ISSN: 2584-0495 Vol. 2, Issue 10, pp. 1286-1301

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

A Narrative Analysis Report on Heart Disease Prediction using Advanced Deep Learning Techniques in Smart Healthcare System

U.A.Ranjni and Mercy Paul Selvan

Cite as: Ranjni, U. A., & Selvan, M. P. (2024). A Narrative Analysis Report on Heart Disease Prediction using Advanced Deep Learning Techniques in Smart Healthcare System. International Journal of Microsystems and IoT, 2(10), 1286-1301. https://doi.org/10.5281/zenodo.14245220

Published online: 30 October 2024

DOI: https://doi.org/10.5281/zenodo.14245220

International Journal of Microsystems and IoT
Vol. 2, Issue 10, pp.1286-1301; DOI: https://doi.org/10.5281/zenodo.14245220

A Narrative Analysis Report on Heart Disease Prediction using Advanced Deep Learning Techniques in Smart Healthcare System

U.A.Ranjni and Mercy Paul Selvan
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, India.

KEYWORDS

Artificial intelligence; Data Mining; Deep Learning; Heart Disease Prediction; Healthcare Sector.

ABSTRACT

In recent years, the advancement of artificial intelligence (AI) and the gradual initiation of AI exploration in the medical industry have allowed people to recognize the promising potential of combining AI with healthcare. Data mining methods are being utilized efficiently in illness detection, which benefits health professionals. A vast amount of data is gathered from the health industry, and classification methods are used to identify new patterns. Heart disorders were chosen for diagnosis and categorization in this article. In this paper, an exhaustive analysis of certain common data mining algorithms is undertaken utilizing many datasets. The hot deep learning discipline, for example, has demonstrated increased potential in applications like as illness prediction and treatment response forecast. The study results will aid in understanding the primary data mining approaches and selecting the appropriate category of algorithms for heart disease analysis. This article provides several fundamental deep learning frameworks and common disorders, as well as a summary of deep learning prediction approaches for cardiac diseases. Point out several flaws in present illness prediction and provide a prognosis for future progress. It seeks to clarify the usefulness of deep learning in illness prophecy, as well as to highlight the high correlation among profession in terms of future growth. Deep learning approaches' unique feature extraction methods can still production an essential part in future medical inquiry.

1. INTRODUCTION

When it comes to the collection and processing of data, the healthcare industry is among the most concerning. Countless pieces of multidimensional patient data, such as clinical variables, hospital resources, disease diagnostic information, patient records, and medical equipment, are being generated by the digital age and technological breakthroughs [1]. In order to make informed decisions, it is necessary to analyze and analyze the massive, dense, and complicated data. One of the many promising applications of medical data mining is the discovery of previously unknown patterns within medical datasets [2]. Data mining and machine learning have transformed healthcare organizations by revealing hidden patterns in massive databases, as well as correlations and interactions between several variables [3]. With its ability to provide and compare current data, it plays a crucial role in the medical industry, guiding future actions. This technology allows for the examination of huge volumes of data by combining numerous analytic approaches with contemporary and complicated algorithms [4]. Its primary function in healthcare is the systematic collection, organization, and analysis of patient data. In addition to providing a foundation for a more thorough understanding of the mechanisms at work in nearly every facet of the medical domain, it has the potential to reveal innate inefficiencies and best practices for delivering better services, which could improve diagnosis, medicine, and treatment outcomes [5]. In general, it searches medical databases for relevant information, which helps in the early diagnosis and inhibition of disease outbreaks.

The phrase "medical diagnosis" describes the procedure by which a person's symptoms and other signs are used to establish a diagnosis. A diagnostic procedure, or series of procedures, including diagnostic tests, are carried out throughout the diagnostic process. Due to the complexity of the symptoms involved, the diagnosis of chronic diseases is a major concern for healthcare providers. Because of the complexity of the process, it often results in erroneous assumptions. The symptoms experienced by the patient and the doctors' level of training and experience form the basis of clinical judgement when making medical diagnoses [6]. It becomes increasingly challenging for doctors and physicians to keep up with the latest advancements in clinical practice as medical systems change and new therapies become accessible [7]. Data mining is crucial in the healthcare sector for the prognosis of this illness. Data for making decisions by sifting through large datasets of historical records. The data could be hidden and unintelligible until data mining is employed. Data mining techniques like categorization allow us to draw conclusions about the future from the past [8]. In order to better anticipate a patient's future health, medical data mining has developed a workable solution to integrate classification methods and provide computerized training on the dataset. This, in turn, enables further inspection of the hidden patterns in medical data sets. Thus, data mining can offer clinical assistance through analysis and is likely to give awareness on a patient's past. These patterns are crucial for clinical examinations of patients. To rephrase, data mining methods are an integral aspect of pre-event detection of heart illness [9]. Predictions on the likelihood that a person may suffer from heart disease will be used to train and evaluate algorithms.

Forecasts about the future may be made by doctors with the use of models are used. Patients' future outcomes are forecast by prediction models using historical data. It will be easier for doctors to aid patients if they can predict who may be at risk of cardiovascular illness in the future [10]. If not eliminated entirely, this can drastically cut down on patient fatalities. Assuming the models are refined, academics will find this large data prediction analysis to be very useful, and it may even find practical application. The goal of applying machine learning techniques to health datasets is to improve disease detection, prognosis, prevention, and treatment by extracting relevant characteristics. This survey discusses the pros and cons of utilizing deep learning for heart disease prediction. This work also includes the analysis for the future.

The remainder of the document is prearranged as: Section 2 gives some history of heart illness; Section 3 talks about the datasets that are now accessible; Section 4 addresses some technical background; and Section 5 talks about related studies. Section 6 presents data mining's healthcare applications, Section 7 discusses the challenges with current methodologies, Section 8 looks at potential future research, Section 9 draws conclusions, and Section 10 provides references.

Table 1 brings out the explanation of diverse kinds of heart diseases or cardiovascular disease (CVD) and the symptoms categorizing each of these.

2. Background of Heart Disease Detection

Types of CVD	**Description**	**Symptoms**	**Risk Factor**
Coronary Heart Disease	Ischemic heart disease (IHD); most common type.	Heart attack, Angina at chronic condition.	Irregular eating habits, high blood pressure, diabetes, and lack of physical exercise.
Stroke	Typical types of cardiovascular disease and their classifications are as follows: ischemic stroke, transient ischemic attack, and hemorrhagic stroke.	There is injury to the brain, which results in weakness, often on one side of the body.	High blood sugar, bad eating habits, lack of exercise, diabetes, and just getting older.
Congenital heart disease	Malformations of the heart or central blood artery that occur either before birth or while the mother is pregnant.	Breathlessness or inability to attain appropriate growth and development. .	Maternal alcohol and medicines use, maternal impurity, poor maternal nutrition,
Other diseases	cardiovascular	Diseases of the heart valves, heart muscle disorders, vascular tumors of the brain, and heart tumors are all examples of heart-related conditions	-

There are two main categories of cardiovascular illness that pose a concern. The following are listed:

1) Controllable factors.
2) Uncontrollable factors.

