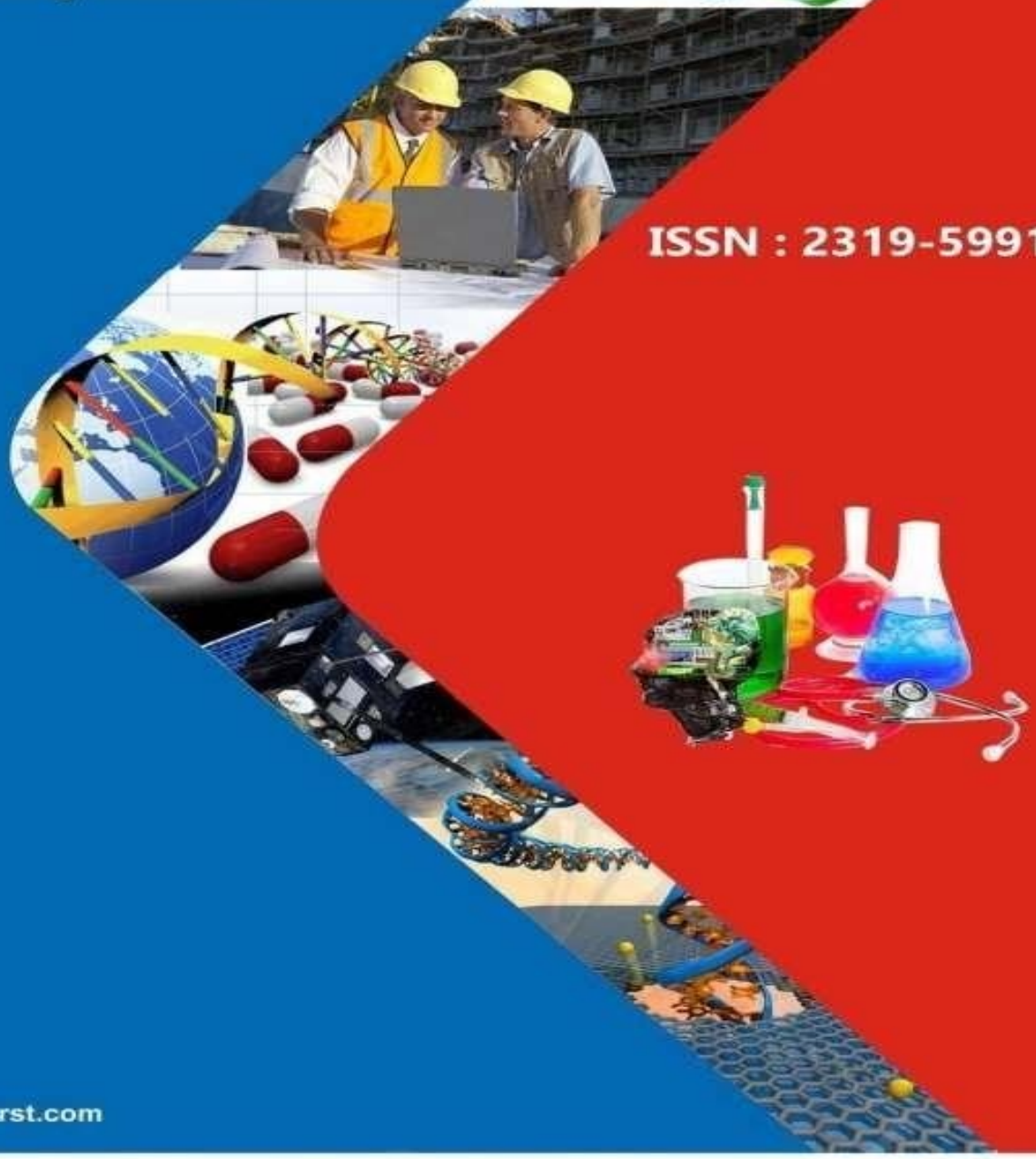
IJERST

International Journal of Engineering Research and Science & Technology



ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

A Comprehensive Review of Machine Learning Methods for Healthcare-Related Illness Prognosis

[1]Dr.Kamel Alikhan siddique [2]MOHAMMAD SAMEER, [3]MOHAMMED GAYAZ, [4]MOHAMMED MAJEED,
[1]Associate Professor, Department of AIML, Lords Institute of Engineering & Technology.
[234] Student, Department of AIML, Lords Institute of Engineering & Technology.

