Machine Learning Algorithms in Healthcare: A Literature Survey

Conference Paper · July 2020

DOI: 10.1109/ICCCNT49239.2020.9225642

3 authors:

Munira Ferdous
Daffodil International University
2 PUBLICATIONS **45** CITATIONS
SEE PROFILE

Jui Debnath
Daffodil International University
2 PUBLICATIONS **45** CITATIONS
SEE PROFILE

Narayan Ranjan Chakraborty
Universitetet i Agder
38 PUBLICATIONS **263** CITATIONS
SEE PROFILE

Machine Learning Algorithms in Healthcare: A Literature Survey

Munira Ferdous*, Jui Debnath† and Narayan Ranjan Chakraborty‡
Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
Email: *munira15-9772@diu.edu.bd, †jui15-9712@diu.edu.bd, ‡narayan@daffodilvarsity.edu.bd

***Abstract*—Machine learning algorithms construct a remarkable contribution to predicting diseases. The generic purpose of this work is to help the researchers and practitioners to choose appropriate machine learning algorithm in health care. Previous research has shown that machine learning algorithms provide the best accuracy in diagnosing diseases but the accuracy of the algorithms and other related issues are hardly available in one complete paper. The necessary information has to be found in separate articles which is most frequently time-consuming and tedious. So, the objective of this work is to provide all the necessary information about the machine learning algorithms used in the healthcare sector. We generated a data table about machine learning algorithms accuracy for different diseases from the literature then finished this process step by step and systematized this survey paper. The output of this work produces a list of best machine learning algorithms with accuracy for predicting diseases. This output will help the researcher and practitioner to know about the contribution of machine learning algorithms in the field of health care with the accuracy of algorithms together in one complete paper.**

***Index Terms*—Healthcare, Machine learning, Algorithm, Data mining, Diseases**

I. Introduction

Science has presented human life much more accessible through technology. In this twentieth era, the entire world is in the time of big data. Machine learning can predict future data or can offer a desirable decision from a big dataset. It is commonly capable of guiding machines to perform tasks by regarding measures of pattern, whereby they should be done [1]. Undoubtedly, machine learning is an enormously realistic technological application that operates with a big dataset [2]. The contribution of machine learning in medical science is undeniable in many fields. Although, machine learning has developed many algorithms. Machine learning algorithms have succeeded in many oppressive things worldwide. Machine learning algorithms are generally mixed up with mathematics and logic that can easily predict from a given dataset. Machine learning algorithms are very accessible to diagnose complex diseases by applying machine learning algorithms. As there's loads of research performed about machine learning algorithms in healthcare for predicting diseases. All over the world thousands of researchers have discovered a lot in predicting diagnosis applying machine learning algorithms. Since there is a lot of information scattered in many papers, whenever some researchers aspire to recognize about machine learning algorithms, what is the predicting accuracy of diagnosis and which algorithms are the best in all the papers, usually, they get exhausted of looking for papers. Since there are lots of diseases and lots of algorithms, it's extremely challenging to find out of the best algorithm for diagnosis most of the time they could not figure out of the exact paper. This paper purposed a literature survey with various machine learning algorithms predicting diseases with accuracy is explained by different papers information in one, a paper by using commonly used machine learning algorithms. This paper extensively presents as much information as possible can get, those can be kept. This work also focused to analyze the best algorithms which give the most reliable accuracy for the prediction of any disease based on the study from existing literature. This work will encourage practitioners and researchers to find the information easily while working in the healthcare sector.

The remainder of the article is structured as follows. In the next section, the related work is described, followed by research methodology and data analysis. Then, we discuss our work before we conclude.

II. Related Work

In previous work, we have seen machine learning algorithms can predict disease quickly. But many of the paper has researched some specific diseases and after doing implementing machine learning algorithms, they shown machine learning algorithms can predict diseases. A study paper has shown the machine learning algorithms definition, and brief about some machine learning algorithms and one of the top algorithms has also shown for a specific disease [2].

Another study performing analysis of a machine learning algorithm on diabetes, they used a big dataset and apply some machine learning algorithm for predicting diabetes and show the best algorithms for diabetics prediction [3]. Another study surveyed the prediction of diseases using a machine learning algorithm, they have shown the best algorithms [4]. Another study shows the best algorithms for liver diseases prediction [5].