People may take action to reduce their risk of cardiovascular disease by controlling modifiable risk factors such as smoking, alcohol use, weight, blood pressure, and cholesterol. People cannot change their sex, age, or family medical history to reduce the risk of cardiovascular disease. In this world, people might suffer from a wide variety of cardiac conditions.

2.1 Congenital heart disease

One form of heart illness, known as congenital heart disease, has been present in humans from the moment of their birth. Examples of congenital cardiac defects include:

1) **Septal Defects:** The septal defect is characterized by a hole that exists between the left and right chambers of the human heart.
2) **Obstruction Defects:** A partial or total blockage of the blood flow amid the heart chambers characterizes the obstruction defects.
3) **Cyanotic Heart Disease:** Cyanosis of the heart occurs when there is insufficient oxygen in the blood.

2.2 Arrhythmia

When the normal heartbeat of a human being is disrupted, arrhythmia occurs. Arrhythmia occurs when human heart is not in sync with the heartbeat rate. Regardless of external factors, the heart's electrical impulses maintain a constant rhythm. Variations in heart rate are so common that most people experience them.

2.3 Coronary Artery Disease

It is the primary role of arteries to provide the heart with oxygen-rich blood and nutrients. Cholesterol can cause damage or illness to the coronary arteries. Reduced oxygen and nutrition delivery to the body is a direct result of cholesterol in the coronary arteries.

2.4 Heart failure

When the heart is powerless to pump blood efficiently through the body, a disorder known as heart failure or congestive heart failure sets place.

2.5 Heart Muscle Disease (Cardiomyopathy)

Cardiomyopathy is another name for heart muscle illness. When the heart's walls thicken or grow, it can lead to heart muscle disease. The main goal of this illness is to lessen blood flow to the entire body, which ultimately leads to heart failure.

2.6 Heart Valve Disease

Hearts in humans have four valves. The four valves of the heart are accountable for regulating the blood's forward flow and distributing it throughout the body. Which of these symptoms a person experiences most often is a strong indicator of whether or not they have heart disease. Typical individuals often miss some of the signs. However, heart palpitations, shortness of breath, and chest discomfort are among the most typical symptoms.

Chest discomfort, often known as angina or angina pectoris, is a symptom of many different kinds of heart disease that occurs when the heart does not receive enough oxygen. Stress or physical activity can trigger angina attacks, which typically subside after 10 minutes. As a result of many forms of cardiac disease, heart attacks can also occur. Similar to angina, the symptoms of a heart attack can worsen when at rest and are indicative of a more serious condition. Indigestion, heartburn, stomachaches, and a heavy sensation in the chest are all signs of cardiac illness, which can often mimic these conditions. Chest discomfort that travels to other parts of the body, such as the arms, neck, jaw, vertigo, nausea, and vomiting are further signs of heart disease [11].

Description of Various Heart Disease Dataset

Heart Disease Dataset (Comprehensive)

This study's data set was compiled from five well-known databases on cardiovascular illness, which include medical and non-medical aspects [12]. This exhaustive dataset contains information from 1,190 individual patient records. Eleven characteristics are slope. You can set it to "normal" or "heart disease" as its target variable. There are both numerical and category elements in the dataset. In Table 2, you can find full explanations of the features.

Table 2 Feature explanation of inclusive heart disease data set.

Variable name	**Values**	**Data category**
Age	Real	Numeric
Sex	1,0	Binary
Resting blood pressure	Real	Numeric
Serum cholesterol	Real	Numeric
Fasting blood sugar	1,0	Binary
Resting electrocardiogram result	0,1,2	Nominal
Concentrated heart level achieved	Real	Numeric
Exercise induced angina	0,1	Binary
Old peak	Real	Numeric

The slope of the peak application ST segment	0,1,2	Nominal
Target	0,1	Binary

3.Open-Access Datasets and Data Source

To ensure that DL and ML learning models can accurately predict CVD, datasets must be trained and verified. Two important data sets were identified in the papers examined for this study: (1) open-access data sets and (2) unpublished critical data sets [13]. Those publications make use of open access data sets from the following projects: Apta CDSS-E, GitHub PCG, Pascal Challenge 2016, and PhysioNet/CinC Challenge 2015. There are 25 research and 5 databases that make up the open-access group. However, several datasets, such as CVD data set obtained from the Kaggle repository, were utilized in more than one study. There were 3,240 individuals with cardiovascular disease in the sample.

Most studies use the sound of CVD to identify if a patient has CVD or not, according to the dataset webpage. Sixteen of the thirty-seven articles cited here examined CVD prediction using the dataset. The Pascal Challenge CVD dataset was the second most used data set for CVD prediction research in the past. There are two categories of audio files in the dataset: normal and aberrant. In doing so, it classifies patients as either healthy or unhealthy. The dataset may be found at the Kaggle Depository and was used in three publications.

.

4. Dataset of HD-2019.

A number of researchers have been using the first "heart disease dataset 2019," which has been utilized for the diagnosis of heart [14]. You may also find the "heart disease dataset 2019" in the open-source Kaggle repository. There are 1025 samples in the heart disease dataset, along with 13 characteristics and a few missing values. Two categories, "normal" and "heart patient," make up the desired output label. These findings

are based on the second "cardiovascular disease dataset 2019" that was analyzed. You may also get the 2019 cardiovascular illness dataset in the Kaggle repository. There are 70,000 patient samples in the 2019 cardiovascular disease dataset, which contains 11 distinct characteristics and a few missing values. The specifics of these two databases are laid forth in Tables 3 besides 4.

Table 3: Explanation of dataset 2019 with feature info

Sr #	Comment	Feature name	Feature code	Description	Category	Value range (min–max)
1	Patient's age in years	Age	Age	Age in years	Numeric	29<age>77
2	1: Male, 0: Female	Sex	Sex	1=Male 0=Female	Nominal	1 0
3	1: Typical angina, 2: Atypical angina,	Chest pain category	Cp	1=strange angina 2=classic angina 3=asymptomatic 4=nonanginal pain	Nominal	1 2 3 4
4	3: Non-anginal pain, 4: Asymptomatic	Resting blood pressure	Trestbps	In mm Hg on admission to the hospital	Numeric	94–200
5	Systolic pressure: Measure of the blood pressure	Serum cholesterol	Chol	In mg/dl	Numeric	126–564
6	Amount of cholesterol in patient's blood (Mg/Dl)	Fasting blood sugar	Fbs	Go on hunger strike blood 120 mg/dl) (1 D true; 0 D false)	Insignificant	1 0
7	1 (True) : ≤>120 mg/dl,	Resting Electrocardiographic results	Rest ECG	0 D normal 1 D having ST-T 2 D hypertrophy	Insignificant	0 1 2
8	0 (False):	Maximum heart rate completed	Thalach	Not comment	Numeric	77–202
9	0: Normal,	Exercise-induced angina	Exang	1=yes 0=no	Insignificant	1 0
10	1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or	Old peak=ST depression induced by exercise relative to rest	old peak	Not mention	Numeric	0–6.2

	depression of >					
11	0.05 mV)	Slope of the peak exercise ST segment	slope	1= up sloping 2 = at 3 = down sloping	Insignificant	1 2 3
12	2: Showing probable or definite left ventricular hypertrophy by Estes' criteria	Number of major vessels (0–3) colored by uroscopy	Ca	Not mention	Insignificant	1 2 3
13	Maximum number of times that heart beats per minute (bpm)	Thallium scan	Thal	3=normal 6=xed defect 7=reversible shortcoming	Insignificant	3 6 7

Table 4: Explanation of dataset 2019 with feature info.

