A Literature Review on Prediction of Chronic Diseases using Machine Learning Techniques

Siddegowda C. J. [1, 2] & A. Jayanthila Devi [3]

[1] Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore, Karnataka, India

[2]Assistant Professor, Dept. of MCA, Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India

ORCID: 000-0001-8206-9367; E-mail: siddegowdacj@gmail.com

[3] Research Professor, Institute of Computer Science & Information Science, Srinivas University, Mangalore– 575001, India.

ORCID: 0000-0002-6023-3899; Email: jayanthilamca@gmail.com

Area/Section: Computer Science.
Type of the Paper: Research Agenda based on Literature review.
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed in: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.6823291
Google Scholar Citation: IJMTS

How to Cite this Paper:

Siddegowda, C. J., & Jayanthila Devi, A. (2022). A Literature Review on Prediction of Chronic Diseases using Machine Learning Techniques. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *7*(1), 28-49. DOI: https://doi.org/10.5281/zenodo.6823291

International Journal of Management, Technology, and Social Sciences (IJMTS)
A Refereed International Journal of Srinivas University, India.

CrossRef DOI: https://doi.org/10.47992/IJMTS.2581.6012.0209

Received on: 25/03/2022
Published on: 14/07/2022

A Literature Review on Prediction of Chronic Diseases using Machine Learning Techniques

Siddegowda C. J. [1, 2] & A. Jayanthila Devi [3]

[1] Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore, Karnataka, India

[2]Assistant Professor, Dept. of MCA, Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India

ORCID: 0000-0001-8206-9367; E-mail: siddegowdacj@gmail.com

[3] Research Professor, Institute of Computer Science & Information Science, Srinivas University, Mangalore– 575001, India.

ORCID: 0000-0002-6023-3899; Email: jayanthilamca@gmail.com

ABSTRACT

Background/ Purpose: *Reviewing of the various work and literature in the proposed areas will help in developing a strong foundation of the domain on which the research is planned. The reason forth for the literature review is to become familiar in the health care domain. Since the area selected is the health care domain, the recent literature review is carried out as it is very important.*

Objective: *A strong background on health care domain is developed and a new problem which is not addressed is discussed. The gaps in the research area are identified. A new solution for solving the problem is designed and developed.*

Design/Methodology: *This work has adapted secondary source of data which is mainly journals, articles and review comments. The relevant literature is selected and a detailed study is conducted. This has helped in drafting the problem statement.*

Findings/Results: *The finding and drawbacks of all the recent work are well studied. The reason for the gap is also well studied and the results of each work are also well analyzed.*

Research Limitations: *A detailed study done on the chronic diseases and its impact has helped to open up the importance of studying about comorbid diseases. The limitations of various machine learning algorithms are also studied.*

Originality/Value: *This paper aims at studying the relevant existing literature that includes research journals, conference papers, technical book chapter and few web sources. All the papers selected were relevant to the proposed work and all papers are recent and from well reputed publisher. The papers are cited by many authors.*

Paper Type: *Literature review paper is carried out on scientific papers, especially from well indexed services.*

Keywords: Health care, Chronic diseases, Machine learning algorithms, Heterogeneous data, ABCD analysis

1. INTRODUTION :

Public health is given more importance and it is of primary concern of most of the country's where every county want to safeguard their community from high risk diseases which will affect the mortality rate of the state. Governments allocate huge amount of GDP for health and welfare and thereby to increase the life expectancy of the citizens. But in the last few years there has been considerable increase in the amount of chronic diseases like diabetic, cardiovascular diseases, cancer, etc. due to various factors which includes life style of the people.

The high availability of electronic health record (EHR) systems has led to tremendous promises in creation of many high risk disease prediction models which helps in prediction of various chronic diseases. As there are lots of hidden patterns available in the medical database, there is a great need for studying the prediction models that will help in reduction of hospital admission and medical cost. Machine learning techniques are used for prediction in several domains as the results obtained are

accurate and satisfactory better results. Various machine learning techniques used for detection of high risk diseases is studied and conferred in this paper.

2. RESEARCH OBJECTIVES :

The main objective of this paper is to identify the various findings and gaps which can be further explored by researchers. The following are objectives of this literature survey.

(1) To develop a strong background on health care domain.
(2) To come up with a new problem which is not addressed.
(3) To identify the research gap.
(4) To design a new solution for solving the identified problem.

3. METHODOLOGY :

The data required for understanding the results of various studies is collected from the existing sources. The data that may be required for simulating the proposed model can be collected form online databases as well as from hospitals. The data needed foe disease prediction is collected from medical professionals in hospitals and scan centres.

4. LITERATURE REVIEW :

The application of machine learning algorithms is at higher pace because of its efficiency to create efficient learning models. It uses many types of statistical and optimization methods to create training module sand to derive pattern from the past data and use the same of prediction effectively [1]. Popular machine learning algorithms are used in different domains like automated text categorization [2], network intrusion detection [3], junk e-mail filtering [4], detection of credit card fraud [5], customer purchase behavior detection [6], optimizing manufacturing process [7] and disease modeling [8] and in so many other areas. Off late supervised algorithms are more preferred than unsupervised algorithms because of various challenges. In supervised learning model, a training model is built based on training the dataset many times. As the data is labeled, it becomes easy to predict the outcome of unlabeled data. Recent studies have found that artificial intelligence (AI) and machine learning (ML) techniques can play a very important role in the prediction of high risk diseases. Machine learning methods can be applied to individual patient data which may fall under different research direction. The research finds are very useful as they help health professionals to create awareness among patients to prevent such high risk diseases. We review here the application of machine learning techniques on three major chronic diseases like diabetic, cardio vascular diseases and breast can.

The authors at [9] proposed a CNN-MDRP algorithm where data of huge volume of data is collected from hospital which is a combination of both structured and unstructured data. Which is trained and tested using Naïve Bayes algorithm .CNN based MDRP model is used to extract features from unstructured data. This has led to improvisation of prediction and the accuracy is around 94.2 percent the major drawback of this work is that this model is not suitable for complex diseases like cerebral diseases.

The authors at [10] has developed a decision-making model which is based on clinical construction of data which is used for the prediction of multiple diseases based on huge amount of historical data. The pattern which is not available in the training model is also considered in this work. Data related to multiple illnesses is also used in this work. The visualization of the results shows an accuracy of 94.5 percent. The major drawback of this work is it did not consider the feature reduction.

M. Chen proposed [11] a machine learning algorithm for prediction of chronic diseases that may affect the community in the form of a major disease outbreak. The real time data is taken from hospitals in China. A latent factor model is designed for reconstructing the data which may go missing sometimes. The prediction algorithm is multimodal in nature and has used the idea of CNN for training the model. The disease selected is a cerebral fraction. Most of the existing works have not used multi modal form of data and this work has achieved a result of 94.8% where the time complexity rate is also very low.