Abstract–

There has been an explosion of healthcare data in the 21st century, along with an increase in chronic illnesses and other health-related issues. Conventional data management solutions struggle to keep up with the sheer volume of healthcare data. The need for intelligent decision support systems that enable doctors to make better decisions via the use of various algorithms brings machine learning into play. This paper summarizes the many kinds of healthcare data and analyses the results of many studies according to how well machine learning approaches forecast illnesses and identify high-risk diseases early on. The most often mentioned Machine Learning techniques were Support Vector Machine (SVM), Random Forest, Decision Trees (DT), K Nearest Neighbor (KNN), Deep Learning, Naive Bayes (NB), Artificial Neural Network (ANN), and Logistic Regression. Primarily considered are algorithms, datasets, accuracy, and knowledge gaps in the field.

Index Terms –Wellness program, Machine Learning, Disease Prediction, Heterogeneous data.

INTRODUCTION

As John Naisbitt put it in Megatrends, the world is "drowning in data, yet famished for understanding," and the quantity of healthcare data has grown rapidly from terabytes to petabytes. The healthcare industry makes use of a wide variety of data sources. When it comes to medical services and apps, one of the biggest challenges is processing massive volumes of data using traditional techniques. In order to discover treatments, specialist medical care, and other services, data is collected from a wide variety of sources and processed in various ways [5]. The field of artificial intelligence (AI) makes use of machine learning (ML) to aid in the early diagnosis of several illnesses. Since machine learning has the potential to drastically cut costs and medical visits, it also enables treatment in distant regions. Using ML-based predictive analytics and effective learning algorithms such as KNN, SVM, Random Forest, Decision Tree, Deep Learning, ANN, Logistic Regression, Naive Bayes, etc., the medical home network prioritizes care management and outreach to patients who are more susceptible to infections [27]. Advanced data analytics tools, Address: Gharuan, Mohali, Chandigarh University Yogesh Banyal Department of Computer Science Machine learning applications in healthcare and clinical decision-making have been developed at a rapid pace due to factors such as open access efforts, healthcare digitalization, and computing. Machine Learning has become the go-to tool for complex analysis, smart decision-making, and innovative problem-solving in the healthcare industry. We live in an algorithmic age, and machine learning algorithms are essential for solving difficult problems in healthcare. These algorithms have shown remarkable results in areas such as segmenting brain tumors [26], identifying organs from medical images [2], reconstructing medical images [1], and classifying interstitial lung illnesses [3]. In addition to identifying cancer cells, diabetes, heart illness, Alzheimer's disease, mental tiredness, stress management, etc., these algorithms currently outperform radiologists in detecting malignant tumors. To improve the quality of therapy, machine learning in healthcare utilizes external data analysis tools like X-rays, CT scans, and other tests and screenings. These tools employ self-learning neural networks. Looking for patterns in the data enables us to make future judgments based on the observations. Examples, real-world experience, and training are all viable avenues for gathering data observations. Using these patterns, we may better understand the present and plan for the future, for instance, by determining who is most likely to get the virus and what variables put them at risk of infection [3]. The purpose of this data is to "train" models. Machine learning is most effective when large amounts of "training" data are accessible and there are many patterns to learn from [3]. Accessible huge databases include picture databases such as "ImageNet," which has around fourteen million images. These databases provide both a source of training data and a benchmark for

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

academics to assess the efficacy of their proposed algorithms. This paper will summarize the ways in which healthcare systems use ML algorithms, specifically how ML approaches help with symptom diagnosis and early detection of chronic illnesses. The methods that were most heavily emphasized were those that include deep learning, decision trees, artificial neural networks, support vector machines, random forests, and logistic regression machine learning.