Type	**Feature name**	**Feature code**	**Description**	**Value range (min–max)**
Numeric	Age	Age	In days	10798<day>23713
Nominal	Gender	Gender	1=women 2=men	1 2
	Height	Height	In cm	55–250
Numeric	Weight	Weight	In Kg	10–200
Numeric	Systolic blood pressure	ap_hi	Not mention	−150–16020
Numeric	Diastolic blood pressure	ap_lo	Not mention	−70–11000
Numeric	Cholesterol	Cholesterol	1=customary 2=above normal 3=well overhead normal	1 2 3
Nominal	Glucose	Gluc	1=normal 2=above normal 3=well above normal	1 2 3
Nominal	Smoke	Smoke	Whether persevering 1=smoke 0=not smoke	1 0
Nominal	Alcohol intake	Alco	whether tolerant take 1=take alcohol 0=not take alcohol	1 0

5. Software/Tools used for the CVD Forecast

It is usual practice to use forecast CVD. The algorithms for the CVD forecasts were developed using a wide variety of tools and programming techniques. The analyses in the examined articles were conducted using a variety of software programs. We grouped these programs and resources into two categories: data mining and programming. Python and related platforms, such as the R programming environment and Jupyter Notebook, are mentioned in the articles as a language for deep and machine learning data investigation. In these research, the data-mining apps most often employed to predict CVD were WEKA, MATLAB, and Minitab.

5.1 Data preprocessing and descriptive statistics

The data collection was pre-processed before any model was applied. Data that was inconsistent or duplicated was the main emphasis. The data collection was cleansed of all those inconsistent and duplicate records during the subsequent data preprocessing phase. The features were displayed independently due to the fact that the dataset contains both numerical and category variables. Categorical features were found for every subcategory, in contrast to numerical features that were shown using minimum and maximum values, median, mean, and standard deviation.

5.2 Technical Background of Deep Learning

In order to treat patients, doctors rely on standard data. After a specific number of cases have been reviewed, examines are summarized where standard information is inadequate. Though examples can be detected earlier with the help of AI, this interaction does take time. A massive quantity of data is needed to make use of AI. Depending on the infection, there is a very limited quantity of data available. There is a large disparity between the number of tests showing an infection and the number of tests showing no disease at all. Consequently, deep learning models for the documentation of heart disease were the exclusive focus of the investigation.

5.3 CNN

An upsurge in the memory cost occurs when the number of parameters in a neural network with only one fully connected layer increases as it goes deeper into the network. A remedy has been progressively investigated to address the issues brought about by some completely linked layers and to encourage the advancement of neural networks to a higher degree. A (CNN) decreases the number of parameters and increases training efficiency by applying the notion of local correlation and weight sharing. Originally proposed in 1998, LeNet laid the groundwork for (CNNs). Included in the fundamental architecture are the following layers: convolutional, pooling, and fully connected. Lastly, obtain output after fc, and make architectural adjustments after convolution and pooling.

However, it has not been developed to a higher level due to the effect of processing power and data volume at that time. After a few objective issues were overcome in 2012, CNN swiftly became the preferred approach in computer vision, thanks to its great outcomes in international competitions. Following that, CNN variant models and better CNN models like VGG net, ResNet, DenseNet, Xception, etc., began to appear. The essential details of several frameworks are shown in Table 5. While (CNNs) were initially developed for computer vision, they have since been shown to be highly successful in a variety of other domains, with medicine being one of the most notable examples. Classification of images, illness prediction, visualization of medical images, etc. Although CNN is capable of processing a wide variety of data, its most notable use is the processing of images including spatial local correlation data, such as CT scans, X-rays of the chest from medical fields, and other similar images. Using technologies of complicated data, the CNN model can handle medical pictures, whether it's a single image for illness diagnosis or a composite of numerous photos for supplementary prediction. Then, when compared to conventional machine learning models, its performance in illness categorization is superior thanks to the completely linked layer.

Table 5: Commonly used CNN perfect and construction

General structure	**Model name**	**Proposal time**
First, stack the greater $3 * 3$ kernel. Then, use the biggest pooling layer, then link the fully connected layer, and lastly use the Softmax function to classify.	VGG (16, 19)	2014
There is a skip connection made between the input and output, and the residual module is used. Even though there are more network levels, the problems caused by gradient dispersion are not made worse.	ResNet	2015
A new idea called "depth separable convolution operation" is used. It can make sure that a lot of data can be handled more quickly while keeping the same number of strictures.	Xception	2016
As an improvement based on ResNet, the main difference is that it makes a dense link between all the front and back layers once the connection is made.	DenseNet	2017

5.4 RNN

Models need to learn sequence characteristics for time series forecasting, task-based discourse, and many other applications. There is an issue with ineffective time series management since neurons in the fully linked layer and every other layer of the neural network cannot be connected to each other. Convolutional neural networks (CNNs) excel at spatial problem-solving but struggle with time-related data. A few of the issues listed above inspired the development of recurrent neural networks (RNN).

Neurons and feedback loops make up an RNN. A situation in which RNN excels is one in which there is a dependence connection between the inputs. In the face of delays and unusual noise in massive data, RNN outperforms the majority of machine learning methods. Improved models LSTM and GRU evolved from RNN, although RNN still has short-term memory issues and can experience gradient explosion and disappearance during training. In contrast to the simple RNN, LSTM incorporates a gate mechanism and a state vector. The three primary building blocks of LSTM are Gate are the sole components of the GRU. The gating mechanism allows one to manage the refreshing and forgetting of information. The input data is represented by xt in this figure, the state vector is denoted by ht, and the change of the state vector following gate control is shown by ht−1 to ht. The activation function is represented by σ, tanh. Different from typical RNN, LSTM demonstrates outstanding performance for certain data with a lengthy timeline. Among the many types of data it may receive are pictures, sounds, and images. Because of the high computational cost and huge number of model parameters associated with LSTM, a simplified GRU model is constructed instead. gates are the only remaining components of GRU's gate control architecture. The construction of GRU is basic, however it performs as well as LSTM in many cases. When presented with data that is comparable to vertical time stamps, both models function admirably. It is important to look at medical records from the past, or to integrate data from different time periods, especially when trying to anticipate certain diseases. RNN is a lifesaver in those kinds of situations. When dealing with real-world medical data, RNN is typically used in conjunction with CNN or other representations to address the spatial and temporal complexity of the problem.