S. Leoni Sharmila, at al. [12] have made an extensive study on liver data set by using different machine learning technique such Fuzzy logic, Fuzzy Neural Network and decision tree. Fuzzy Neural network is used for classification of data and the study could achieve an accuracy of 91% from their experimentation.

Logistic regression (LR) is a method used for classification which is an extension for ordinary regression. This model is capable of modeling only a dichotomous variable that can be implemented on the occurrence or nonoccurrence of an event. This method is helpful for analyzing the probability of the condition on which a new instance may belong to a specific cluster of class [13]. Prediction of chronic diseases has gained prominence among the research community as it opened lots of scope for research and development.

Naïve Bayes is a classification technique used by research community which is mainly used for training and classification and this algorithm is basically derived from the Bayes' theorem [14]. This theorem is applied in finding the probability of an event from the knowledge base that is derived from the training model which is related to the condition to that event. This classifier derives its idea based on the assumption if a particular feature in a class are not related to any other feature of the same class, even though features belonging to that class are interdependent of themselves [15].

Decision tree is a support tool used for decision making and for predictive analytics. A decision tree model uses the concept of decision logic and test the data based on the logic and a tree-like structure is used for interpreting the results for classifying data items. The nodes are segregated into multiple levels where the first or the node in the highest position is represented as the root node [16].

A latent factor model is proposed by authors in [17] which is used for the reconstruction of health care data which is cerebral infraction data. As sometimes we may get incomplete data in health care, the model can handle incomplete data. This work has focused on multimodal data which is trained by using CNN.

Breast Cancer

Any Machine learning algorithm requires preprocessing, feature selection or extraction followed by classification [18]. Diagnosis and prognosis of cancer depends on Feature extraction, which can be extended to detect cancer setting to benign and malignant tumors [19] MS Uzer, et al. [20] has developed a breast cancer detection model which uses the concept of Artificial Neural Networks. The classification mode used is a combination of Backward and Forward Sequential Searching with PCA, using which an accuracy of 98.75 percent was achieved.

Author [21] presented usage of a neural wavelet network which looks at a linear and localized approach to detect breast cancer. The algorithm is trained on its parameters using a Recursive Least Squares (RLS) technique. The findings when compared to previous trials, demonstrated that the suggested technique is highly efficient and offers an excellent categorization [22-23]. To increase the accuracy of CAD-based procedures, boosting has been used by the author [24]. A hybrid boosting approach was developed, to approach the problem of characterization of breast cancer, this combines the benefits of multiple boosting techniques. The results of testing the various Boosting approaches on actual breast cancer show that the hybrid boosting algorithms surpasses the other boosting techniques by 48%.

An intelligent system was constructed by combining a classifier based on Support Vector (SVM) networks and a neural network (ANN) for automating breast cancer diagnosis [25]. The dataset used Wisconsin diagnostic records to train the SVM, it helped to distinguish between clusters which are benign from cancerous areas.

The dataset utilised in these experiments include Fine Needle Aspirates-based measures (FNA). Most of the work comparing various traditional statistics-based approaches along with traditional Machine Learning (ML) classification procedures has been published in order to highlight the benefits of (ML) and its potential [26].

In this work, the decision tree classifier is applied. This classifier appears to be using recursive splitting of the sample space [27]. The predictive approach here functions as a mapping between the object's characteristics and values [28]. A two phase support vector network was used which combined the two-phase clustering approach with a probability based SVM that evaluated the Wisconsin Breast Cancer Diagnosis (WBCD) dataset [29]. It achieved a classification model accuracy of 99.10 percent. Unlike other approaches, this strategy can recognize the figure of the masses and provide quick assessments for large bodies.

Cardio Vascular Diseases

Multiple conditions influence normal functioning of the heart which otherwise termed as heart or cardio vascular disease. Heart failure, cardiac arrest are the multiple terms used for explaining the abnormal

function of the heart, conditions leading to affecting the functioning of heart is referred to as a heart disease.

It is difficult to detect heart illness because of the existence of several health issues, such as high blood pressure, diabetes, irregular pulse rate, excessive cholesterol. A variety of neural networks-based approaches, data analysis has been employed to assess the severity of heart illness in humans.

A range of strategies, like Genetic Algorithm (GA), K-Nearest Neighbor (KNN) algorithm, DT, Naive Bayes (NB) algorithm, are used to classify the severity of sickness [30-31].

Several research have been conducted, as well as various machine learning models, with the aim of categorising and forecasting heart disease diagnoses. In the medical industry, ANNs were created to obtain the highest prediction accuracy feasible [32].

ANNs are utilized to estimate cardiovascular infection by means of back proliferation multi-facet perceptron (MLP). The subsequent discoveries are contrasted with those of recently distributed models in a similar region and viewed as altogether improved [33].

This examination work utilized the Artificial Neural Network (ANN) [34] a powerful method for planning the coronary illness expectation framework. Comparatively [35] uses DMT, a framework for infection detection in cerebral-vascular regions. It involves classifiers such as a Decision Tree or a Bayes network in an arrangement together with a Multi-layer neural network with Back Propagation (BP) for preparing the model. Therefore, it is shown that DMT can forecast CVD much more efficiently. Progressive DMT helps in recognizing the secret examples, and connections that are frequently unnoticed.

According to WHO estimates, over 23 million people would die from heart disease by 2030 [36]. Even though CVD is the leading cause, it is manageable and preventable. Analysis is critical for reducing the severity of the damage. The most difficult element is determining the correct illness [37]. Paper [38] presents a system model that can read human behaviors and properly predict trends in healthcare information. Naive-Bayes (NB) serves as the foundation for other algorithms and data processing methods. The technique used in [39] utilizes a Bayes Rule that computes prediction skills using a probabilistic approach. This aids in the investigation of novel methods of knowledge-based training, categorization, and prediction. The authors suggested a multimodal illness risk prediction system based on convolutional neural networks. For evaluating the proposed method, real-world hospital data from central China was collected between 2013 and 2015. The chronic illness of cerebral infraction was the subject of an experiment. The results of the experiments reveal that Naive Bayes works better for structured data; the suggested method performs better when combined.

The paper [40] presents proof-of-concept research. The datasets were created with python, and it used a Naïve Bayes and Decision Trees that exhibited increased precision in detecting heart illness. A machine learning based prediction model is proposed by authors in [41] where the heart data set is experimented on seven machine learning and feature selection and reduction algorithms are also compared. The ML model can easily classify healthy people from unhealthy ones. The results are validated with performance metrics. Two major algorithms Naïve Bayes and Decision Trees are used to create a model based on medical sets and the accuracy of both algorithms are simulated and compared, and naïve bayes has given better results than decision tree [42].

Dinesh, K. G. et al [43] have discussed heart disease assumption and performed data pre-getting ready uses methodologies like the expulsion of uproarious information, removal of missing information, using defaults where appropriate, and enumerating characteristics for forecast and decisions at various levels. Observing a model uses methods like request, precision, affectability, and identity assessment. The proposal can gauge whether individuals have heart disease or not based on the assessment.

