An Efficient Multi-Disease Diagnosing In Clinical decision Making Using Healthcare Machine Learning Techniques

Ramasamy R[1] , G Pattabirani[2], S Parasuraman[3]

[1]Research Scholar, Department of Information Technology, Annamalai University, Chidambaram, Tamil Nadu, India, Email: rams.ms01@gmail.com
[2]Assistant Professor, Department of Information Technology, Annamalai University, Chidambaram, Tamil Nadu, India, Email: prvijayaraja@gmail.com
[3]Co-Guide and Professor, Department of ECE, KarpagaVinayaga College of Engineering and Technology, Chengalpattu, Tamil Nadu, India, Email: parasuhodece78@gmail.com

Received: 16.04.2024 Revised : 18.05.2024 Accepted: 27.05.2024

ABSTRACT
Healthcare uses Machine Learning techniques (HMLT)can be for better application of knowledge and identifying successful prescription patterns for diseases. Usage of computer aided diagnosis for expert opinion learning have definite advantage. Integrated Machine Learning (IML)with forecasting can provide a dependable and a high quality desirable outcome. Prediction of diseases using machine learningtechniques is a motivating task for augmenting diagnostic accuracy. Hence the objective of this research is usage ofHMLT/IML methodology that can take less time and which can be more economical. The methodology can be useful to predict healthcare diseases. Hence to understand the usage of this research work is to identify the methodology to predict Healthcare diseases from patient's records and suggest a non-invasive machine learning model.

Keywords: HMLT, IML,CART, Clustering, Diagnosis, Data mining, Decision Making, Healthcare.

1. INTRODUCTION

Healthcare delivery system generates and stores enormous quantum of primary data. While technological advancements in the form of computer-based patient record software and personal computer hardware are making the collection of and access to health care data more manageable, few tools exist to evaluate and analyze this clinical data after it has been gathered and filed.

Analysis of primary data available can enhance the better management of disease progression. Better search modes need to be developed for the task. Past efforts in this area have been limited primarily to epidemiological studies on aData Mininginitiative and claims databases. Discovery in Databases or Knowledge Discovery in Databases (KDD), is the search for relationships and global patterns that exist in large databases but are 'hidden' among the vast amounts of data [1]. The typical machine learning process involves transferring data originally collected in production systems into a data warehouse, cleaning or scrubbing the data to remove errors and check for consistency of formats, and then searching the data using statistical queries, neural networks, or other machine learning tools [2]. Though many applications of KDD have focused on discovering novel data patterns to solve business related problems, they have also been used extensively in the healthcare study and researches. Machine learninghas been used is used to discover subtle factors affecting the success and failure of therapeutic modalities which led to improvements in patient care [3].

MATLAB can work with matrices, deleting a row, a column, transposing a matrix, calculating the determinant…etc. Similarly Machine learningsystems can be used for identifications and intervention strategies of diseases that were likely to cut costs and thereby reducing the economic burden. Thus, the eventual goal of knowledge recovery effort is to identify factors that can improve the quality and reduce costs in mining thehealthcare information. This research work analysis themachine learningtechniques which are tested with MATLAB.

2. LITERATURE SURVEY

Alanazi, et al., (2022) postulated about prognostic analytics in health care and used six ML techniques on the data set. Evaluation was performed and compared with different ML models for predicting diabetes. The performance of SVM and KNN showed high accuracy for PIDD dataset. However, the work did not consider hyper-parameter tuning models for obtaining high accuracy.

Rastogi M et al., (2022) demonstrated that breast cancer is one of the most common and dangerous disease these days which costs many lives. Machine learning and its techniques are very useful for early detection of breast cancer. So far, extreme machine learning gives better results with accuracy and time.
Phasinam, K. et. al., (2022) has mentioned in the algorithm that is supposed to provide Healthcare precision with the overall classification as being mentioned in the traditional system. The proposed machine learning approaches has enabled the understanding and predicting of heart diseases quickly. Machine learning the exactness which has been traced using the random model has resulted in computing the coronary disease at the earliest.
Amiri Zet. al., (2023) analyzed the use of machine learning techniques to predict and prevent hypoglycemia in patients. The algorithm used real-time data from a blood glucose monitoring device to calculate the appropriate insulin dose required to maintain blood glucose levels within a normal range. Their results demonstrated the effectiveness of the machine learning model in predicting hypoglycemic events with high accuracy.
Shukur, B. S et al., [2023] have involved machine learning techniques to diagnose early-stage Healthcare disease by developing accurate and effective predictions based on digital patient records collected from the Kaggle platform. The performance of the techniques is compared to the accuracy of disease prediction.