5.5 Autoencoder (AE)

The fact that there is a mountain of unlabeled data out there. It encourages the development of algorithms for unsupervised learning. Data acquired in real-life medical situations is notoriously difficult to categorize. In order to find additional possible value in the data, we need technology that can automatically categorize or aggregate it. In addition, the real medical setting is relevant, and there are several ways to get medical data. Occasionally, data will be corrupted or missing, and information will be skewed. Autoencoder is a model for neural networks that efficiently reduces feature dimensionality, removes noise, and restores original data. It also performs well in learning.

Two parts make up the traditional AE model: encoding and decoding. Given that the input data x does not include any label information, AE utilizes x as both the input and the output simultaneously. The two components of the network are the encoding progression (f1:x>z) on the left and the decoding progression (f:z) on the right, when put into practice. In encoding, we get a lower-dimensional variable z, which represents the dimensionality decrease; in decoding, we get a new x, representing an increase in the z-dimensionality. In order to produce the optimized objective function, the primary goal of this method is to minimize the difference among the output x. Reduce L to its minimum value, which is given by dist. In order to improve the representation data obtained, the functions f1 and f2 are trained using deep neural networks and then integrated with additional deep learning models. It is possible to retrieve the original input x more effectively than using the machine learning model. The number of possible configurations for the AE model is large, and it is compatible with many others. Sparse Autoencoder, Variational Autoencoder (VAE), and other AE architectures may be built using changing medical data for practical applications. These architectures efficiently analyze low-recognition medical data from numerous angles.

5.7 Generative Adversarial Networks (GAN)

One type of model that uses game theory to train synthetic images is the GAN. Network training mostly represents the concept of game theory. A generation network and a discriminant network are each put up independently. The backpropagation after they employ a technique that is comparable to a competition. Afterwards, get better at helping each other out. The primary responsibility for learning the actual lies with the generating network, while the

Table 6: Heart disease prediction

discriminating network ensures that the models produced by the generating network are accurately differentiated from the actual samples. Both the discriminating and generating networks rely on convolutional neural networks, but the deconvolutional network is fundamental to both.

Nevertheless, of the picture size or GAN's presentation has been continuously increasing and breaking new ground since its 2014 proposal. The image processing industry is where GAN is most often used, and its many uses range from image-to-image translation to image restoration, resolution enhancement, medical picture fusion, and more. The use of Gan in medical imaging has been on the rise as of late, with impressive outcomes. The stability issues that traditional Gan often faces during training include the loss of gradients and mode collapse. Many recently proposed solutions to these issues exhibit promising results. The medical world has already seen the benefits of GAN, even if it is still in its early phases of medical imaging. Resolving these remaining issues will allow GAN to contribute significantly to the medical industry.

5.6 Other Models and Procedures

Aside from the models have also made several appearances in medical-related research in the past few years. Improving image fusion is one use case when processing medical pictures. The analysis of speech information by DBN can aid in the diagnosis of Parkinson's disease and other similar conditions. DBN has dual utility: it can function as a classifier in unsupervised learning, it can mimic the effects of AE to help doctors process medical data more thoroughly.

Combining numerous models improves the performance of research techniques when dealing with practical applications. In order to approach problems from different directions, many models are employed. By integrating CNN for horizontal data processing with RNN for vertical issue solving, followed by Bagging, the accuracy outperforms that of the individual models.

Over the past several years, transfer learning has exploded in popularity, and its integration with deep learning is a clear trend. Data reliance is a major issue with deep learning as it necessitates a substantial quantity of data. Little data is needed to complete transfer learning, which is comparable to a "polymorphic" concept; this concept offers ideas for reusing models, which not only saves time but also helps to develop the model even more. It is recommended that more diverse DL models be used in conjunction with transfer learning [15]. By utilizing transfer learning in medical research, new issues may be addressed using the trained model. The initial model was significantly adjusted for use in cancer picture lesion detection, and it performed admirably. Now that medical imaging technology is getting better, the medical industry may benefit from transfer learning and conduct more extensive research.

5. Related Studies on Deep learning for HD prediction

Author's Name	**Technique used**	**Merits**	**Dataset**	**Performance analysis**	**Demerits**
Bizimana, P. C et al. [16] (2023)	Machine Learning-Based Prediction Model (MLbPM)	Addresses the challenge of early and accurate prediction of heart disease, aiming for prevention and timely treatment.	a University of California Irvine HD dataset t	Achieves an accuracy of 96.7% when considering logistic worsening	The model's robustness to variations in the dataset or external factors is not addressed
Chandrasekhar, N., & Peddakrishna, S. [17] (2023)	Random Forest K-nearest Neighbor Logistic Regression Naïve Bayes Gradient Boosting AdaBoost Classifier	GridSearch CV enhances models performance by tuning parameters for optimal results.	Cleveland Dataset and IEEE Data port Dataset	Cleveland Dataset: Logistic Regression: 90.16% Soft Voting Ensemble: 93.44% IEEE Data port Dataset: AdaBoost: 90% Soft Voting Ensemble: 95%	Different algorithms may perform differently based on the dataset; the choice of algorithms might not be universally applicable.
Ahmad, A. A., & Polat, H. [18] (2023)	ML is employed as an artificial intelligence technology for disease detection, specifically heart disease prediction. Support Vector Machine (SVM) for Classifier.	Jellyfish Optimization Algorithm efficiently addresses the curse of dimensionality by reducing the dataset to a lower-dimensional subspace	Cleveland heart disease dataset is used.	Sensitivity: 98.56% Specificity: 98.37% Accuracy: 98.47% Area Under Curve: 94.48%	The study focuses on SVM classifier performance, and the generalizability of the results to other datasets is not explored
Ansari, G. A., et al. [19] (2023)	Logistic Regression (LR) K-Nearest Neighbor (KNN) Support Vector Machine (SVM) Naive Bayes (NB) Random Forest (RF) Decision Tree (DT)	The use of six feature selection algorithms adds depth to the analysis, indicating an effort to identify the most relevant features for heart disease prediction.	Cleveland Heart Dataset from UCI Machine Learning Repository	K-Nearest Neighbor (KNN) and Random Forest (RF) achieves accuracy Rate: 99.04%	The choice of top-performing algorithms may be sensitive to the characteristics of the specific dataset, limiting generalizability.