A study created a dataset and from that, they applying some approach and method and predict breast cancer also shows which algorithms can predict breast cancer, they just showed the prediction for breast cancer [6]. One of the studies Classifying Ear Disorders and they only Using Support Vector

Machines algorithms, they showed only one algorithm there [7].
Different papers have predicted for a specific disease. We studied all the details and taken the best algorithm showed the best for the disease and make a dataset.
It is difficult to find out all the topmost algorithm in one compilation. So, we have decided to do a survey paper where all the topmost algorithms have organized. In the future, if someone wants to do research machine learning algorithms in the field of healthcare they can get all the information together which saves significant research time.

III. Research Method

This paper displays the participation of machine learning algorithms in the healthcare sector. To achieve the research objectives, we followed the following steps.

A. Initial studies selection

At first, based on the search key we picked some paper as an initial study. The study of research results that would assist to attain the ends. The platforms searched for the title, abstract, index key, is:
• IEEE Xplore digital library
• Google Scholar

B. Inclusion and Exclusion criteria

Secondly, We created some meaningful search key for machine learning algorithm prediction only in healthcare, got 4123 papers in total. The maximum of those papers does not summarize the aim of this research paper. Therefore, some exclusion and inclusion criteria were needed.
• Criteria for Exclusion: Didn't match with our research objective. Paper focusing or economic or any other section.
• Criteria for Inclusion: We also found a few pieces of literature that did not come up with the search string but we must add those papers in the study. We used a snowball search strategy to add those articles to the repository.
After finishing the exclusion and inclusion part we got 357 papers in total that summarized resembling our research and found a possible maximum paper that contains proper information of the machine learning algorithm's future prediction.

Thirdly, From the selecting 357 papers, we categorized data into some parts for finding quality assessment of the research paper's Context, Qualitative and Quantitative data. These data help in finding out related research papers. Then we got 174 papers for final study after completing all inquiries.

Fourthly, after checking the quality assessment of the research paper we took 28 papers. From these 28 papers, we have arranged the paper and find out the best machine learning algorithm gives the best accuracy for diagnosis.

IV. Data Analysis

To do the analysis, a data-set has been created with machine learning algorithms accuracy for diagnosis. From the data-set same tables have been made for each disease where the machine learning algorithm's accuracy for diagnosis is available. So that researchers can easily find out which algorithms

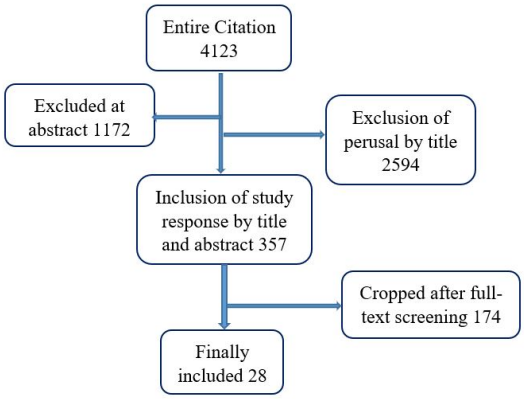

Fig. 1: The process of paper selection

are best for diagnosis. Their accuracy has been described consecutively. Total of 13 diseases and 17 algorithms found in that data-set. Each paper has different accuracy through different tests but we have taken the best from there. For each disease, a separate table is drawn with the topmost algorithms that have provided the best accuracy.

Presented in table 1, the best algorithm is Logistic Regression as per the accuracy provided in the literature. The article described that it prepared and pre-processed data for applying three algorithms to predict Liver disease and get the best algorithm is Logistic Regression with 95.8 percent accuracy [5]. The logistic regression algorithm is used for a classification method based on the probability concept and the logic model is mapped between 0 and 1 [8].

Table 1: Liver Disease

SL	Algorithm	Accuracy (%)	Year
1	Logistic Regression [5]	95.8	2018
2	Decision Tree [5]	94.9	2018
3	Fuzzy Neutral Network [9]	91	2017
4	Support Vector Machine [5]	82.7	2018
5	Support Vector Machine [10]	71.97	2018

Table 2 exhibits, for predicting diabetics a list of algorithms with accuracy. A study has utilized various algorithms for prediction Diabetics, they applied Random forest, support vector machine, and Linear discriminant algorithm considering Kappa metrics, Sensitivity, and Specificity metrics [3]. From the inquiry, we got one of the algorithms at first which accuracy is best among all implemented algorithms. This paper gets the best algorithm for predicting diabetes is the Random forest algorithm with 100 percent accuracy. Random forest generally makes decision trees for every attribute with avoiding overfitting of training data also it is conducted for both classification algorithms and regression algorithms [11]. Another study uses the classification for applied data and gets 98 percent accuracy from the Random forest algorithm [12].