Golande, P. Kumar T. [44] has considered both male and female individuals for study and this extent might vary according to the locale also this extent is considered for the people old enough bundle 25-70. It is not clear that people with one more in the age limit does not affect heart infirmities. Discussions on various estimations and gadgets used for the gauge of heart ailments have been carried out.

Prasad, P. et al. [45] have anticipated heart sicknesses by utilizing AI methodologies by spanning the two or three rhythmic movement investigates. They have utilized the determined relapse is used and the therapeutic administrations data which orchestrates the patients whether or not patients are having heart ailments according to the information on record and made information model that can predict probability of the patient would have heart sickness.

Khourdifi, Y. & Bahaj, M. [46] have discussed coronary sickness, mishandled the "Fast Correlation-Based Feature Selection (FCBF)" technique that can direct overabundance of features and estimate the parameters of a coronary ailment request. They have done request reliant upon different plan estimations, for instance, Support Vector Machine, K-Nearest Neighbor, Random Forest, Naïve Bayes, and a Multilayer Perception, Artificial Neural Network and Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) approach. The work utilizes a combination of algorithms on the heart disease dataset to achieve a high request accuracy of 99.65%.

In [47] another effort has been made to use a combination of several machine learning techniques, the methodology and goal is about finding basic incorporates by using Machine Learning. The model is composed of several features and a few known course of action systems. The model showed a high level of accuracy - 88.7% using the forecast model for coronary illness with a unique model that is similar with regards to Diverse information mining approaches and assumption strategies, such as, KNN, LR, SVM, NN, and Vote are extremely popular to recognize and foresee coronary illness. Sonam Nikhar et al [48] developed an algorithm that uses a combination of choice tree and Naïve Bayes classifiers. Experiments have gleaned insightful information mining methodology on identical datasets, and it has been shown that Decision Tree algorithm out performs a Bayesian network.

In [49] techniques for Heart Disease Prediction considering the contributing components like coronary illness (circulatory strain, diabetes, momentum smoker, elevated cholesterol, etc.). Various frameworks assist in mining data and neural network are used to find probability of illness in individuals.

Support vector machine (SVM) characterize structured and non-structured information. It can guide information into an n-layered space where n is the quantity of highlights. Recognizing the hyperplane helps to isolate the information into separate classes and augment the distance between the classes and helps in limiting error [50].

Diabetics:

In this day and age diabetes has become perhaps the most perilous and simultaneously most normal illnesses in Indium as well as across the world. It is found in all age categories nowadays and the cause is attributed to the way of life, hereditary, stress, and other factors. The causes may be many, the result might be serious whenever left unseen. Presently different strategies are being utilized to anticipate diabetes and diabetic caused illnesses. In this part, we discussed about the arrangement and expectation calculations for diabetes forecast in medical care.

The Pima Indians Diabetes Dataset (PIDD) was utilized in a few tests to foresee diabetes. [51-57] utilizing AI calculations and different techniques. The use of BLE sensors and AI calculations is critical in self-observing diabetes mellitus in medical care. AI is a major influence in the medical industry by giving the ability to express causes to medical experts to investigate, analyze clinical information [58-65]. Additionally, savvy medical services frameworks are giving constant clinical consideration to penniless patients [66-67]. Remarkably, we tweaked MLP for grouping because of its promising presentation in medical services, explicitly in diabetes forecast [68-71]. The creators [72] proposed a strategic relapse model in light of photoplethysmogram examination for diabetes detection. The proposed framework accurately calculated 92% as nondiabetic.

Maniruzzaman et al. [73] utilized an AI worldview to characterize and anticipate diabetes. Four AI calculations, i.e., credulous Bayes, choice tree, AdaBoost, and arbitrary timberland, was used to detect diabetes. Ahuja, et al [74] used different AI algorithms, like Naïve Bayes, Decision Tree and MLP, on PIMA dataset for diabetes detection. MLP was recommended approach for diabetes detection according to this study. In [26] utilizing MLP, an accuracy of 77.5% has been achieved on the PIMA dataset.

In [75] a credulous Bayes with a backing vector network calculations for diabetes characterization utilizing the PIMA dataset. Furthermore, utilizing component choice-based methodology and k-overlap cross-approval that works on the accuracy of the model. Experimental results have shown that the vector machine outperforms the guileless Bayes network. Be that as it may, cutting edge correlation is absent alongside accomplished precision. Choubey et al. [76] introduced a similar examination of arrangement strategies for diabetes grouping using PIMA Indian information gathered from the UCI Machine Learning Repository. Using AdaBoost, K-closest neighbor relapse, and spiral premise capacity to categorise patients as diabetic or not was performed.

In [77], medical services forecast framework in light of Naïve Bayes calculation is introduced. Proposed framework finds and concentrates stowed away information connected with various infections from

illness data set. This framework permits clients to express their wellbeing concerns and there after utilizes Naive Bayes to anticipate the right disease.

In [78] utilized datasets from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). The dataset contained in this exploration incorporates data connected with systolic circulatory strain (sBP), diastolic pulse (dBP), HDL, fatty substances (TG), BMI, fasting glucose (FBS), and orientation. Using Bootstrap conglomerating, Adaptive Boosting, and choice tree model, improved precision was observed. Adaboost when applied can foresee illnesses like diabetes, coronary illness, and hypertension. In [79], various AI techniques such as NB, SVM, DT on PIDD is studied. Later NB classifier was used and which could improve the precision by 76.30%. In [80] it is shown that diabetic forecast is calculated using relapse on PIDD. It was able to track down the quantity of pregnancies, BMI, and glucose level as important factors for diabetes forecast. The concept of feature ranking is used and Cleveland clinic database is used. The Particle Swarm Optimization as well as Neural Network Feed Forward Back-Propagation is used for training the model. A total of 13 effective factors are reduced to 8 optimized features which makes classification accurate. The results from the experiments proves that the selected features of the classified methods also showed that PSO method along with Neural Networks of Feed Forward Back-Propagation gives the accurate output of the rate of 91.94% on these features [81]. The high availability of electronic health record (EHR) systems has led to tremendous promises in creation of many high risk disease prediction models which helps in the prediction of various chronic diseases. As there are lots of hidden patterns available in the medical database, there is a great need for studying the prediction models that will help in reduction of hospital admission and medical cost. Machine learning techniques are widely used for prediction as it gives better results. Various machine learning techniques used for detection of high risk diseases is studied and conferred in this paper.

Table 1: Review on Related work

Reference no.	Year	Title	Methodology	Findings	Gaps
1	2014	Lesion Type Classification by Applying Machine-Learning Technique to Contrast-Enhanced Ultrasound Images	Adaboost machine learning	One Dimensional Haar-like features are used to describe the intensity changes in a TIC. Adaboost machine learning technique, is used for feature extraction. Hyperparameters of weak classifiers, e.g., the step size of a Haar-like filter length and threshold for output of the filter, are optimized by searching for those parameters that give the best	Performance becomes very low when the total scale of data increases.
2	2002	Machine learning in automated text categorization	These algorithms can be applied to various applications like automated text categorization	We review here the application of machine learning techniques on three major chronic diseases like diabetic, cardio vascular diseases and breast can and the results obtained is satisfactory	If the quality of the data is not good, the results turn out to be biased.