3. PROBLEM STATEMENT

Human population explosion has resulted in the manifestation of many noval and hither by undocumented diseases, where certain diseases may not have a permanent cure. Treating diseasesare a major challenge to Health care providers as the Signs and symptoms of diseasesimulate other disorders. Management process in the healthcare facility are a major challenge to healthcare providers, thus making complex issue. A patient's feelings of distress, guilt, and anxiety occur dueto their negative social experiences when they fail to understand the intensity of their illness [5]. Moreover, patients experience unique challenges personally, when diagnosed with an invisible illness that threaten their quality of life [6]. To improve a patients' emotional and physical health, an awareness on Healthcare diseases in patients and proper direction in diagnosis of diseaseand management are needed. Patients are vulnerable to psychological apprehension due to the decreased quality of living. Patients need treatment quickly and prompt intervention to reduce the morbidity to identify and take care of exhibited effectively. Thus, though disease can be treated effectively with the modern parts, the diagnosis early is a major problem due to its manifestations simulate other diseases. In processing data for identifying disease, machine learning techniques need previous history of patients adequately. Though current improvements in Medicare have enhanced patient's survival rates, fear of morbidity and mortality persists.

4. PROPOSED WORK

The goal of early disease prediction is to provide an early warning signal of impending diseases, enabling timely intervention to reduce errors in diagnosis and management. By designing and proposing new algorithms and techniques, healthcare systems can more accurately predict diseases at an early stage, allowing for quicker response and treatment. Analyzing existing algorithms and comparing the proposed methods' effectiveness ensures continuous improvement in prediction models. This, in turn, reduces the time lag in diagnosis, ultimately lowering healthcare costs and improving patient outcomes. Early prediction also helps extend life spans while enhancing the quality of life for afflicted patients.

5. METHODOLOGY

The data storage, processing and retrieval can enhance the better management for which complete comprehensive details of every individual is needed.
The following methodology is used

1. Dataset
2. Data Cleaning
3. Missing Value Prediction
4. Dataset Preparation
5. Decision Making

5.1 Dataset

The Dataset used in this work is the Chicago Lupus Database (CLD). This is a registry ofindividuals with lupus used for lupus research with a probable or definite lupus symptom. The Chicago Lupus Database (CLD)is a comprehensive data repository aimed at improving the understanding, diagnosis, and management of systemic lupus erythematous (SLE), an autoimmune disease. It is part of a larger initiative

to study lupus in diverse populations and provide crucial insights into the disease's progression, impact, and treatment outcomes.CLDattributes are Age, Gender, Test Sample, Disease Activity, Symptoms, Severity, Involved
Organs, Tests conducted and follow-ups.

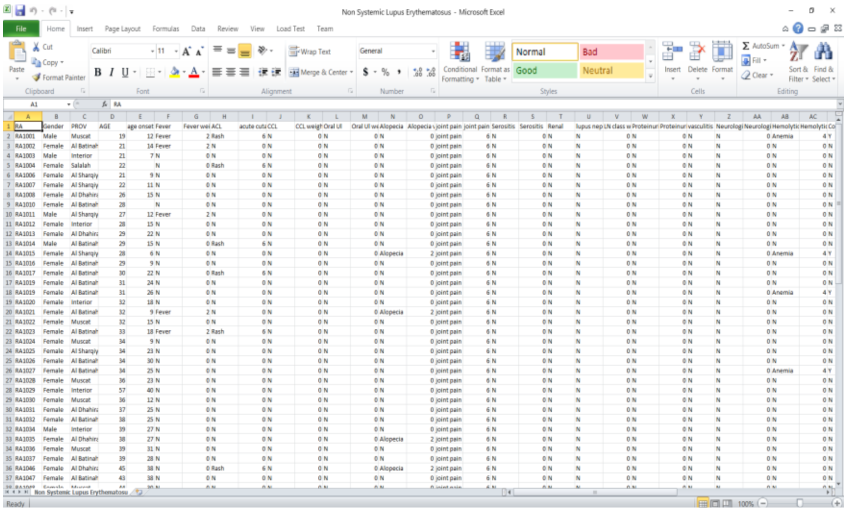

Fig. 5.1 Sample Dataset

5.2 Data Cleaning

Data cleaning is an important part of machine learning for its significance in model building. Data cleaning can make or break an analysis. Professional data analysts spend a lot of time in this step. A clean dataset can get desired results even with a simple algorithm, which is beneficial. Figure 5.2 depicts the flow of Data Cleaning.