Table 3 explains, all the best-performed for prediction and detection of Breast Cancer. In 2018, a research paper calculated 99 percent accuracy bu using the Support Vector

Table 2: Diabetes Mellitus

SL	Algorithm	Accuracy (%)	Year
1	Random forest [3]	100	2017
2	Random forest [12]	98	2018
3	Support Vector Machine [3]	96.9	2017
4	Decision Tree [13]	96.5	2020
5	Naive Bayes [2]	95	2018
6	Linear discriminant analysis [3]	94.5	2017

Machine algorithm [4]. In 2010 another study showed that Support vector machine obtained to provide an accuracy of 98.80 percent for predicting breast cancer [6].

Table 3: Breast Cancer

SL	Algorithm	Accuracy (%)	Year
1	Support Vector Machine [4]	99	2018
2	Support Vector Machine [6]	98.8	2010
3	Adaptive Boosting [14]	98.5	2018
4	Artificial Neural Network [15]	98.5	2018
5	Decision Tree [2]	98.14	2018
6	Artificial Hydrocarbon Networks [16]	97.8	2017

Table 4 shows, the best four algorithms accuracy for Sickle Cell Disease. The best algorithm is Sequential Minimal Optimization (SMO) for predicting Sickle Cell Disease. For the implementation of these four algorithms, the MATLAB tool with SVMLAB has been used and receives the best accuracy for SMO, Rules JRIP, Decision Tree, and Naïve Bayes algorithms [17].

Table 4: Sickle Cell Disease

SL	Algorithm	Accuracy (%)	Year
1	SMO [17]	99.5	2017
2	Rules JRIP [17]	95.9	2017
3	Decision Tree [17]	93.1	2017
4	Naive Bays [17]	70.6	2017

A study was proposed a genetic algorithm based trained recurrent Fuzzy Neural Network accuracy and get the best accuracy is 97.78 percent from the testing set [18]. Table 5 shows, the best algorithms for predicting heart diseases. Artificial Neural Network accuracy is 97.5 percent which is the most desirable machine learning algorithm for predicting heart disease [19]. The decision tree algorithm shows 91.48 percent accuracy [13]. The different paper shows different accuracy but we took the best accuracy for heart disease prediction. A decision tree is a mostly used algorithm in the machine learning field used for classification purposes, every tree contains an internal node, one leaf node, the root node, and branches [2] and the decision tree doesn't require any parameters because it's a structure very easy to follow.

Table 6 exhibits, In 2019, a study applied algorithm (ANN, Random forest, LR, SVM, Decision tree) on trained data for predict cardiac arrest and get the best accuracy is 85.00 percent in the Artificial Neural Network algorithm [8]. Artificial neural networks usually used to Classify input data and get these all

Table 5: Heart Disease

SL	Algorithm	Accuracy (%)	Year
1	Fuzzy Neural Networks [18]	97.7	2017
2	Artificial Neural Network [19]	97.5	2018
3	Support Vector Machine [20]	94.6	2012
4	Decision Tree [13]	91.4	2020
5	Naïve Bayes [13]	87.19	2020

Table 6: Cardiac Arrest

SL	Algorithm	Accuracy (%)	Year
1	Artificial Neural Network [8]	85.0	2019
2	Random Forest [8]	59.4	2019
3	Logistic Regression [8]	56.31	2019
4	Support Vector Machine [8]	53.8	2019
5	Decision Tree [8]	49.0	2019

into the desired output [21]. Decision Tree algorithms give less accuracy among all five for predicting Cardiac arrest. The second best-given accuracy for predicting Cardiac arrest is the Random Forest algorithm. Random Forest generally makes decision trees for every attribute with avoiding overfitting of training data also it is conducted for both classification algorithms and regression algorithms [11].

Table 7: Thyroid Disease

SL	Algorithm	Accuracy (%)	Year
1	Support Vector Machine [2]	98.6	2018
2	Support Vector Machine [22]	92.9	2020
3	Random Forest [22]	56.31	2020
4	Naive Bayes [22]	74.3	2020

Table 7 shows, the Support Vector Machine algorithm gives the best accuracy for predicting Thyroid Disease [2].