3	1999	An application of machine learning to network intrusion detection	network intrusion detection	The research finds are very useful as they help health professionals to create awareness among patients to prevent such high risk diseases.	Fatal errors are high
4	2018	Disease Risk Prediction by Using Convolutional Neural Network	Risk analysis of diseases	Artificial intelligence (AI) and machine learning (ML) techniques can play a very important role in the prediction of high risk diseases.	Low performance for Handling enormous amount of data
5	2015	An Improved Method for Disease Prediction Using Fuzzy Approach	Prediction model for liver diseases	A fuzzy c-means clustering algorithm is used in combination with KNN classifier. These model could classify the diseases more accurately.	Computational complexity is high
6	2017	Chronic disease risk prediction using distributed machine learning classifiers.	Kidney disease prediction	The authors have used distributed ML Classifiers and the paper to predict Chronic-Kidney-Disease. The application of this algorithm is very useful in the domain of medical bioinformatics.	Low accuracy while there is increase in data
7	2013	A novel method for disease prediction: hybrid of random forest and multivariate adaptive regression splines	optimizing manufacturing process	The authors have applied machine learning algorithms to read and train the patient data and has failed to address the problem of fault tolerance	Low fault tolerance
8	2013	Machine learning: an artificial intelligence approach.	disease modeling	The research finds are very useful as they help health professionals to create awareness among patients to prevent such high risk diseases. We review here the application of machine learning	Not recommended when missing values are more

				techniques on three major chronic diseases like diabetic, cardio vascular diseases and breast cancer.	
9	2017	Disease classification using machine learning algorithms-a comparative study	Fuzzy logic, Fuzzy neural network, decision tree	The major challenge is diagnose the disease at a novice stage with the help of prediction models. The authors have studied and compared different machine learning techniques such Fuzzy logic, Fuzzy Neural Network and decision tree in classifying the presence of liver disease from the data set.	Needs more machine learning technique to address in comparative analysis
10	2017	Data mining and visualization for prediction of multiple diseases in healthcare.	has developed a decision-making model	Has developed a decision-making model which is based on clinical construction of data which is used for prediction of multiple diseases based on huge amount of historical data. The pattern which is not available in the training model is also considered in this work. Data related to multiple illnesses is also used in this work. The visualization of the results shows an accuracy of 94.5 percent. The major drawback of this work is it did not consider the feature reduction.	Produces tight clusters when number of variables are large
11	2019	Various preprocessing methods for neural network based heart disease prediction	Neural network based heart disease prediction	Have used neural network for prediction of heart diseases that affect majority of people with a disease outbreak. The real	Increases computational cost.

				time data is taken from hospitals. The prediction algorithm is multimodal and has used the idea of NN for training the model. The disease selected is heart disease. Most of the existing works have not used this kind of data and this work has achieved a result of 96.7% where the time complexity rate is also very low.	
12	2017	"Disease Classification Using Machine Learning Algorithms - A Comparative Study"	machine learning technique such Fuzzy logic, Fuzzy Neural Network and decision tree	Have made an extensive study on liver data set by using different machine learning technique such Fuzzy logic, Fuzzy Neural Network and decision tree. Fuzzy Neural network is used for classification of data and the study could achieve an accuracy of 91% from their experimentation.	Computational cost cannot be brought down
13	2013	Applied logistic regression	Logistic regression (LR) is a method used for classification which is an extension for ordinary regression.	Logistic regression (LR) is a method used for classification which is an extension for ordinary regression. This model is capable of modelling only a dichotomous variable that can be implemented on the occurrence or non-occurrence of an event. This method is helpful for analysing the probability of the condition on which a new instance may belong to a specific cluster of class.	Not robust if training data has too much of noise
14	1958	Fiducial distributions and Bayes' theorem	Naïve Bayes is a well-known machine learning	Prediction of chronic diseases has gained prominence among the research	It works well only with large datasets

				community as it opened lots of scope for research and development.	with high dimensionality
15	2001	An empirical study of the naive Bayes classifier	Bayes classifier This theorem is applied in finding the probability of an event from the knowledge	This theorem is used in this work for tracing the probability of an event from the existing knowledge base. The knowledge base is built from training the data set and build a model which is correlated to the condition to that event. This classifier is based on the assumption if a specific feature in a class is not in relation to any other feature of the class even though features belonging to that class are interdependent of themselves	Target features were imbalanced which affected the accuracy of the work
16	1986	Induction of decision trees.	A decision tree model uses the concept of decision logics and test the data	Decision tree is a support tool used for decision making and for predictive analytics. A decision tree model uses the concept of decision logics and test the data based on the logic and a tree-like structure is used for interpreting the results for classifying data items. The nodes are divided into different levels where the first or top-most node is represented as root node	Risk level parameter is ignored in this work
17	2017	Disease Prediction by Machine Learning Over Big Data From Healthcare Communities	A latent factor model is proposed and CNN	A latent factor model is proposed by which is used for reconstruction of health care data which is cerebral infraction data. As sometimes we may	The sample size selected is less which affected the accuracy of the work

				get incomplete data in health care, the model can handle incomplete data. This work has focused on multimodal data which is trained by using CNN	
20	2013	A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS, and PCA	designed a breast cancer detection method which is based on Artificial Neural Networks	Designed a breast cancer detection method which is based on Artificial Neural Networks. The classification mode used a combination of Backward and Forward Sequential Searching with PCA, using which an accuracy of 98.75 percent was achieved.	The hidden layer calculation would have been used
21	2013	Local linear wavelet neural network for breast cancer recognition	The algorithm is trained on its parameters using a Recursive Least Squares (RLS) technique.	Presented usage of a neural wavelet network which looks at a linear and localized approach to detect breast cancer. The algorithm is trained on its parameters using a Recursive Least Squares (RLS) technique. The findings when compared to previous trials, demonstrated that the suggested technique is highly efficient and offers an excellent categorization	Parameter selection is not elaborate
29	2019	Edge and fog computing in critical infrastructures: analysis, security threats, and research challenges	SVM is used	A two phase support vector network was used which combined the two-phase clustering approach with a probability based SVM that evaluated the Wisconsin Breast Cancer Diagnosis (WBCD) dataset. It achieved a classification model accuracy of 99.10	Parameter selection is not elaborate

				percent. Unlike other approaches, this strategy can recognize the figure of the masses and provide quick assessments for large bodies.	
32	2015	"Predictions in heart disease using techniques of data mining,	ANNs were created to obtain the highest prediction accuracy feasible	Several research have been conducted, as well as various machine learning models, with the aim of categorising and forecasting heart disease diagnoses. In the medical industry, ANNs were created to obtain the highest prediction accuracy feasible.	feature selection technique used did not have an ability to enhance the performance of machine learning models.