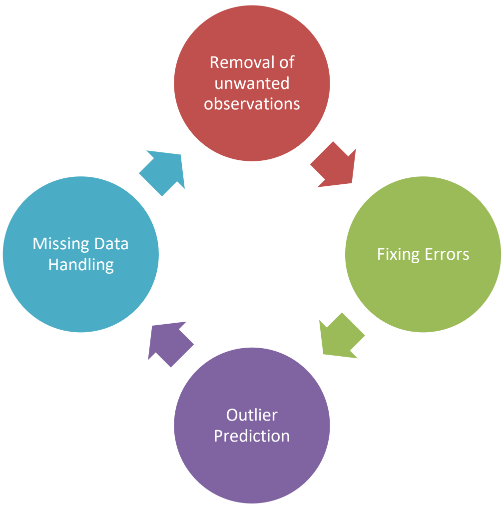

Fig. 5.2 Data Cleaning Flow

Data cleaning involves different steps for different data. The steps followed in this research work is Missing Value Prediction, Redundancy Avoidance, Filtering (Fill mean mode value) and Attribute Reduction.

5.3 Missing Value Prediction

Missing data can be categorized into three main types based on the underlying cause and its relationship with other features. Missing Completely At Random (MCAR)occurs when the missing values are purely random, with no relationship to any other feature in the dataset, making it the highest level of randomness. In this case, the likelihood of a value being missing is independent of both observed and unobserved data. Missing At Random (MAR)refers to situations where the missing data is dependent on other observed features. In this scenario, the missingness can be predicted or explained by values of other variables in the dataset, meaning that the pattern is not entirely random but linked to certain characteristics. Finally, Missing Not At Random (MNAR) implies that the missing data is related to the value of the variable itself or the data collection process. In such cases, the missingness may be influenced by unobserved factors, and the data gathering process should be reviewed for potential biases or issues that might have caused the missing values. Understanding these distinctions helps in selecting the appropriate method for handling missing data.

5.4 Dataset Preparation

Adataset is a comprehensive collection of patient's data. The data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set. Machine Learning methods use a training data set where actual data is used to train the proposed model for performing various actions. The training data set applies concepts like neural networks for learning and expected results. It includes both input and expected output data. Training sets make up the majority nearly 70% of the total data. The testing model adapts to fit to parameters in a process called adjusting weights. The test data set is then used to evaluate how well a machine learning technique was programmed with the training data set. Testing sets represent remaining30% of the data and ensures the input data grouped together is verified with correct outputs.

5.5 Decision Making

Making the right decision is often a challenge. A simple and quick approach for taking a decision is following past experiences in similar situations. The human brain decision is based on two factors namely logical and intuitive. Most decision are an automatic response as the logical part invents a reason for the decision. The intuitive system based on an entity from several plausibleconclusions which need to be assessed. Tools like decision matrix can help in unbiased decision making. It is an advanced approach for making decision and scores each possible option against certain criteria or feature. This approach results in creating decision matrix for analysis of possible options. Machine Learning Techniques (MLT) help to improve decision making and can be viewed as assigning or predicting correct label based on data features (Classification Problem).

6. EXPERIMENTAL RESULTS

6.1 Data Cleaning

Data cleaning is a crucial step in the machine learning process, as the quality of data directly impacts the accuracy and performance of the model. It involves detecting and correcting or removing inaccurate, incomplete, irrelevant, or corrupt data from the dataset to ensure that the model is trained on clean and reliable information.