TYPES OF HEALTHCARE DATA

In healthcare, there are several places to find data. Several data mining methods are used to extract the most important elements from this data, and in order to make accurate predictions in the future, algorithms need to be trained. Below, we will go over the many types of healthcare data that ML algorithms use for early illness identification. The vast amount of healthcare-related electronic data is a consequence of the many resources shown in Fig. 1. 1.Clinical Data: Information collected from a patient's Electronic Health Records (EHR) over the course of their treatment, including results from laboratory testing, radiologic imaging, allergies, and other diagnostic procedures. The likelihood of survival for breast cancer patients may be estimated with the use of ML algorithms that have been created [2] and applied to data obtained from Multiparametric Magnetic Resonance Imaging (mpMRI). 2. Sensor Data: Machine learning techniques have been developed to forecast the survival rates of breast cancer patients using data collected from multiparametric magnetic resonance imaging (mpMRI) [2]. Whether it's more complex data or just numerical and category values, computing tools process these data bits. Researchers studying Parkinson's disease (PD) use machine learning algorithms to discover streams of data from wearable sensors. Thirdly, omics data is complicated and high dimensional; it mostly consists of genomic, transcriptomic, and proteomic information. Genomic data is a collection of information on DNA sequence, copy number variation, gene expression, and other related topics; it is used in bioinformatics [2]. 5. Transcriptomic data: This refers to the data compiled from the many mRNA transcripts found in a biological sample [2]. The purpose of analysing and extracting these samples is to build different databases. 6. Genomic information: A proteome is a catalog of all the proteins that an organism, cell, or tissue has expressed. Within the cell, it depicts actual, practical molecules.



Fig. 1: An image of the diverse sources that contribute to healthcare data [26]

LITERATURE SURVEY

Several articles have shown the use of various Machine Learning methods in healthcare for illness prediction. For the purpose of early illness prediction, many hybrid techniques are used. Below is a survey that summarizes the many publications that were reviewed in order to determine which ML approaches are utilized for illness prediction [12] [21]. Also specified are the rates of accuracy attained by various ML algorithms. Here are some examples of the many ML approaches used for chronic illness prediction, as shown in Table 1. To aid in future studies of these disorders, we have also determined the research gaps of all the articles that were considered. Disease Prediction Using Machine Learning Methods 1. Heart Disease Detection One common use of ML methods is the very accurate diagnosis of cardiac illness. With the help of the UCI heart disease patient repository's Cleveland database, the dataset was created. Using Decision Tree, KNN, Naive Bayes, and Random Forest, Devansh Shah et. al. [13] sought to predict the probability of cardiac disease using classification techniques. While Random Forest achieved 91.6% accuracy, Naïve Bayes achieved 83.49%, Decision Tree achieved 71.43%, and KNN achieved 83.16%. 2. Examining Conditions Related to Diabetes This takes into account sets found in the UCI Machine