Table 2: An insight to review on machine learning techniques

Source	Year	Advantage	Limitation	Result
[82]	2019	Accuracy is high	-	LR Is 87 Percent
[83]	2019	Accuracy is high	Sensitivity is low	For Artificial Neural Network-77 Percent,C4 Is 57,6 Percent And Hybrid-Decision Tree Is 78 Percent
[84]	2015	Accuracy is high	-	chronic liver disease (odds ratio [OR] 3.71), high alanine aminotransferase (OR 2.26), esophageal reflux (OR 1.85), and history of acute bronchitis (OR 1.45)
[85]	2019	Accuracy is high	Blackbox is of importance	Naïve bayes-84 Percent, Support Vector Machine-85 Percent, Voting-87 Percent
[86]	2018	Accuracy is high	High computation time	Lr-89 Percent K-Nearest Neighbour-80 Percent Artificial Neural Network-77 Percent Support Vector Machine-87 PercentNaïve bayes-85 Percent Decision Tree-73 Percent Random Forest-85 Percent
[87]	2018	Accuracy is high	Training Time is high	Artificial Neural Network-84 Percent, Support Vector Machine-82 Percent Naïve Bayes-83 Percent
[88]	2018	Accuracy is high	High Computation Time	NN-87 Percent, Gini Model-77 Percent,J48-71 Percent, Bagging-76 Percent, ProposedEFM-79 Percent
[89]	2017	Accuracy is high	Training Time is high	CNN-94 Percent

[90]	2017	Accuracy is high	Minimum number of features are used for training	FuzzyAHP-Artificial Neural Network-83 Percent
[91]	2017	Accuracy is high	High Computation Time	PCA-K-Nearest Neighbour-79 Percent, Pca-Support Vector Machine-89 Percent, EM-PCA-Fuzzy-Rule-92 Percent

Table 3: Review on Disease dataset based machine learning

Source	Disease Dataset Considered	Method Applied	Accuracy
[92]	Kidney	Pre-processing and K-Nearest Neighbour, K-Nearest Neighbour	96 % 81 %
[93]	Heart	Vote Technique	87.4 %
[94]	Kidney	Support Vector Machine	98.5 %
[95]	Heart	Rule-Based Fuzzy Classifier	78 %
[96]	Tumor in Breast	Relief F-Support Vector Machine	90 %
[97]	Kidney	Naïve bayes, Support Vector Machine	80.75 % 82.35 %
[98]	Diabetes	Support Vector Machine	100 % 100 %
[99]	Lungs	Genetic Algorithm Based Feature Selection	99 %
[100]	Breast	SFSP+NN, SBSP+NN	97.57 % 98.57 %
[101]	Hepatitis	FDA+Support Vector Machine	96.77 %
[102]	Lymph	PCA+Fuzzy Weighting Pre-Processing+ANFIS	88.83

Table 3: Review on Disease dataset based machine learning

Source	Disease Dataset Considered	Method Applied	Accuracy
[92]	Kidney	Pre-processing and K-Nearest Neighbour, K-Nearest Neighbour	96 % 81 %
[93]	Heart	Vote Technique	87.4 %
[94]	Kidney	Support Vector Machine	98.5 %

[95]	Heart	Rule-Based Fuzzy Classifier	78 %
[96]	Tumor in Breast	Relief F-Support Vector Machine	90 %
[97]	Kidney	Naïve bayes, Support Vector Machine	80.75 % 82.35 %
[98]	Diabetes	Support Vector Machine	100 % 100 %
[99]	Lungs	Genetic Algorithm Based Feature Selection	99 %
[100]	Breast	SFSP+NN, SBSP+NN	97.57 % 98.57 %
[101]	Hepatitis	FDA+Support Vector Machine	96.77 %
[102]	Lymph	PCA+Fuzzy Weighting Pre-Processing+ANFIS	88.84

5. PRESENT STATUS :

In the recent days, there are many research activities for the automatic prediction of diseases in the health care domain. Most of the existing methods lack accuracy due to various technical reasons. The recent research also focuses on minimum classification error in disease prediction automatically. So that it helps the medical professional to do the needful.

6. RESEARCH GAPS AND PROPOSAL :

After an extensive review of the related literature, it is observed that there is still a huge scope for exploring the area of chronic disease detections. There are not sufficient papers for studying the automatic prediction of diseases in health care of humans. This survey finds the following research gaps and also suggest solutions for overcoming these research gaps.

(1) **Research Gap1:** To identify the chronic diseases with minimum computation time.
(2) **Research Gap2:** Classify the different types of chronic diseases in human.
(3) **Research Gap3:** To optimize the computation efficiency.

7. RESEARCH PROPOSAL :

(1) The proposed disease detection framework can be implemented using Matlab.
(2) Various machine learning algorithms have to be identified and analysed and the disease has to be accurately classified.

8. RESEARCH AGENDAS :

(1) Which technology can be used for preprocessing?
(2) What is the use of segmentation techniques?
(3) How to find efficient feature extraction time?
(4) How to identify the efficient machine learning technique with minimum computation time?
(5) What type of methods can be used to automatically predict the diseases?
(6) Which machine learning technique can be used to minimize the classification errors?

9. ANALYSIS OF RESEARCH AGENDAS :

The first step is to collect the primary data from hospitals in discussion with health experts. The consent form from the health professional and patients also need to be collected. This gives clarity on the quality of data collected. Apart from primary data, the proposed prediction model has to be validated with

benchmark data also Noise removal techniques has to be applied as a part of preprocessing techniques. The image has to be enhanced if required. The relevant simulation tool has to be selected to implement the learning model. Finally we have to use validation metrics to justify the accuracy of the results achieved.

10. LIMITATIONS OF THE PROPOSAL :

The scope of this paper is limited to health care domain. A detailed study can be done to use it in various image processing fields. This proposal gives a detailed study about normal and the abnormal status of human. A detailed study can be done to suggest a solution in abnormal cases of disease prediction.

11. ABCD ANALYSIS :

ABCD analysis is associated with advantages, benefits, constraints, and disadvantages in a systematic manner [57-59]. The ABCD listing of prediction of chronic diseases using Machine Learning Techniques is shown in table 4. The benefits of the ABCD analysis can be effectively used to find the optimal results from the proposed scientific model. It has wide scope for applications in various research areas. Though it is used more to reduce the loss in the inventory management system, it can also be used in image processing areas to reduce errors while simulating the proposed model.

Table 4: ABCD Listing of prediction of chronic diseases using Machine Learning Techniques

ADVANTAGE	BENEFITS
• Efficient Result • Accurate classification • Minimum cost	• Increasing demand in health care • Increased concern in medical professions in health care • Efficient prediction of diseases
CHALLENGES	**DISADVANTAGE**
• Risky datas like eye diseases, heart diseases • Classification errors while predicting disease	• Mis classification errors. • High computation time

12. SUGGESSTIONS TO IMPLEMENT RESEARCH ACTIVITES ACCORDING TO PROPOSAL :

As a first step, data has to be primarily collected from health centres as well the secondary source of open data bases available. A relevant preprocessing technique has to be used to clean the data. If the collected data has noise it has to be removed. A technique for image enhancement has to be introduced. A segmentation method or region of interest model has to be used. Finally a machine learning classification approach has to be used to classify whether the disease may lead to comorbidity or not. The result of the proposed model has to be validated with performance parameters.