The fields taken and first analyzed for Missing values. Missing values can change the course and direction of a result, if not handled proper. Hence, first step is to addressthe missing value. Figure 5.3 depicts the output of Missing Values

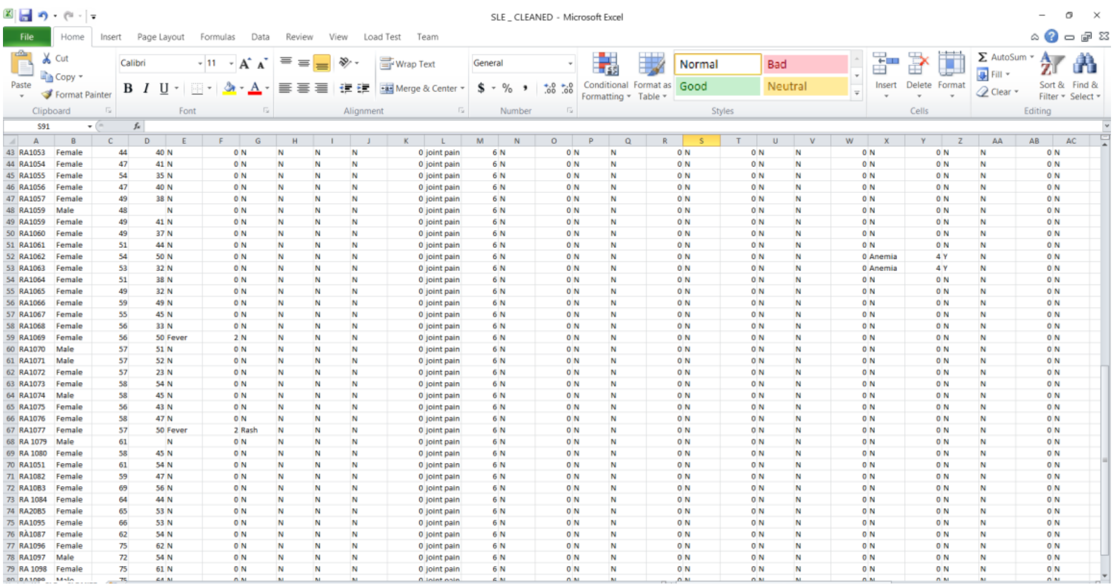

Fig. 6.1 Data Cleaning Results

6.2 Decision Making

The decision making tree is one of the better known decision making techniques, probably due to its inherent ease in visually communicating a choice, or set of choices, along with their associated uncertainties and outcomes. Their simple structure enables use in a broad range of applications. They can be drawn by hand to help quickly outline and communicate the critical elements in a decision. Alternatively, a decision tree's simple logical structure enables it to be used to address complex multiple decision scenarios and problems with the aid of computers. The proposed work uses the CART Algorithm for forming its decision making tree based on the criteria listed in Table 6.1.

Table 6.1 Decision Making Criteria

Attribute	Decision Weight
General Attributes	1
Disease Activity	2
Symptoms	3
Test Results	4