Table 8: Alzheimer's disease

SL	Algorithm	Accuracy (%)	Year
1	Neural Network [23]	98.3	2018
2	Random Forest [23]	97.8	2018
3	Support Vector Machine [23]	97.5	2018
4	Gradient Boosting [23]	97.2	2018
5	K-Nearest Neighbor [23]	95.0	2018

Table 8 describes the top five best algorithms for predicting Alzheimer's disease. The author's of one of the article created a data-set for finding out, at first, who have Alzheimer Diseases and who have not, [23] then and applied algorithms (NN, Random Forest, SVM, Gradient Boosting, K-NN) get the best accuracy for predict Alzheimer disease is 98.36 percent. Many modifications have been shown in Neural Networks and for training intent, it provided a set of data links and classes [24].

In Table 9, A study explained after training the data-set they applied Naïve Bayes and Artificial Neural Network to get 100 percent from the Naïve Byes machine learning algorithm [25].

A research paper described two different machine learning algorithms, they applied from the same data-set for predicting

Table 9: Chronic Kidney Disease

SL	Algorithm	Accuracy (%)	Year
1	Naive Bayes [25]	100.0	2016
2	Artificial neural network [25]	72.3	2016

Ear disorder, they got different accuracy for SVM and MPNN algorithm and the best accuracy 92.5 percent has Support vector machine [7]. This phenomenon is presented in table 10.

Table 10: Ear Disorder

SL	Algorithm	Accuracy (%)	Year
1	Support vector machine [7]	92.5	2010
2	MPNN [7]	77.5	2010

In table 11, Our investigation determined simply one of the algorithm's accuracy for predicting Eczema. A study did show their research goals remained to improve eczema detection models and they used a novel deep learning approach that uses an ensemble of the AdaBoost algorithm and checked whether the disease is eczema or not and achieved an accuracy of 97.5 percent [26].

Table 11: Eczema

SL	Algorithm	Accuracy (%)	Year
1	AdaBoost [26]	97.5	2019

Table 12 shows, Convolutional Neural Network algorithms for General Disease prediction. A study created a data-set from patient's records and utilized KNN and CNN to classify patient data, compare these two algorithms and get the CNN algorithm to give the best accuracy for predicting General Disease Prediction [27].

Table 12: General Disease Prediction

SL	Algorithm	Accuracy (%)	Year
1	Convolutional neural network [27]	84.5	2019

In 2019, a research paper showed for predicting Chronic disease they categorized into two parts structured and unstructured data and using evolution method accuracy and get the best accuracy for the Recurrent Neural Network algorithm [28]. Table 13 presented this analysis.

Table 13: Chronic Disease

SL	Algorithm	Accuracy (%)	Year
1	Recurrent Neural Network [28]	97.6	2019

V. Discussion

The purpose of this study was to explore the best machine learning algorithms based on the accuracy used in health care research.

According to our inquiry, it is observed and obvious that many researchers are inspired to research machine learning algorithms in the health care sector. But it is always difficult to choose the best algorithm to predict the disease based on the data-set created by the researcher. Sometimes researchers wasted the time and resources by choosing the incorrect algorithm. Some researches have already been done to identify machine algorithms based on health care data. But often sometimes all information is not presented in one single article. Due to that, new researchers are usually become confused and put a lot of efforts to identify the information from the literature. This research is exactly focused on this point. So, our main intention is to bring all information in one single paper. For this, we have designed a data-set of a different relevant research paper, regarding these, those research paper have already composed about machine learning algorithms accuracy prediction for diagnosis. Figure 2 provided a graphical demonstration of all machine learning algorithms based on their accuracy level.

Next from the data-set, we have arranged a list of the most applicable and providing the best accuracy in machine learning algorithms. Our work delivers topmost algorithms of machine learning which have the best accuracy. If someone fancies knowing concerning the best machine learning algorithm for the diagnosis, they can quickly find out the most desirable algorithm. This work has analyzed about 14 diseases. For each disease, there is a different table individually shows the best algorithm list for prediction disease as if researchers haven't faced any difficulty when they search for something about machine learning algorithms in healthcare. The main objective of this work is to help researches to find out the participation of machine learning algorithms in healthcare knowledge.