13. CONCLUSION :

This paper has studied in detail about how machine learning algorithms can be used in the health care domain for prediction of chronic diseases like diabetics, breast cancer, and cardio vascular diseases. The primary focus of this study is to analyze how different algorithms are used by researchers in research for predicting different types of diseases using machine learning. From the study, we have concluded that most of the existing works were only predicting the chronic disease and very less work is done for the prediction of chronic diseases using Machine Learning Techniques.

REFERENCES :

[1] Takagi, K., Kondo, S., Nakamura, K., & Takiguchi, M. (2014). Lesion type classification by applying machine-learning technique to contrast-enhanced ultrasound images. *IEICE TRANSACTIONS on Information and Systems*, *97*(11), 2947-2954. Google Scholar↗

[2] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, *34*(1), 1-47. Google Scholar↗ CrossRef↗

[3] Sinclair, C., Pierce, L., & Matzner, S. (1999, December). An application of machine learning to network intrusion detection. In *Proceedings 15th Annual Computer Security Applications Conference (ACSAC'99)*, IEEE, *8(1)*, 371-377. Google Scholar↗

[4] Ambekar, S., & Phalnikar, R. (2018, August). Disease risk prediction by using convolutional neural network. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)* IEEE., *5*(3), 1-5. Google Scholar↗ CrossRef↗

[5] Chetty, N., Vaisla, K. S., & Patil, N. (2015, May). An improved method for disease prediction using fuzzy approach. In *2015 Second International Conference on Advances in Computing and Communication Engineering* IEEE., *6*(4), 568-572. Google Scholar↗

[6] Jena, L., & Swain, R. (2017, December). Work-in-progress: Chronic disease risk prediction using distributed machine learning classifiers. In *2017 International Conference on Information Technology (ICIT)* IEEE, *1(8),* 170-173. Google Scholar↗

[7] Yao, D., Yang, J., & Zhan, X. (2013). A Novel Method for Disease Prediction: Hybrid of Random Forest and Multivariate Adaptive Regression Splines. *J. Comput.*, *8*(1), 170-177. Google Scholar↗

[8] Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). *Machine learning: An artificial intelligence approach*, *Springer Science & Business Media*, *5*(2), 23-27. Google Scholar↗

[9] Sharmila, S. L., Dharuman, C., & Venkatesan, P. (2017). Disease classification using machine learning algorithms-a comparative study. *International Journal of Pure and Applied Mathematics*, *114*(6), 1-10. Google Scholar

[10] Kunjir, A., Sawant, H., & Shaikh, N. F. (2017, March). Data mining and visualization for prediction of multiple diseases in healthcare. In *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)* IEEE, *3(8),* 329-334. Google Scholar↗

[11] Burse, K, Kirar, VPS, Burse, A & Burse, R, (2019). Various preprocessing methods for neural network based heart disease prediction. Smart Innovations in Communication and Computational Sciences, *Advances in Intelligent Systems and Computing, Springer, Singapore, 851*(1), 55-65. Google Scholar↗ CrossRef↗

[12] Sharmila, S. L., Dharuman, C., & Venkatesan, P. (2017). Disease classification using machine learning algorithms-a comparative study. *International Journal of Pure and Applied Mathematics*, *114*(6), 1-10. Google Scholar↗ CrossRef↗

[13] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons. Google Scholar↗

[14] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 7(1), 102-107. Google Scholar↗

[15] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence,* 3(22), 41-46. Google Scholar↗

[16] Quinlan, J. R. (1986). *Induction of decision tree,* 6(7), 81-106. Google Scholar↗ CrossRef↗

[17] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *Ieee Access*, 5(1), 8869-8879. Google Scholar↗

[18] Castaneda, C, Nalley, K, Mannion, C, Bhattacharyya, P, Blake, P, Pecora, A & Suh, KS (2015). Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *Journal of Clinical Bioinformatics*, *5*(1), 1-4. Google Scholar↗

[19] Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. *Journal of healthcare engineering*, *8*(1), 1-17. Google Scholar↗

[20] Uzer, M. S., Inan, O., & Yılmaz, N. (2013). A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA. *Neural Computing and Applications*, *23*(3), 719-728. Google Scholar↗

[21] Senapati, M. R., Mohanty, A. K., Dash, S., & Dash, P. K. (2013). Local linear wavelet neural network for breast cancer recognition. *Neural Computing and Applications*, *22*(1), 125-131. Google Scholar↗ CrossRef↗

[22] Garg, H., & Lal, N. (2018, April). Data analysis: Opinion mining and sentiment analysis of opinionated unstructured data. In *International Conference on Advances in Computing and Data Sciences*, Springer, Singapore, *4*(1), 249-258. Google Scholar↗

[23] Kaur, N., & Lal, N. (2018, April). Clustering of Social Networking Data using SparkR in Big Data. In *International Conference on Advances in Computing and Data Sciences*, Springer, Singapore. *6(5)*, 217-226. Google Scholar↗

[24] Hilal, A. R., & Basir, O. (2009, December). Combination of enhanced AdaBoosting techniques for the characterization of breast cancer tumors. In *2009 International Conference on Future BioMedical Information Engineering (FBIE)*, IEEE. *5(2)*, 568-571. Google Scholar↗

[25] Maglogiannis, I., Zafiropoulos, E., & Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Applied intelligence*, *30*(1), 24-36. Google Scholar↗ CrossRef↗

[26] Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998, July). A Bayesian approach to filtering junk e-mail. In *Learning for Text Categorization: Papers from the 1998 workshop*, *62*(1), 98-105. Google Scholar↗

[27] Mohammed, M. A., Abd Ghani, M. K., Hamed, R. I., & Ibrahim, D. A. (2017). Analysis of an electronic methods for nasopharyngeal carcinoma: Prevalence, diagnosis, challenges and technologies. *Journal of Computational Science*, *21(1)*, 241-254. Google Scholar↗ CrossRef↗

[28] DeMántaras, R. L. (1991). A distance-based attribute selection measure for decision tree induction. *Machine learning*, *6*(1), 81-92. Google Scholar↗

[29] Tedeschi, P., & Sciancalepore, S. (2019, June). Edge and fog computing in critical infrastructures: analysis, security threats, and research challenges. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* IEEE, 3(2), 1-10. Google Scholar↗

[30] Khourdifi, Y., & Bahaj, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *International Journal of Intelligent Engineering and Systems*, *12*(1), 242-252. Google Scholar↗ CrossRef↗

[31] Durairaj, M., & Revathi, V. (2015). Prediction of heart disease using back propagation MLP algorithm. *International Journal of Scientific & Technology Research*, *4*(8), 235-239. GoogleScholar↗