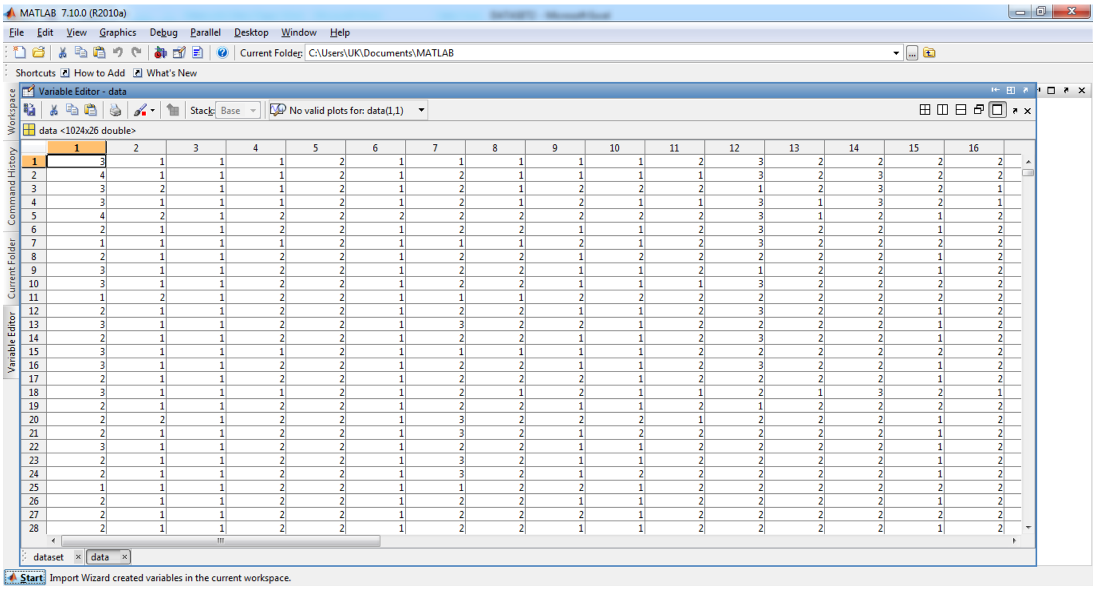

Fig. 6.2 Decision Making Criteria Results

6.3 Hybrid K-Means and CART

The Hybrid K-Means and CART algorithm combines two powerful machine learning techniques: K-Means clustering and CART (Classification and Regression Trees). This hybrid approach is useful in scenarios where both unsupervised and supervised learning techniques are needed to enhance model performance, particularly for segmentation and prediction tasks.

The hybrid approach combines the strengths of both K-Means and CART. The key idea is to use K-Means to pre-process or segment the data and then use CART for building a predictive model within each cluster. Here's how it typically works:

6.4 Steps in the Hybrid Algorithm

Step 1: Apply K-Means Clustering:

- First, apply K-Means clustering on the dataset to divide it into K clusters.
- This clustering step allows for the discovery of natural groupings in the data that might not be immediately obvious.

Step 2: Apply CART to Each Cluster:

- After clustering, apply CART (decision tree) within each cluster to build a predictive model for classification or regression.
- The rationale is that each cluster might represent a unique segment of the data, and building a separate model for each cluster can increase the overall model's accuracy.
- CART is used to predict either the class labels or the continuous target values for the instances within the clusters.

6.5 Proposed Algorithm Steps

Input: The Healthcare CLD dataset with n entities.
Output: A set of optimal clusters K_i.
Step 1: Likewise any abnormal or out of range values are also not considered for preprocessing.
Step 2: Records of incorrect and missing data are not considered and assign mean mode value.
Step 3: If Multi variant attributes or more than one instance are there then
Step 4: Remove redundant value using deletion query operation
Step 5: Normalization of missing attribute instances is done by filtering.
Step 6: Processed, filtered value converted as MAT file and stored in separate database.
Step 7: Initialize dataset $\sum$ (F) = { f_1,f_2,f_3 ... f_n} attributes.
Step 8: Identify the Outliers in the considered column $\sum$ (F') = { f_1',f_2',f_3' ... f_n'}
Step 9: Repeat, formulate the rules for identifying the similar attributes.
Step 10: do until, Identify the frequent itemsets.
Step 11: Specify the threshold Mean, proportion value.
Step 12: Identify the K initial mean vector from the attributes
Step 13: Identify the distance between f_i attributes and the centroid value f_j.
Step 14: Recalculate until new centroid f_j identified.
Step 15: Identify the end convergence.
Step 16: Fin the neighborhood active attribute rule set.
Step 17: Generate recommendations from most frequent itemset.
Step 18: Identify the disease threshold mean prediction value.

Table 6.2 Comparative Performances by Machine learning Techniques

Algorithms	Sensitivity	Specificity	Accuracy
CART	96.58	97.33	96.94
K-Means	94.33	94.13	94.24
Proposed K-Means +CART	98.56	98.87	99.14