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Learning Repository. Machine learning algorithms have been used in several research to predict the condition of diabetes, as stated in article [1]. Among the algorithms used are Naive Bayes (79.56% accuracy), Decision Tree (76.95%), and Logiboost (77.47%), all of which are part of the CART framework. Breast cancer disclosure using machine learning Among female malignancies, breast cancer ranks high in incidence and mortality rates in both Asia and the US. Breast cancer is predicted using machine learning methods. These datasets are sourced from the UCI machine learning repository and are based on data from Wisconsin. By using a j48, Naive Bayes model and the WEKA tool, Williams et al. [19] were able to detect breast cancer risks. In comparison to Naive Bayes' 82.6% accuracy, j48's 94.2% makes it the superior algorithm for predicting breast cancer. A decision tree with a 94% accuracy rate was discovered to be able to predict breast cancer using a common pattern [20]. 4. Recognizing a Thyroid Condition In order to diagnose thyroid issues, machine learning techniques are used. Data mining methods such as decision trees, SVMs, and fuzzy maps are used for classification purposes. The data sets originate from the UCI repository. In order to provide accurate predictions about thyroid illness, K. Shankar et al. [22] developed a multi-kernel support vector machine that used the IGWO computation to choose features optimally. 98.65% accuracy was achieved with MKSVM. 5. Use eye tracking to identify mental fatigue using machine learning techniques The relevance of mental fatigue has grown due to the correlation between it and declines in cognitive performance and health consequences, especially among the elderly. Using SVM, the authors of research [15] were able to identify mental fatigue in both younger and older participants in realistic viewing circumstances with a 91.0% accuracy rate.

TABLE 1: Summary of ML methodologies used in prediction of different chronic diseases

Sno.	Contribution	Dataset	Results	Research Gaps
1	Tuppad, A, et al.[27] in year 2022 states Clinical decision support for type 2 diabetes is acknowledged and identified. the medical knowledge, advice, and practise gaps in diabetes that machine learning (ML) can fill, -a review of current research on the use of machine learning in diabetes risk assessment, diagnosis, and prognosis.	The sharing of data is not relevant because no datasets were produced or examined for this subject.	The possible research gaps in terms of medicine and technology that need to be filled in order to develop reliable and a review of current research on the use of machine learning in diabetes risk assessment, diagnosis, and prognosis. applications in clinical decision support aimed at type 2 diabetes management and prevention.	It reflects holes in pre-diabetes diagnostic standards, dangers, and the potential for developing pre-diabetes risk screening tools — complications using ML to avoid T2D. Pre-diabetes screening and therapeutic recommendations for its diagnosis and treatment are challenging to comprehend.
2	Ahsan, M.M.et al. [28] reviewed the publications on illness diagnosis based on machine learning published between 2012 and 2021. (MLBDD).	The Parkinson, PIMA, and Cleveland Heart disease datasets are the most frequently utilised datasets in the field of illness diagnosis.	Due to its outstanding performance with both image and tabular data, CNN is one of the most sophisticated ML approaches.	No Research Gaps were found.
3	R. Bharti et al.[12] states comparison of several deep learning and machine learning techniques	Hungary, Cleveland, Long Beach V and Switzerland Public Health Dataset	Logistic regression - 83.3%, KNN-84.8% SVM - 83.2% Random Forest-80.3% Decision Tree- 82.3% Deep Learning-94.2%	ML techniques cannot be used on a large dataset. If dataset size is increased then DL will produce more promising results.
4	L. J. Muhammad et al.[7] explains the development of supervised machine learning models for COVID-19 logistic regression, decision trees, SVM, Nave Bayes, and ANN.	Epidemiological Surveillance Systems' Information on Viral Respiratory Diseases epidemiology dataset contains cases of COVID-19 from Mexico with both positive and negative test results.	Decision tree obtained highest accuracy-94.99%, Logistic Regression - 94.41, Naïve bayes - 94.36, SVM-92.4%, ANN-89.2%	No Research Gaps were found.
5	D. Shahet al. [13] tells about supervised learning techniques such as Random Forest	Cleveland database of UCI repository	Naïve Bayes provides 83.49% accuracy, Decision Tree	For early heart disease prediction to be more accurate, more complicated and