VI. Conclusions and Limitations

To facilitate practitioner and machine learning researchers we have tried to create an organized data-set from all of the study which contains machine learning algorithms that contributed to healthcare. We have found a list of best algorithms from all the papers we read. This work is not biased towards the methodology presented in the reviewed literature. We only focused on the algorithms used and their accuracy level. We believe that the outcome of this study will reduce the fatigue of the practitioner and researcher to choose the best algorithm while predicting the diseases we discussed here.

A common point is that the application of machine learning at the first machine has to learn the data then give a prediction. Usually, it depends on how the machine is being taught. Since we have not implemented any application in our work and we have shown the most desirable accuracy for diagnosis from the data-set which implemented by another research work. We didn't get 100 percent accuracy for all of the diagnosis. This is one of the deficiencies in our work. Another limitation is database selection. We only focused on google scholar and IEEE explore database to get the relevant literature. Choosing a few other databases may produce the diversity of the results we presented here.

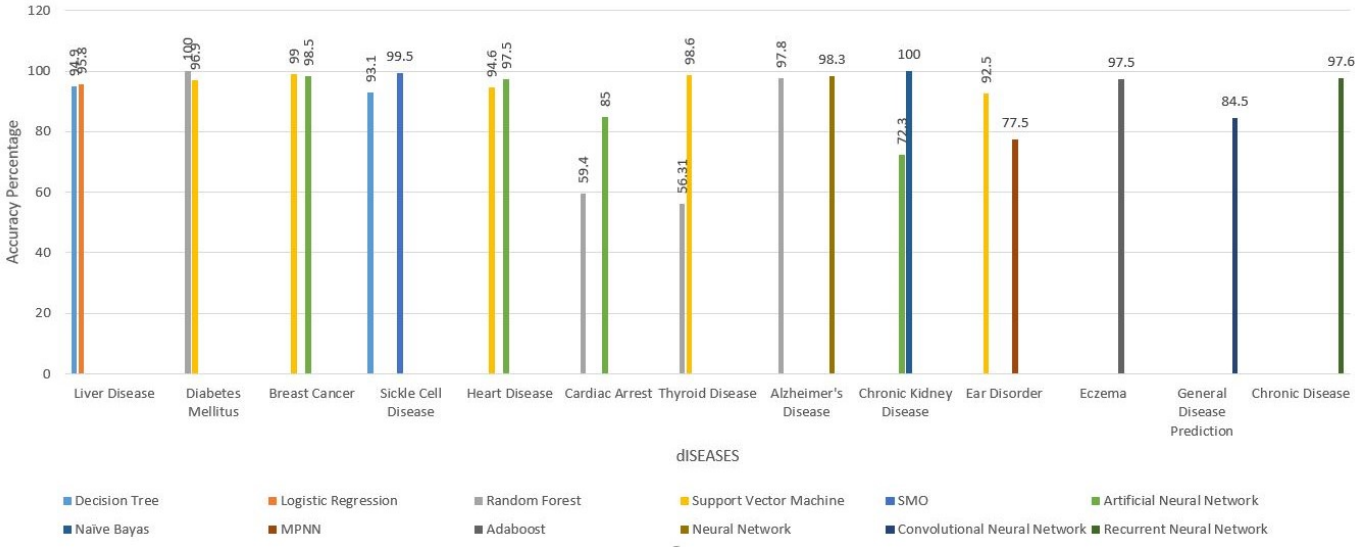

Fig. 2: Accuracy of Machine learning algorithms in various diseases

REFERENCES

[1] B Dhomse Kanchan and M Mahale Kishor. Study of machine learning algorithms for special disease prediction using principal of component analysis. In *2016 international conference on global trends in signal processing, information computing and communication (ICGTSPICC)*, pages 5–10. IEEE, 2016.

[2] K Shailaja, B Seetharamulu, and MA Jabbar. Machine learning in healthcare: A review. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 910–914. IEEE, 2018.

[3] P Suresh Kumar and S Pranavi. Performance analysis of machine learning algorithms on diabetes dataset using big data analytics. In *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS)*, pages 508–513. IEEE, 2017.

[4] N Marline Joys Kumari and Kishore KV Krishna. Prognosis of diseases using machine learning algorithms: A survey. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pages 1–9. IEEE, 2018.

[5] Vyshali J Gogi and MN Vijayalakshmi. Prognosis of liver disease: Using machine learning algorithms. In *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)*, pages 875–879. IEEE, 2018.