Shafiq, M., et al. [20] (2023)	Utilized for data collection through wireless sensor networks and digital stethoscopes. Deep Convolutional Neural Network (CNN) algorithm used for classification.	Leveraging IoT devices and cloud storage for efficient data collection, processing, and analysis.	PASCAL Dataset	Achieves high accuracy (98.4%), precision (99%), recall (98.7%), and F1-score (98.9%) using the deep CNN model.	The success of the proposed strategy may be sensitive to the characteristics of the PASCAL dataset, raising questions about its generalizability to diverse patient populations
Dileep, P., et al. [21] (2023)	Cluster-based Bi-directional Long Short-Term Memory (C-BiLSTM) used for heart disease prediction	C-BiLSTM leverages cluster-based techniques to improve the accuracy of traditional methods.	UCI heart disease dataset and real-time heart disease dataset utilized for testing deep learning techniques.	UCI Dataset Accuracy: 94.78% Real-Time Dataset Accuracy: 92.84%	The performance of deep learning methods, including C-BiLSTM, may vary based on the dataset, raising questions about generalizability
Shukur, B. S., & Mijwil, M. M. [22] (2023)	Utilizes machine learning techniques, including logistic regression, random forest, artificial neural network, support vector machines, and k-nearest neighbors, for heart disease diagnosis	Conducts tests to compare the diagnostic accuracy of different techniques, identifying the most effective method	For this experiment, we use the Cleveland Clinic dataset that we got from the UC Irvine machine learning repository and the Kaggle platform.	SVM achieves the most profitable performance in diagnosing heart disease, with a diagnostic accuracy of 96%.	The article lacks a discussion on the generalizability of the machine learning models.
Kadhim, M. A., & Radhi, A. M. [23] (2023)	• Random Forest • Support Vector Machines • K-Nearest Neighbor • Decision Tree	Applies hyperparameter optimization using random search to enhance the classification results, indicating an effort to fine-tune the model for improved accuracy.	Heart disease datasets obtained from the IEEE Data Port, integrating data from five well-known cardiac disease datasets (Long Shoreline VA, Hungarian, Cleveland, Starlog, and Switzerland).	Initial Classification Accuracy: 94.958% After Hyperparameter Optimization: 95.4%	The performance of machine learning algorithms can be sensitive to the characteristics of the dataset. The study does not explore how well the chosen algorithms generalize to new and unseen data.
Menshawi, A., et al. [24] (2023)	• LR • SVM • RF • XGBoost	Demonstrates the pipeline's ability to be	Public dataset from Kaggle known as the	Achieves a high accuracy of 95.6%	The paper does not explicitly discuss how well the proposed

	• DL	retrained and adapted for other datasets collected using different measurements and distributions	UCI heart disease dataset, collected from 1988.		framework generalizes to different healthcare systems with varying data collection practices and standards.
Shrivastava, P. K., et al. [25] (2023)	Incorporates Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) for prediction.	Addresses challenges related to missing data and imbalanced data in the dataset, enhancing the robustness of the model.	Heart disease Cleveland UCI dataset collected from Kaggle.	Achieves a high accuracy level of 96.66%.	The evaluation is conducted on a single dataset, and the generalizability of the model to diverse datasets with different characteristics is not extensively discussed.
Saboor, A., et al. [26] (2022)	Utilizes nine machine learning classifiers: AB (AdaBoost), LR (Logistic Regression), ET (Extra Trees), MNB (Multinomial Naive Bayes), CART (Decision Tree), SVM (Support Vector Machine), LDA (Linear Discriminant Analysis), RF (Random Forest), and XGB (XGBoost).	Improves classifier accuracy through hyperparameter tuning, enhancing the overall predictive capabilities.	The Cleveland heart disease dataset	Precision: Positive Classifiers: Ranges from 85% to 98%. Negative Classifiers: Ranges from 78% to 100%. Recall: Positive Classifiers: Ranges from 91% to 100%. Negative Classifiers: Ranges from 61% to 94%. F-Measure: Positive Classifiers: Ranges from 89% to 98%. Negative Classifiers: Ranges from 69% to 94%. Accuracy: Ranges from 83.66% to 96.72%.	The evaluation is performed on a single heart disease dataset, and the generalizability of the models to diverse datasets is not extensively discussed.
Subahi, A. F et al. [27] (2022)	Offers a Modified Self-Adaptive Bayesian Algorithm (MSABA) to accurately evaluate the risk of heart disease using data collected from sensors.	Uses Kernel Discriminant Analysis (KDA) for retrieving features from the sensor data,	The “Heart Disease Dataset” from the UCI Machine Learning Repository	Achieves 90% accuracy.	Limited insight into MSABA's generalizability across diverse datasets or populations, potentially hindering

		contributing to a more nuanced analysis.			applicability and reliability in varied scenarios.

5.8 Application of data mining in smart healthcare

Here we have covered some of the data mining uses in healthcare, including illness prediction, improved therapy, and more. What follows are:

- **Disease Diagnosis and Prediction:** A primary justification for utilizing knowledge in social insurance is the need to infer and anticipate pain in relation to social security firms.
- **Classification of Various Hospitals:** Taking into account the end goal of classification for each health center, the data mining technique zeroes in on their specific areas of interest. Organization's rate medical centers based on their capacity to treat actual patients; in other words, top-tier facilities can handle an influx of high-risk patients when the need arises.
- **Successful Treatments:** On the other hand, mining is employed to dissect the effectiveness of medications into its component parts, including causes, symptoms, signs, and cost.
- **Infection Prevention in Hospital sites:** In order to uncover such unexpected scenarios in disease control data, information mining is employed to examine contamination. For pollution control purposes, a qualified person conducts additional investigations into these cases.
- **Identifying Patients at High-Risk:** With the use of American Health approaches, clinics can improve efficiency and save administrative expenses for patients with heart disease [28].

6. Existing Problems and Technical Complications

Insufficient Interpretability

Despite the deep learning representation's stellar results in categorization and extraction, it has taken a lot of heat for being difficult to understand and work with. Research in deep learning and medicine, two distinct scientific disciplines, is certain to be diverse in every respect. There are now illness prediction algorithms that use deep learning techniques in a way that is distinct from medical domain knowledge. Paying close attention to the model's interpretability is crucial for correctly integrating these two types of subjects. The degree to which people can comprehend the decision-making rationale is reflected in the model's interpretability. In most cases, the model has to explain its prediction process, or the "why" issue. There is a strong integration of medical expertise into this method, as it is based on hand picked characteristics that are part of classic statistical approaches. Since deep learning relies solely on data, it pays little attention to prior domain experience, as well as a few recognized risk indicators.

An example of a black box model would be a deep neural network. Once the model has completed its training, it may be fed fresh data and then used to make predictions. People lose faith in the model since it does not teach them how to forecast, but rather returns the prediction based on the input data. We need to overcome the basic trustworthiness problem and make the model believable before we can apply the deep learning approach to actual illness forecast and the healthcare scheme. Only then can we realize the intelligence of disease prediction. For deep learning representations to be credible, they must meet specific requirements for interpretability. Deep learning will not be able to accomplish its goal of merging with real-world applications in the medical industry unless the issue of model interpretability is resolved. Finally, future studies should focus more on investigating model interpretability. In order to make the model more "humanized," it is important to choose the right methodologies and integrate them with actual needs, regardless of whether it is the criteria, interpretability. This is the only method to improve the credibility of deep learning concepts. This manner, we can provide patients a diagnosis that is both accurate and quick, and we can also give them a foundation for their diagnosis that is easy to understand.