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

	heart disease.		Random Forest provided 91.6% accuracy.	
	J. Sneghaet al. [23] in year 2020 states that in order to distinguish between chronic kidney diseases, two data mining algorithms—Random Forest and Back Propagation Neural Network—have been contrasted.	Data Source: kaggle	For predictin g chronic renal illness, Back Propagati on Neural Network surpasses Random Forest Classifier . For predictin g chronic renal illness, Back Propagati on Neural Network surpasses Random Forest Classifier . yields a 98.40% level of assurance .	Different classifiers can be used for diagnoses of chronic kidney disease.
	L. J. Muhammad et al. [24] in year 2020 states that Data mining techniques were created with the goal of predicting patients with COVID-19 infection's prognosis for recovery.	Available on Kaggle is epidemiologic al data on South Korean COVID-19 patients that was gathered from the Korea Centers for Disease Control and Prevention (KCDC).	The algorithm s SVM, DT, LR, NB, K-NN, and RF were used. and DT gave highest accuracy i.e., 99.85%.	The number of attributes (here 5) can be increased and can be used with different models for prediction.
8	Y. Yamada et al.[15] in year 2018 developed a model to identify adult viewers of all ages experiencing mental weariness during real-world watching circumstances.	29 individuals (aged 29 and older) who watched videos both before and after performing cognitive activities were tracked using eye-tracking technology.	The model, which made use of an SVM classifier, was able to detect with 91.0% accuracy whether or not a person was weary after seeing a film for only 30 seconds.	- A few people took part. This approach can be used to calculate the level of mental tiredness during different cognitive tasks. -In older persons, subjective and objective assessments of mental tiredness frequently diverge.
9	K. Shankar et al. [22] in year 2018 forecasted about thyroid disease, a multi-kernel support vector machine was created utilising the IGWO calculation for the best feature selection.	UCI repository	MKSVM produces 98.65% accuracy.	We can create more potent methods that enhance performance and enable thyroid illness diagnosis.
10	K. Shailaja et al. [1] in year 2018 presents numerous machine learning (ML) strategies for predicting a range of ailments, including heart disease, breast cancer, diabetes, and thyroid disease.	UCI Machine Learning Repository	Heart illness is detected with 86% accuracy using Nave Bayes. SVM provides 96.40% accuracy in detecting breast cancer, while CART provides 79% accuracy in detecting diabetes.	Building effective decision-support systems for healthcare applications is necessary.

STUDY DEPARTURES Healthcare organizations have access to a wide variety of data from a number of sources, including numerical data on infectious cases, pictures from radiography, textual data, biomedical data, 2. The integration of embedded sensor data, speech sound data, natural language to biological data is a significant challenge, and there are several inconsistencies and errors in health data. 3. Inadequate Data: Due to missing and noisy information, the existing real-life data cannot be considered completely reliable. Much of the data is still paper-based and so unavailable in the common database. 4. The Reliability and Accuracy of AI/ML: Artificial intelligence and machine learning are not being used to their full potential since it is difficult to analyze data patterns in such dynamic data. Because of this, the results are not as accurate or reliable. 5. The development of a reliable and accurate decision-support system for use in healthcare is essential. [2] in

6. More complex and combination models are needed to improve the accuracy of early cardiac disease prediction. [13]. 7. Disease prediction in new patients is done using information of old patients. It is necessary to construct physical infrastructure that provides an integrated system that enables the real-time transfer of data [5]. [7]. Choose the right

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

machine learning algorithm might be challenging as it's not always possible to predict which one would perform best. [5].

CONCLUSION

The healthcare business is in dire need of machine learning, which has immense promise. One example of the new types of data that provide both benefits and challenges is data from wearable devices. Additional data sets allow for the use of deep learning and other enhancements, leading to improved results. Practical research in computer science and statistics will be stimulated by machine learning that makes use of data from electronic health records. This is because machine learning affects many different areas of healthcare. For the betterment of both patients and physicians, there are alternative ways to merge specific multimedia with machine learning and deep learning models that have been trained on diseases. Machine learning (ML) has many potential applications, including but not limited to illness detection, diagnosis, treatment, and care, as well as the search for a cure for certain diseases. A comprehensive examination of the prognosis of long-term conditions such as diabetes, breast cancer, and cardiovascular issues has been undertaken in connection to the healthcare industry's use of machine learning algorithms. Disease prediction is the primary aim of this research. Since much of the previous research focused on chronic illness prediction, the logical conclusion is that machine learning methods were seldom used for this purpose. Since there is a great deal of unrealized potential, further investigation into this field is required.

