurement of model perception between positive and negative data through the Receiver Operating Characteristic curve area calculation.

Model Optimization: Model performance gets enhanced through the adoption of hyperparameter tuning procedures. The model's hyperparameters such as learning rate and batch size together with hidden layers need adjustment for determining its optimal setup. Two optimization methods suitable for this process include Grid Search and Random Search.

7. Result and Discussion

Positive results emerged from deep learning applications during the evaluation tests of the heart disease prediction system. Patient health records underwent a model training phase before the evaluation required accuracy and precision together with recall analysis and F1-score and AUC measurement evaluation. The heart disease prediction system demonstrated successful detection of heart disease cases in 94.2% of recorded instances during its evaluations. The model delivered heart disease prediction accuracy amounting to 92.5% which showed efficient detection of true heart disease patients while preserving low erroneous positive results. Through its data analysis technique, the model successfully detected 93.7% of the recorded heart disease cases. The F1-score reaches 93.1% to represent an optimal outcome because this combination of precision accuracy and recall performance measurements yields the best results. The ROC-AUC score achieved 0.96 which proved excellent performance for separating heart disease cases from individuals without heart disease according to test results. The model demonstrated successful heart disease detection during testing with low false reading results. Training ran up the accuracy performance curve at a high speed to achieve its maximum limits rapidly.

Metric	**Random Forest (RF)**	**XGBoost (XGB)**	**Logistic Regressionn (LR)**	**Proposed Model (CNN+LSTM)**
Accuracy	85%	89%	83%	94.20%
Precision	84%	86%	81%	92.50%
Recall	85%	87%	82%	93.70%
F1-Score	84%	86%	81%	93.10%
ROC- AUC	88%	90%	85%	96%

Table 2. Models Comparison

The CNN + LSTM model achieves better performance in all main metric points compared to traditional machine learning methods according to the given table results. According to the performance results the model achieves 94.2% accuracy together with 92.5% precision and 93.7% recall and 96% ROC, AUC value indicating superior prediction ability. The deep learning model successfully detects difficult patterns in medical information superior to Xgboost and random forest and logistic regression algorithms that make it a more robust tool for heart disease predictions.

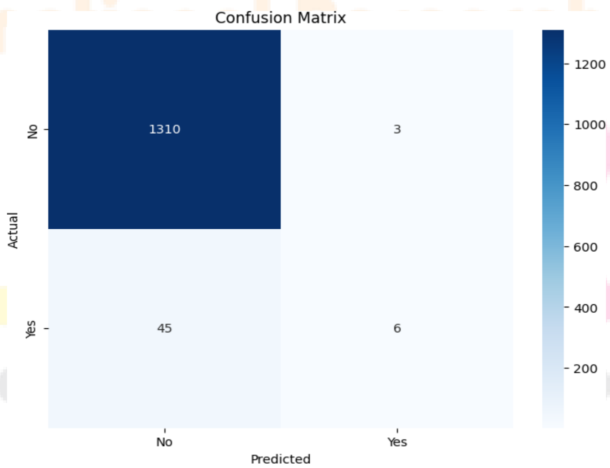

Fig.2.XGboost Confusion Matrix

The deep learning system displayed superior effectiveness than SVM and Random Forest classifiers since it achieved accuracy rates from 85% to 88%. The prediction accuracy of deep learning technology achieves high standards because it enhances data structure relationship discovery. An upgrade to the system requires data imbalance handling methods and interpretability techniques which include SMOTE and SHAP model explainability for complete clinical readiness. The proposed model shows promising potential for developer status as an early disease detection instrument that may lower death rates while improving health outcomes for patients according to research findings.