6.1 Clinical Implementation Issues

Currently, there is a lack of clinical use for most deep learning approaches that mix illness prediction with other fields. This predicament is caused by several factors..

Generalization Ability

Some have questioned the model's usefulness and accuracy in real-world applications due to its weak generalization capabilities. For instance, it could be challenging to match the model's fundamental data needs in regions with subpar medical equipment, making it impossible to extract high-quality medical images. The most important factor is that DL necessitates more data. On, convolutional neural network (CNN) or other denoising-functional deep learning representations can be developed for this purpose. Furthermore, by utilizing the right filters, image enhancement technologies can be utilized to boost the image's overall or localized features, accentuate important details, and make the image more useful. Also, combine different types of picture data to strengthen the foundation of the diagnosis, and make the most of the current top-notch image fusion technology. There are still numerous issues with class imbalance and the sample is tiny when it comes to medical data. Research on data availability should receive greater focus going forward in order to enhance the model's capacity for generalization.

6.2 Stability

We still don't know if the deep learning model holds up in real-world scenarios. One of the requirements for genuinely applying applications is a high level of stability. The truth is that medical data are inconsistent and incomplete. It is possible for the neural network model to encounter a scenario where correspond to the real sample when it is applied to real-life medical situations. It is possible for the model to have aberrant features, inadequate classification accuracy, underfitting, or lack robustness when it processes actual samples after being modified using training samples. The algorithm's performance and efficiency will suffer, illness predictions will be off, clinical safety issues may arise, and patients' lives might be in jeopardy if the model's stability is not assured. It follows that guaranteeing the model's stability should underpin the realization of the shift to clinical application.

6.3 Privacy Security

Because deep learning illness prediction often makes use of patients' medical records and other personally identifiable information, it stands to reason that patients' right to privacy and data security must be carefully examined. Particularly pressing is the question of how to safeguard patients' privacy and stop the disclosure of sensitive information. Though progress has been made in data protection with dispersed technology in fresh years, and the objective of keeping data at the unique medical facility accomplished. Unfortunately, distributed systems are still in their early stages of development, and there has been insufficient investigation into how to integrate them with deep learning systems. Keeping patients' and physicians' personal information private should be a higher priority in future studies. Due to the impact of various medical problems, there exist several kinds of medical data, each with its own unique characteristics and varying degrees of quality. So, we need to make better use of current technology while also creating application technologies that can protect the privacy and security of medical data. The faster we can bring therapeutic applications to fruition, the less anxious patients will be about the disclosure of their personal information.

Several other elements also have a role, in addition to the aforementioned three. In illness prediction, for instance, the model relies on data that has been manually tagged. Some people's prejudices are bound to be present, and such biases might lead to unethical behavior. A worldwide agreement on ethics is obviously not simple to achieve, and therefore necessitates the creation of a practical ethical outline to prevent the impact of individual emotions. Not to mention that most regions lack the necessary prerequisites, there is a scarcity of skilled workers in relevant industries, and the application barrier for deep learning technology is very high. Deep learning technology has a long way to go before it can be used in clinical practice.

6.4 The Contest of Perfect Training

Some of the aforementioned learning representations, including AE and GAN, still retain their limitations. The veracity of synthetic data is one of the issues. It fails to persuade due to missing essential linkages and mappings. We must immediately address the issue of instability and the confusion surrounding evaluation indicators. Resolving these issues and shortcomings is crucial for the technology to gain confidence in the healthcare industry. Also, when training a model using medical data, over-fitting and under-fitting are more common. These issues will arise because to the aforementioned duplication or absence of a training data set. If you start with the model and make some structural changes, such expanding or contracting the dataset, you should be able to fix it.

Also, the most significant obstacle to overcome while training a model is poor data quality. It is on top-notch medical data that DL models do so well when it comes to illness prediction. Under the current system, medical records may be easily obtained; nevertheless, the data quality is inadequate. Experts in the medical field with extensive training are required to assign a private label to a large amount of medical data. There are several privacy concerns associated with medical data sets, which are kept in separate institutions, and the processing of picture characteristics is of utmost importance. Many databases are inaccessible and cannot be utilized for real-world research because they are closed and not open. The creation of several creative models is also impeded by the difficulty to acquire proper training.

7. Discussion and Further Recommendations

New developments in DL-based diagnostic procedures have piqued the interest of academics and experts from all over the world, particularly in the healthcare sector. The identification and categorization of CAD has made heavy use of these approaches during the last five years. The current research included 54 papers on DL's application to CAD data analysis and used tables and figures to outline the outcomes of the studies' comparisons and examinations. Research that used ECG, laboratory, and clinical data to identify CAD likely aimed to provide a model for reducing the need for expensive and invasive procedures like CCVT, echocardiography, and angiography. But if data like Echo were employed for CAD diagnosis, the goal would have been to provide a system for automated detection or consultation for detailed conclusions. By carefully reviewing previous studies and their obstacles and issues, we may offer suggestions for future study and address the sixth research topic.

While deep learning (DL) models and algorithms are becoming increasingly popular for early coronary artery disease (CAD) diagnosis and accurate clinical decision-making, we think future research should take certain issues into account. Some of the areas that are missing include:

(i) Data Pre-processing.

Data preparation and preprocessing is a significant step towards the efficiency and convergence of learning models, conferring to the texts. The presentation of DL models is not only dependent on the topology and assembly of networks. Research on data preprocessing—including data standardization, segmentation, feature engineering, and classification—has been lacking in certain investigations. For improved model performance, it is recommended to do preprocessing and feature selection activities based on the data type.

(ii) Sufficient Data.

Massive datasets are crucial to the performance of DL algorithms. The pattern detection of heart disorders is greatly impacted by the use of bigger datasets in DL approaches. The validity of DL models can be compromised if these approaches use inadequate data, which raises the chance of overfitting. Therefore, they are not reliable resources for addressing difficulties in clinical diagnosis.

(iii) Escape Overfitting Technique.

Several works on CAD detection using DL approaches failed to address overfitting ideas or provide solutions to this major problem with DL methods, according to the review of these papers. Nevertheless, it is important to take precautions when implementing DL techniques that are known to cause overfitting. When it comes to deep learning, this review recommends using larger datasets and the dropout method.

(iv) Comparing Different ML Techniques or CNN Pretrained Network.

The pattern identification and finding of CAD are areas where different DL networks and ML algorithms demonstrate differing levels of effectiveness. Therefore, it is suggested that future research use many methods and then choose the most efficient one. Similarly, when it comes to CNNs, we recommend testing out several pretrained networks and then tweaking the hyperparameters of the one that performed the best.

(v) Free Platform.