[32] Gandhi, M., & Singh, S. N. (2015, February). Predictions in heart disease using techniques of data mining. In *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, *6(1)*, 520-525. Google Scholar↗

[33] Baccour, L. (2018). Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets. *Expert Systems with Applications*, *99(1)*, 115-125. Google Scholar↗

[34] Lutimath, N. M., Ramachandra, H. V., Raghav, S., & Sharma, N. (2022). Prediction of Heart Disease Using Genetic Algorithm. In *Proceedings of Second Doctoral Symposium on Computational Intelligence* Springer, Singapore, *8*(1), 49-58. Google Scholar↗ CrossRef↗

[35] Sudha, A., Gayathri, P., & Jaisankar, N. (2012). Effective analysis and predictive model of stroke disease using classification methods. *International Journal of Computer Applications*, *43*(14), 26-31. Google Scholar↗

[36] Swathy, M., & Saruladha, K. (2022). A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques. *ICT Express*, *8*(1), 109-116. CrossRef↗

[37] Hazra, A., Mandal, S. K., & Gupta, A. (2016). Study and analysis of breast cancer cell detection using Naïve Bayes, SVM and ensemble algorithms. *International Journal of Computer Applications*, *145*(2), 39-45. Google scholar↗

[38] Vinothini, S., Singh, I., Pradhan, S., & Sharma, V. (2018). Heart Disease Prediction. *International Journal of Engineering & Technology, 7 (3.12)*, 750-753. Google Scholar↗

[39] Singh, G., Bagwe, K., Shanbhag, S., Singh, S., & Devi, S. (2017). Heart disease prediction using Naïve Bayes. *Int. Res. J. Eng. Technol*, *4*(3), 4-6. Google Scholar↗

[40] Taylor, R. A., Pare, J. R., Venkatesh, A. K., Mowafi, H., Melnick, E. R., Fleischman, W., &Hall, M. K. (2016). Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach. *Academic emergency medicine*, *23*(3), 269-278. Google Scholar↗

[41] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, *6*(5), 300-310. Google Scholar↗

[42] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in genetics*, *4*(1), 515-516. Google Scholar↗

[43] Krishnan, S., & Geetha, S. (2019, April). Prediction of Heart Disease Using Machine Learning Algorithms. In *2019 1st international conference on innovations in information and communication technology (ICIICT)*, IEEE. *4(4)*, 1-5. Google Scholar↗

[44] Gupta, S., Verma, H. K., & Bhardwaj, D. (2021). Classification of diabetes using Naive Bayes and support vector machine as a technique. *In Operations Management and Systems Engineering, 5*(2). 365-376. Google Scholar↗

[45] Osman, A. H. (2017). An enhanced breast cancer diagnosis scheme based on two-step-SVM technique. *Int. J. Adv. Comput. Sci. Appl*, *8*(4), 158-165. Google Scholar↗

[46] Das, R., Turkoglu, I., & Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. *Expert systems with applications*, *36*(4), 7675-7680. Google Scholar↗ CrossRef↗

[47] Cotter, M. (2019). *A Systems View of Cardiovascular Disease*, *7*(1), 14-29. Google Scholar↗

[48] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*(2), 81542-81554. Google Scholar↗

[49] Nikhar, S., & Karandikar, A. M. (2016). Prediction of heart disease using machine learning algorithms. *International Journal of Advanced Engineering, Management and Science*, *2*(6), 239-484. Google Scholar↗

[50] Farhadi, A., Chen, D., McCoy, R., Scott, C., Miller, J. A., Vachon, C. M., & Ngufor, C. (2019, October). Breast cancer classification using deep transfer learning on structured healthcare data. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* IEEE, *1*(4), 277-286). Google Scholar↗

[51] Sisodia, D., Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science. 132*(1), 1578-1585. Google Scholar↗ CrossRef↗

[52] Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, *10(1)*, 100-107. Google Scholar↗ CrossRef↗

[53] Jacobs, I. S. (1963). Fine particles, thin films and exchange anisotropy. *Magnetism*, *1*(2), 271-350. Google Scholar↗

[54] Choubey, D. K., & Paul, S. (2017). GA_RBF NN: a classification system for diabetes. *International Journal of Biomedical Engineering and Technology*, *23*(1), 71-93. Google Scholar↗ CrossRef↗

[55] Tigga, N. P., & Garg, S. (2021). Predicting type 2 diabetes using logistic regression. In *Proceedings of the Fourth International Conference on Microelectronics, Computing and Communication Systems*, Springer, Singapore, *1*(3), 491-500. Google Scholar↗

[56] Nai-arun, N., & Moungmai, R. (2015). Comparison of classifiers for the risk of diabetes prediction. *Procedia Computer Science*, *6(1)*, 132-142. Google Scholar↗ CrossRef↗

[57] Aithal, P. S., Shailashree, V., & Kumar, P. M. (2015). A new ABCD technique to analyze business models & concepts. *International Journal of Management, IT and Engineering*, *5*(4), 409-423. Google Scholar↗

[58] Aithal, P. S., & Aithal, S. (2018). Factor & Elemental Analysis of Nanotechnology as Green Technology using ABCD Framework. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *3*(2), 57-72. Google Scholar↗

[59] Aithal, P. S. (2017). ABCD Analysis as Research Methodology in Company Case Studies. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *2*(2), 40-54. Google Scholar↗

[60] Maheswari, S., & Rani, P. A. J. (2016). RVM-based human action classification through Gabor and Haar feature extraction. *International Journal of Computational Vision and Robotics*, *6*(1-2), 54-64. Google Scholar↗

[61] Wadhwa, S., &Babber, K. (2020, July). Artificial intelligence in health care: predictive analysis on diabetes using machine learning algorithms. In *International Conference on Computational Science and Its Application*, Springer, Cham, *1*(7), 354-366. Google Scholar↗ CrossRef↗

[62] Fairooz, A. (2021). *Coronary Illness Prediction Using Machine Learning*, *1*(4), 457-465. Google Scholar↗

[63] van der Schaar, M., Alaa, A. M., Floto, A., Gimson, A., Scholtes, S., Wood, A., ... & Ercole, A. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Machine Learning*, *110*(1), 1-14. Google Scholar↗ CrossRef↗

[64] Arnold, M. H. (2021). Teasing out artificial intelligence in medicine: an ethical critique of artificial intelligence and machine learning in medicine. *Journal of bioethical inquiry*, *18*(1), 121-139. Google Scholar↗

[65] Kim, S. K., & Huh, J. H. (2021). Artificial intelligence based electronic healthcare solution. In *Advances in Computer Science and Ubiquitous Computing*, Springer, Singapore. 4(2), 575-581. Google Scholar↗

[66] Ali, F., El-Sappagh, S., Islam, S. R., Ali, A., Attique, M., Imran, M., & Kwak, K. S. (2021). An intelligent healthcare monitoring framework using wearable sensors and social networking data. *Future Generation Computer Systems*, *114*(1), 23-43. Google Scholar↗