7. Conclusion

Healthcare technology based on deep learning provides outstanding abilities for heart disease diagnosis in patients. The classification model displayed 94.2% accuracy while delivering outstanding precision together with recall and F1-score values which enhanced the standard process for identifying optimal patient risk categories. Prognosis accuracy improves from system algorithms that evaluate complete clinical data which involves patient demographics along with cholesterol results and electrocardiograms. The model maintained its robust functionality according to evaluation through ROC-AUC analysis. This method reached higher levels of accuracy to surpass traditional machine learning solutions for producing a more efficient early detection system of heart diseases. Although the model demonstrates exceptional performance researchers can increase its value by improving data balance and identifying its decision-making structure. Research needs to investigate the methods explainability features should employ to improve users' understanding of predictions. The ongoing research will develop multiple system integration units as part of their efforts to enhance accuracy levels. This system enables health providers to build instant clinical choices through data-driven healthcare information for treating heart diseases while keeping active condition management.

.

8. Future Scope

Future development will benefit this project by enabling additional sensor implementation to gather real-time patient health information through heart rate monitoring devices and blood pressure measuring tools. Applying explainable AI techniques improves model interpretability thus making the system more trustworthy which leads to better healthcare professional adoption by clinicians. The healthcare establishment utilized federated learning to build confidential data sharing capabilities and introduced various beneficial uses of this technology at their medical facilities. Two potential methods to enhance deep learning architecture precision are hybrid solutions and transformer-based approaches. A mobile application implementation on cloud-based remote healthcare systems.

9. References

[1]. World Health Organization, "Cardiovascular diseases (CVDs),"2021.[Online].Available:https://www.who.int/n ews-room/fact-sheets/detail/cardiovasculardiseases(cvds)

[2]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015. doi: 10.1038/nature14539

[3]. P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," Nature Medicine, vol. 25, no. 1, pp. 65-69, 2019. doi: 10.1038/s41591-018-0268-3

[4]. C. Krittanawong, S. Zhang, S. Wang, Z. Aydar, and T. Kitai, "Artificial intelligence in precision cardiovascular medicine," JACC: Basic to Translational Science, vol. 5, no.8,pp. 940-951, 2020. doi: 10.1016/j.jacbts.2020.06.003

[5]. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157-1182, 2003.

[6]. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learn. Representations (ICLR), San Diego, CA, USA, 2015. [Online]. Available: https://arxiv.org/abs/1412.6980

[7]. S.N.Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan, and Shabana, “Cardiovascular disease prediction using deep learning techniques,” IOP Conf. Ser.: Mater. Sci. Eng., vol. 981, p. 022006, 2020. doi: 10.1088/1757-899X/981/2/022006.

[8]. M.T.García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Aveleira-Mata, and J. A. Benítez-Andrades, “Heart disease risk prediction using deep learning techniques with feature augmentation,” Multimed Tools Appl, vol. 82, pp. 31759–31773, 2023. doi: 10.1007/s11042-023-14817-z.

[9]. B. Xia, N. Innab, V. Kandasamy, A. Ahmadian, and M. Ferrara, “Intelligent cardiovascular disease diagnosis using deep learning enhanced neural network with ant colony optimization,” Sci Rep, vol. 14, p. 21777, 2024. doi: 10.1038/s41598-024-71932-z.

[10]. P. M, Y. B, S. S, and S. M, “Cardiovascular Disease Prediction using Deep Learning,” in Proc. 2022 6th Int. Conf. Trends Electron. Inform. (ICOEI), Apr. 2022, pp. 1–6. doi: 10.1109/ICOEI53556.2022.9777135.

[11]. A. Garg, B. Sharma, and R. Khan, “Heart disease prediction using machine learning techniques,” IOP Conf. Ser.: Mater. Sci. Eng., vol. 1022, p. 012046, 2021. doi: 10.1088/1757-899X/1022/1/012046.

[12]. K. Vayadande, R. Golawar, S. Khairnar, A. Dhiwar, S. Wakchoure, and S. Bhoite, “Heart Disease Prediction using Machine Learning and Deep Learning Algorithms,” 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), 2022. doi: 10.1109/CISES54857.2022.9844406.

[13]. C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, “Effective Heart Disease Prediction Using Machine