Quite a few studies have limited their model implementation to the realm of R&D. Nonetheless, PDAs, smartphones, and tablets can run models trained using DL networks and CNN algorithms. So, to analyze CAD devices that allow software and are accessible to clinical professionals, we suggest employing CNN networks, particularly lightweight networks.

8. CONCLUSION

Medical intervention effectiveness evaluations benefit greatly from data mining for healthcare, a multidisciplinary field that developed out of database statistics. Thanks to the internet, a wealth of medical history records are at our fingertips. Predicting illnesses now requires extracting and analyzing medical history data. The number of people losing their lives to heart attacks is steadily rising, particularly in the field of cardiovascular illness. We can reduce this rate by using medical history data to anticipate illness in cardiac patients. A research that compares HDP utilizing prominent categorization methods is proposed in this project. The inimitable feature structure of deep learning have become the major drivers of future progress in the face of the large dimensionality data. This article presents a number of well-known frameworks and approaches for deep learning illness prediction, as well as the best way to choose amongst them using data from typical medical diagnoses. Raised awareness of several current medical practical issues and deep learning's own limits. In the future, it is planned to work together with precision medicine to study diseases at the genetic to create smart medical platforms and devices, and with the particularity of medical data to suggest ideas for improving models. From the auxiliary diagnosis of today to decision-making diagnosis of tomorrow, we anticipate that illness diagnosis and deep learning will continue to grow in a additional diverse way in the future. There has been a proliferation of DL models for many illnesses. A more robust DL system network is the outcome of the models' interdependence and mutual learning. As a result, medical diagnosis and clinical application will advance, and the medical industry as a whole will benefit, from this intricate system.

REFERENCES

1. Nancy, A. A., Ravindran, D., Raj Vincent, P. D., Srinivasan, K., & Gutierrez Reina, D. (2022). Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. Electronics, 11(15), 2292. https://doi.org/10.3390/electronics11152292

2. Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments, 7(3), 263-275. https://doi.org/10.1007/s40860-021-00133-6

3. Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., & Masood, M. (2021). Prediction of heart disease using deep convolutional neural networks. Arabian Journal for Science and Engineering, 46(4), 3409-3422. https://doi.org/10.1007/s13369-020-05105-1

4. Nadeem, M. W., Goh, H. G., Khan, M. A., Hussain, M., & Mushtaq, M. F. (2021). Fusion-Based Machine Learning Architecture for Heart Disease Prediction. Computers, Materials & Continua, 67(2). DOI:10.32604/cmc.2021.014649

5. El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-label active learning-based machine learning model for heart disease prediction. Sensors, 22(3), 1184. https://doi.org/10.3390/s22031184

6. Waqar, M., Dawood, H., Dawood, H., Majeed, N., Banjar, A., & Alharbey, R. (2021). An efficient SMOTE-based deep learning model for heart attack prediction. Scientific Programming, 2021, 1-12. https://doi.org/10.1155/2021/6621622

7. Thirumalraj, Arunadevi and K, Aravinda and V, Revathi and Balasubramanian, Prabhu Kavin, Designing a Modified Grey Wolf Optimizer Based Cyclegan Model for Eeg Mi Classification in Bci. Available at SSRN: https://ssrn.com/abstract=4642989 or http://dx.doi.org/10.2139/ssrn.4642989

8. Javeed, A., Khan, S. U., Ali, L., Ali, S., Imrana, Y., & Rahman, A. (2022). Machine learning-based automated diagnostic systems developed for heart failure prediction using different types of data modalities: A systematic review and future directions. Computational and Mathematical Methods in Medicine, 2022. doi: 10.1155/2022/9288452

9. Mienye, I. D., & Sun, Y. (2021). Improved heart disease prediction using particle swarm optimization based stacked sparse autoencoder. Electronics, 10(19), 2347. https://doi.org/10.3390/electronics10192347

10. Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. Malaysian Journal of Computer Science, 132-148. DOI: https://doi.org/10.22452/mjcs.sp2022no1.10

11. Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. Algorithms, 16(2), 88. https://doi.org/10.3390/a16020088

12. Siddhartha, Heart Disease Dataset (Comprehensive), IEEE Dataport, 2020, https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive#files. DOI: https://doi.org/10.52866/ijcsm.2023.02.02.004

13. Abunasser, B. S., Al-Hiealy, M. R. J., Zaqout I. S. and Abu-Naser, S. S. "Breast Cancer Detection and Classification using Deep Learning Xception Algorithm" International Journal of Advanced Computer Science and

14. Applications (IJACSA), Vol. 13. No. 7, 2022, pp.223- 228, http://dx.doi.org/10.14569/IJACSA.2022.013072

15. Y. Khan, U. Qamar, N. Yousaf and A. Khan, "Machine learning techniques for heart disease datasets:A survey," in Int. Conf. on Machine Learning and Computing, Zhuhai, China, pp. 27–35, 2019. https://doi.org/10.1145/3318299.3318343

16. Sekar, J., Aruchamy, P., Sulaima Lebbe Abdul, H., Mohammed, A. S., & Khamuruddeen, S. (2022). An efficient clinical support system for heart disease prediction using TANFIS classifier. Computational Intelligence, 38(2), 610-640. https://doi.org/10.1111/coin.12487

17. Bizimana, P. C., Zhang, Z., Asim, M., El-Latif, A., & Ahmed, A. (2023). An Effective Machine Learning-Based Model for an Early Heart Disease Prediction. BioMed Research International, 2023. https://doi.org/10.1155/2023/3531420

18. Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. Processes, 11(4), 1210. https://doi.org/10.3390/pr11041210

19. Ahmad, A. A., & Polat, H. (2023). Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm. Diagnostics, 13(14), 2392. https://doi.org/10.3390/diagnostics13142392

20. Ansari, G. A., Bhat, S. S., Ansari, M. D., Ahmad, S., Nazeer, J., &

Eljialy, A. E. M. (2023). Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction. Computational and Mathematical Methods in Medicine, 2023. https://doi.org/10.1155/2023/8191261

21. Shafiq, M., Du, C., Jamal, N., Abro, J. H., Kamal, T., Afsar, S., & Mia, M. S. (2023). Smart E-Health System for Heart Disease Detection Using Artificial Intelligence and Internet of Things Integrated Next-Generation Sensor Networks. Journal of Sensors, 2023. https://doi.org/10.1155/2023/6383099

22. Dileep, P., Rao, K. N., Bodapati, P., Gokuruboyina, S., Peddi, R., Grover, A., & Sheetal, A. (2023). An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm. Neural Computing and Applications, 35(10), 7253-7266. https://doi.org/10.1007/s00521-022-07064-0

23. Shukur, B. S., & Mijwil, M. M. (2023). Involving machine learning techniques in heart disease diagnosis: a performance analysis. International Journal of Electrical and Computer Engineering, 13(2), 2177. DOI: 10.11591/ijece.v13i2.pp2177-2185