[67] Ali, F., Islam, S. R., Kwak, D., Khan, P., Ullah, N., Yoo, S. J., & Kwak, K. S. (2018). Type-2 fuzzy ontology–aided recommendation systems for IoT–based healthcare. *Computer Communications*, *119*(1), 138-155. Google Scholar↗

[68] Simi, M. S., Nayaki, K. S., Parameswaran, M., & Sivadasan, S. (2017, October). Exploring female infertility using predictive analytic. In *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, IEEE. *115(7)*, 1-6. Google Scholar↗

[69] Golpour, P., Ghayour-Mobarhan, M., Saki, A., Esmaily, H., Taghipour, A., Tajfard, M., ... & Ferns, G. A. (2020). Comparison of support vector machine, naïve Bayes and logistic regression for assessing the necessity for coronary angiography. *International journal of environmental research and public health*, *17*(18), 6449, 1-9. Google Scholar↗

[70] Valupadasu, R., & Chunduri, B. R. R. (2019, May). Automatic classification of cardiac disorders using MLP algorithm. In *2019 Prognostics and System Health Management Conference (PHM-Paris)*, IEEE. *1*(2), 253-257. Google Scholar↗

[71] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, *17*(8), 43-48. Google Scholar↗

[72] Pethunachiyar, G. A. (2020, January). Classification of diabetes patients using kernel based support vector machines. In *2020 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE. *6*(1), 1-4. Google Scholar↗

[73] Maniruzzaman, M., Rahman, M., Ahammed, B., & Abedin, M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. *Health information science and systems*, *8*(1), 1-14. Google Scholar↗

[74] Ahuja, R., Sharma, S. C., & Ali, M. (2019). A diabetic disease prediction model based on classification algorithms. *Annals of Emerging Technologies in Computing (AETiC)*, *3*(3), 2947-2954. Google Scholar↗

[75] Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. *Journal of healthcare engineering*, *3*(2), 1-4. Google Scholar↗

[76] Gupta, S., Verma, H. K., & Bhardwaj, D. (2021). Classification of diabetes using Naive Bayes and support vector machine as a technique. In *Operations Management and Systems Engineering*, Springer, Singapore. *2*(1), 365-376. Google Scholar↗

[77] Choubey, D. K., Kumar, M., Shukla, V., Tripathi, S., & Dhandhania, V. K. (2020). Comparative analysis of classification methods with PCA and LDA for diabetes. *Current diabetes reviews*, *16*(8), 833-850. Google Scholar↗

[78] Sai, P. M. S., & Anuradha, G. (2020, March). Survey on Type 2 diabetes prediction using machine learning. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE. *1*(5), 770-775. Google Scholar↗

[79] Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science*, *82(1)*, 115-121. Google Scholar↗

[80] Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, *132(1)*, 1578-1585. Google Scholar↗ CrossRef↗

[81] Feshki, M. G., & Shijani, O. S. (2016, April). Improving the heart disease diagnosis by evolutionary algorithm of PSO and Feed Forward Neural Network. In *2016 Artificial Intelligence and Robotics (IRANOPEN)* (pp. 48-53). IEEE. Google Scholar↗

[82] Kannan, R., & Vasanthi, V. (2019). Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease. In *Soft computing and medical bioinformatics*, Springer, Singapore, *1*(2), 63-72. Google Scholar↗

[83] Maji, S., & Arora, S. (2019). Decision tree algorithms for prediction of heart disease. In *Information and communication technology for competitive strategies*, Springer, Singapore, *5(2)*, 447-454. Google Scholar↗

[84] Razavian, N., Blecker, S., Schmidt, A. M., Smith-McLallen, A., Nigam, S., & Sontag, D. (2015). Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. *Big Data*, *3*(4), 277-287. Google Scholar↗

[85] Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, *36(1)*, 82-93. Google Scholar↗

[86] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, *5*(1), 225-235. Google Scholar↗

[87] Dwivedi, A. K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing and Applications*, *29*(10), 685-693. Google Scholar↗

[88] Mathan, K., Kumar, P. M., Panchatcharam, P., Manogaran, G., & Varadharajan, R. (2018). A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease. *Design automation for embedded systems*, *22*(3), 225-242. Google Scholar↗

[89] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., & San Tan, R. (2017). A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine*, *89(1)*, 389-396. Google Scholar↗

[90] Vivekanandan, T., & Iyengar, N. C. S. N. (2017). Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. *Computers in biology and medicine*, *90(1)*, 125-136. Google Scholar↗

[91] Nilashi, M., bin Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An analytical method for diseases prediction using machine learning techniques. *Computers & Chemical Engineering*, *106(1)*, 212-223. Google Scholar↗

[92] Akben, S. B. (2018). Early stage chronic kidney disease diagnosis by applying data mining methods to urinalysis, blood analysis and disease history. *IRBM*, *39*(5), 353-358. Google Scholar↗

[93] Abdar, M. (2015) Using Decision Trees in Data Mining for Predicting Factors Influencing of Heart Disease. *Carpathian Journal of Electronic & Computer Engineering*, *8*(2), 31-36. Google Scholar↗

[94] Polat, H., Danaei Mehr, H., & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. *Journal of medical systems*, *41*(4), 1-11. Google Scholar↗

[95] Reddy, G. T., & Khare, N. (2017). An efficient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model. *Journal of Circuits, Systems and Computers*, *26*(04), 1750061; 1-21. Google Scholar↗

[96] Pang, Z., Zhu, D., Chen, D., Li, L., & Shao, Y. (2015). A computer-aided diagnosis system for dynamic contrast-enhanced MR images based on level set segmentation and ReliefF feature selection. *Comput. Math. Methods Medicine*, *2014*(1), 450531, 1-10. Google Scholar↗

[97] Vijayarani, S., & Dhayanand, S. (2015). Data mining classification algorithms for kidney disease prediction. *Int J Cybernetics Inform*, *4*(4), 13-25. Google Scholar

[98] Gürbüz, E., & Kılıç, E. (2014). A new adaptive support vector machine for diagnosis of diseases. *Expert Systems*, *31*(5), 389-397. Google Scholar↗

[99] Lu, C., Zhu, Z., & Gu, X. (2014). An intelligent system for lung cancer diagnosis using a new genetic algorithm based feature selection method. *Journal of medical systems*, *38*(9), 1-9. Google Scholar↗

[100] Uzer, M. S., Inan, O., & Yılmaz, N. (2013). A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA. *Neural Computing and Applications*, *23*(3), 719-728. Google Scholar↗

[101] Chen, H. L., Liu, D. Y., Yang, B., Liu, J., & Wang, G. (2011). A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis. *Expert systems with applications*, *38*(9), 11796-11803. Google Scholar↗

[102] Polat, K., & Güneş, S. (2007). Automatic determination of diseases related to lymph system from lymphography data using principles component analysis (PCA), fuzzy weighting pre-processing and ANFIS. *Expert Systems with Applications*, *33*(3), 636-641. Google Scholar↗

****