[6] Alireza Osareh and Bita Shadgar. Machine learning techniques to diagnose breast cancer. In *2010 5th international symposium on health informatics and bioinformatics*, pages 114–120. IEEE, 2010.

[7] Mahsa Moein, Mohammad Davarpanah, M Ali Montazeri, and Mehrnaz Ataei. Classifying ear disorders using support vector machines. In *2010 Second International Conference on Computational Intelligence and Natural Computing*, volume 1, pages 321–324. IEEE, 2010.

[8] Utsav Chauhan, Vikas Kumar, Vipul Chauhan, Sumit Tiwary, and Amit Kumar. Cardiac arrest prediction using machine learning algorithms. In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, volume 1, pages 886–890. IEEE, 2019.

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

[10] Ajay S Singh, Md Irfan, Abhishek Chowdhury, et al. Prediction of liver disease using classification algorithms. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pages 1–3. IEEE, 2018.

[11] Kumar G Dinesh, K Arumugaraj, Kumar D Santhosh, and V Mareeswari. Prediction of cardiovascular disease using machine learning algorithms. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pages 1–7. IEEE, 2018.

[12] Shweta Ganiger and KMM Rajashekharaiah. Chronic diseases diagnosis using machine learning. In *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, pages 1–6. IEEE, 2018.

[13] P. S. Mung and S. Phyu. Effective analytics on healthcare big data using ensemble learning. In *2020 IEEE Conference on Computer Applications(ICCA)*, pages 1–4, 2020.

[14] Pahulpreet Singh Kohli and Shriya Arora. Application of machine learning in disease prediction. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pages 1–4. IEEE, 2018.

[15] Naresh Khuriwal and Nidhi Mishra. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In *2018 IEEMA Engineer Infinite Conference (eTechNxT)*, pages 1–5. IEEE, 2018.

[16] Hiram Ponce and Ma de Lourdes Martinez-Villaseñor. Interpretability of artificial hydrocarbon networks for breast cancer classification. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3535–3542. IEEE, 2017.

[17] Dhafar Hamed Abd and Intisar Shadeed Al-Mejibli. Monitoring system for sickle cell disease patients by using supervised machine learning. In *2017 Second Al-Sadiq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)*, pages 119–124. IEEE, 2017.

[18] Kaan Uyar and Ahmet İlhan. Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. *Procedia computer science*, 120:588–593, 2017.

[19] Debabrata Swain, Santosh Kumar Pani, and Debabala Swain. A metaphoric investigation on prediction of heart disease using machine learning. In *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)*, pages 1–6. IEEE, 2018.

[20] G Parthiban and SK Srivatsa. Applying machine learning methods in diagnosing heart disease for diabetic patients. *International Journal of Applied Information Systems (IJAIS)*, 3(7), 2012.

[21] Md Razu Ahmed, SM Hasan Mahmud, Md Altab Hossin, Hosney Jahan, and Sheak Rashed Haider Noori. A cloud based four-tier architecture for early detection of heart disease with machine learning algorithms. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pages 1951–1955. IEEE, 2018.

[22] Priyanka Duggal and Shipra Shukla. Prediction of thyroid disorders using advanced machine learning techniques. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 670–675. IEEE, 2020.

[23] Priyanka Lodha, Ajay Talele, and Kishori Degaonkar. Diagnosis of

alzheimer's disease using machine learning. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–4. IEEE, 2018.

[24] Dharavath Ramesh, Pranshu Suraj, and Lokendra Saini. Big data analytics in healthcare: A survey approach. In *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, pages 1–6. IEEE, 2016.

[25] Veenita Kunwar, Khushboo Chandel, A Sai Sabitha, and Abhay Bansal. Chronic kidney disease analysis using data mining classification techniques. In *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*, pages 300–305. IEEE, 2016.

[26] Yash Kumar Arora, Amish Tandon, and Rahul Nijhawan. Hybrid computational intelligence technique: Eczema detection. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pages 2472–2474. IEEE, 2019.

[27] Dhiraj Dahiwade, Gajanan Patle, and Ektaa Meshram. Designing disease prediction model using machine learning approach. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1211–1215. IEEE, 2019.

[28] Anandajayam P, Aravindkumar S, Arun P, and Ajith A. Prediction of chronic disease by machine learning. pages 1–6, 03 2019.