24. Kadhim, M. A., & Radhi, A. M. (2023). Heart disease classification using optimized Machine learning algorithms. Iraqi Journal For Computer Science and Mathematics, 4(2), 31-42. DOI: https://doi.org/10.52866/ijcsm.2023.02.02.004

25. Menshawi, A., Hassan, M. M., Allheeib, N., & Fortino, G. (2023). A Hybrid Generic Framework for Heart Problem diagnosis based on a machine learning paradigm. Sensors, 23(3), 1392. https://doi.org/10.3390/s23031392

26. Shrivastava, P. K., Sharma, M., & Kumar, A. (2023). HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms. Measurement: Sensors, 25, 100657. https://doi.org/10.3390/s23031392

27. Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M. F., & Ullah, N. (2022). A method for improving prediction of human heart disease using machine learning algorithms. Mobile Information Systems, 2022. https://doi.org/10.1155/2022/1410169

28. Subahi, A. F., Khalaf, O. I., Alotaibi, Y., Natarajan, R., Mahadev, N., & Ramesh, T. (2022). Modified Self-Adaptive Bayesian algorithm for smart heart disease prediction in IoT system. Sustainability, 14(21), 14208.

29. Guzzo, A. Rullo, E. Vocaturo, Process mining applications in the healthcare domain: a comprehensive review, Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 12 (2) (2022) e1442. https://doi.org/10.3390/su142114208

30. Mohamed, S., & Ejbali, R. (2023). Deep SARSA-based reinforcement learning approach for anomaly network intrusion detection system. *International Journal of Information Security*, *22*(1), 235-247.

31. Hnamte, V., & Hussain, J. (2023). DCNNBiLSTM: An efficient hybrid deep learning-based intrusion detection system. *Telematics and Informatics Reports*, *10*, 100053.

32. Wang, Y. C., Houng, Y. C., Chen, H. X., & Tseng, S. M. (2023). Network Anomaly Intrusion Detection Based on Deep Learning Approach. *Sensors*, *23*(4), 2171.

33. Abuali, K. M., Nissirat, L., & Al-Samawi, A. (2023). Advancing Network Security with AI: SVM-Based Deep Learning for Intrusion Detection. *Sensors*, *23*(21), 8959.

34. Alzboon, K., Al-Nihoud, J., & Alsharafat, W. (2023). Novel Network Intrusion Detection Based on Feature Filtering Using FLAME and New Cuckoo Selection in a Genetic Algorithm. *Applied Sciences*, *13*(23), 12755.

35. Alosaimi, S., & Almutairi, S. M. (2023). An Intrusion Detection System Using BoT-IoT. *Applied Sciences*, *13*(9), 5427.

36. Xiao, Y., & Xiao, X. (2019). An intrusion detection system based on a simplified residual network. *Information*, *10*(11), 356.

37. Wang, Z.; Wu, X.; Wang, H.; Wu, T. Prediction and analysis of domestic water consumption based on optimized grey and Markov model. *Water Supply* **2021**, *21*, 3887–3899.

38. Wu, J.; Wang, Z.; Dong, L. Prediction and analysis of water resources demand in Taiyuan City based on principal component analysis and BP neural network. *J. Water Supply Res. Tech.* **2021**, *70*, 1272–1286.

39. Wu, X.; Wang, Z.; Wu, T.; Bao, X. Solving the Family Traveling Salesperson Problem in the Adleman-Lipton model based on DNA computing. *IEEE Trans. Nanobiosci.* **2022**, *21*, 75–85.

40. Wang, Z.; Bao, X.; Wu, T. A Parallel Bioinspired Algorithm for Chinese Postman Problem Based on Molecular Computing. *Comput. Intel. Neurosc.* **2021**, *2021*, 8814947.

41. Austin, C.; White, R.; Buuren, V. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. *Can. J. Cardiol.* **2020**, *37*, 1322–1331.

42. Chang, C.; Deng, Y.; Jiang, X.; Long, Q. Multiple imputation for analysis of incomplete data in distributed health data networks. *Nat. Commun.* **2020**, *11*, 5467.

43. Bao, Z.; Chang, G.; Zhang, L.; Chen, G.; Zhang, S. Filling missing values of multi-station GNSS coordinate time series based on matrix completion. *Measurement* **2021**, *183*, 109862. https://doi.org/10.1016/j.measurement.2021.109862

44. Gia, B.; Te, A.; Bksa, C. Evaluating missing value imputation methods for food composition databases. *Food Chem. Toxicol.* **2020**, *141*, 111368.

45. Deng, A.; Wang, Z.; Liu, H.; Wu, T. A bio-inspired algorithm for a classical water resources allocation problem based on Adleman-Lipton model. *Desalin. Water Treat.* **2020**, *185*, 168–174.

46. Li, R.; Chang, Y.; Wang, Z. Study on optimal allocation of water resources in Dujiangyan irrigation district of China based on improved genetic algorithm. *Water Supply* **2021**, *21*, 2989–2999.

47. Ji, Z.; Wang, Z.; Deng, X.; Huang, W.; Wu, T. A new parallel algorithm to solve one classic water resources optimal allocation problem based on inspired computational model. *Desalin. Water Treat.* **2019**, *160*, 214–218.

48. Yu, M. Short-term wind speed forecasting based on random forest model combining ensemble empirical mode decomposition and improved harmony search algorithm. *Int. J. Green Energy* **2020**, *17*, 332–348.

49. Arvanaghi, R.; Danishvar, S.; Danishvar, M. Classification cardiac beats using arterial blood pressure signal based on discrete wavelet transform and deep convolutional neural network. *Biomed. Signal. Proces. Control.* **2022**, *71*, 103131.

50. Yan, S. Understanding LSTM and Its Diagrams. Available online: https://blog.mlreview.com/understanding-lstm-and-itsdiagrams-37e2f46f1714 (accessed on 8 February 2021). https://blog.mlreview.

51. Fang, Z.; Wang, Y.; Peng, L.; Hong, H. Predicting flood susceptibility using LSTM neural networks. *J. Hydrol.* **2020**, *594*, 125734.

52. Chen, Y.; Lin, M.; Yu, R.; Wang, T. Research on simulation and state prediction of nuclear power system based on LSTM neural network. *Sci. Technol. Nucl. Ins.* **2021**, *2021*, 8839867.

https://doi.org/10.1016/j.jhydrol.2020.125734

AUTHORS

U.A.Ranjni is working as a Assistant Professor in the CSE dept, Sathyabama of Science and Technology She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India. Her research interests include Artificial Intelligence, Machine learning, Deep learning.

Dr. Mercy Paul Selvan is working as a Professor in the CSE dept, Sathyabama of Science and Technology. She pursued her master's in computer science and engineering at St. Joseph's College of engineering and Ph.D. in Sathyabama Institute of Science and Technology. She has published 70 articles in reputed journals and conferences. She received 3 granted patents and three international awards in the field of education. Her research interests include AI, Machine learning, Deep learning, Data Science and GEN AI.