REFERENCES

[1]. K. Shailaja, B. Seetharamulu and M. A. Jabbar. (2018) "Machine Learning in Healthcare: A Review." Proceedings of the 2nd International Conference in Electronics, Communications and Aerospace Technology (ICECA 2018). vol. 18(42487), p.910-913.

[2]. A. Dhillon and A. Singh. (2018) "Machine Learning in Healthcare Data Analysis: A Survey." Journal of Biology and Today's World. vol. 8(2), p.1-10.

[3]. J. Wiens and E. S. Shenoy. (2018) "Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology ." Clinical Infectious Diseases . vol. 66, p.149-153.

[4]. D. I. Dogaru and I. Dumitrache, "Big Data and Machine Learning Framework in Healthcare," 2019 E-Health and Bioengineering Conference (EHB), 2019, pp. 1-4, doi: 10.1109/EHB47216.2019.8969944.

[5]. S. S. Aljameel, I. U. Khan, N. Aslam, M. Aljabri, and E. S. Alsulmi, "Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients," Hindawi Scientific Programming, vol. 2021, no. 5587188, pp. 1-10, Apr. 2021.

[6]. Y. Zoabi, S. Deri-Rozov, and N. Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," Nature Partner Journals, vol. 4, no. 3, pp. 1-5, 2021.

[7]. L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset," SN Computer Science, vol. (2021)2, no. 11, pp. 1-13, Nov. 2020.

[8]. C. Sáez, N. Romero, J. A. Conejero, and J. M. García-Gómez, "Potential limitations in COVID-19 machine learning due to data source variability: A case study in the nCov2019 dataset," Journal of the American Medical Association, vol. 2, no. 28, pp. 360-364, Oct. 2020.

[9]. B. Maram, G. Padmapriya, and A. R. Satish, "A Framework for Performance Analysis on Machine Learning Algorithms using COVID-19 Dataset," Advances in Mathematics: Scientific Journal, vol. 9, no. 10, pp. 8207-8215, 2020.

[10]. M. Shouman, T. Turner, and R. Stocker, "Using Data Mining Techniques in Heart Disease Diagnosis and Treatment," 2012 Japan-Egypt Conference on Electronics, Communications and Computers, pp. 173-177, 2012.

[11]. B. Bahrami, and M. H. Shirvani, "Prediction and Diagnosis of Heart Diseases by Data Mining Techniques," Journal of Multidisciplinary Engineering Science and Technology, vol. 2, no. 2, pp. 164-168, Feb. 2015.

[12]. R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of Heart Disease using a Combination of Machine Learning and Deep Learning," Hindawi Computational Intelligence and Neuroscience, vol. 2021, no. 8387680, pp. 1-11, July 2021.

[13]. D. Shah, S. Patel, and S. K. Bharti, "Heart Disease Prediction using Machine Learning Techniques," SN Computer Science, vol. 1, no. 345, pp. 1-6, Oct. 2020.

[14]. A. Martinho, M. Kroesen, and C. Chorus, "A Healthy Debate: Exploring the views of medical

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

doctors on the ethics of artificial intelligence," Artificial Intelligence In Medicine, vol. 121, no. 102190, pp. 1-10, Oct. 2021.

[15]. Y. Yamada, and M. Kobayashi, "Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluate in younger and older adults," Artificial Intelligence In Medicine, vol. 91, pp. 39-48, June 2018.

[16]. M. A. Ahmad, C. Eckert M.D, and A. Teredesai, "Interpretable Machine Learning in Healthcare," ACM BCB, vol. 18, pp. 559-560, Sep. 2018.

[17]. S. K. Sen. (2017) "Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms." International Journal of Engineering and Computer Science (ISSN). vol. 6(6), p.21623-21631.