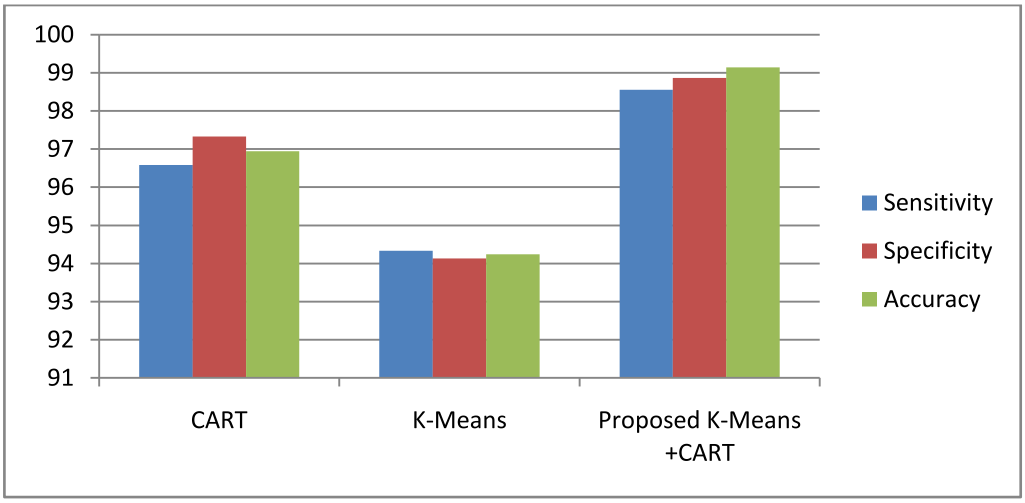

Fig. 6.3 Performances of Proposed Work

6.6 Discussions

The management of data and utility in healthcare sector is a difficult task. But the usage of computer to store and retrieval can help. However the data generated by an individual / or a community can be enormous unless sorted out and categories can get lost in the research.
The research identifies 5 system of approach namely

1. Dataset
2. Data Cleaning
3. Missing Value Prediction
4. Dataset Preparation
5. Decision Making

Every system has its own advantage over other. Each system addresses a particular need.Putting all together one by one or at the same time, the advantages are H/c Count 98% semantic, specific and secure.

The 2% error can happen in any system which is negotiable.Further study could eliminate ever this.However H/c can't compared with Human intelligence and analytical mind can enhance the efficiency to nearly 100%. Nearly 100% is become difficult as every individual can be unique.This system approach can help to predict the disease at an early stage, so management and near cure is achievable. (Near cure be return to 100% normally is as of now impossible) but this system helps in early diagnosis and proper management.The Hospital days can be cut short and economic burden of the patent can be greatly reduced.The research work can be put to proper use in the health care management system.

7. CONCLUSION

Computer assisted management system had pervaded through all the walks of life. The technology and associated AI help a lot in the proper healthcare management. However, managing the data accumulated and stored need to be put to proper use.

To achieve the effective management of the data and to use the same in health sector can help to reduce the hospital days the cost and enhance better management.

REFERENCES

[1] Alanazi, Abdullah. "Using machine learning for healthcare challenges and opportunities." Informatics in Medicine Unlocked 30 (2022): 100924.

[2] Rastogi, M., Vijarania, D. M., &Goel, D. N. (2022, August). Role of machine learning in healthcare Prediction. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC).

[3] Phasinam, K., Mondal, T., Novaliendry, D., Yang, C. H., Dutta, C., &Shabaz, M. (2022). Healthcare Analyzing the Performance of Machine Learning Techniques in Disease Prediction. Journal of Food Quality, 2022(1), 7529472.

[4] Amiri, Z., Heidari, A., Darbandi, M., Yazdani, Y., JafariNavimipour, N., Esmaeilpour, M., ...&Unal, M. (2023). The personal health applications of machine learning techniques in the internet of behaviors. Sustainability, 15(16), 12406.

[5] Shukur, B. S., &Mijwil, M. M. (2023). Involving machine learning techniques in heart disease diagnosis: a performance analysis. International Journal of Electrical and Computer Engineering, 13(2), 2177.

[6] Siddiq, M. (2022). Healthcare prediction Use of Machine Learning to predict patient developing a disease or condition for early diagnose. International Journal of Multidisciplinary Sciences and Arts, 1(1).

[7] Kavitha, C., Mani, V., Srividhya, S. R., Khalaf, O. I., &Tavera Romero, C. A. (2022). Early-stage Alzheimer's disease prediction using machine learning models. Frontiers in public health, 10, 853294.

[8] Kute, S. S., ShreyasMadhav, A. V., Kumari, S., &Aswathy, S. U. (2022). Machine learning–based Healthcare disease diagnosis and prediction for E-healthcare system. Advanced analytics and deep learning models, 127-147.

[9] Ahmed, D., Neema, R., Viswanadha, N., &Selvanambi, R. (2022). Analysis and Prediction of Healthcare Sector Stock Price Using Machine Learning Techniques: Healthcare Stock Analysis. International Journal of Information System Modeling and Design (IJISMD), 13(9), 1-15.

[10] Hasanova, H., Tufail, M., Baek, U. J., Park, J. T., & Kim, M. S. (2022). A novel blockchain-enabled heart disease prediction mechanism using machine learning. Computers and Electrical Engineering, 101, 108086.

[11] Bhardwaj, A. (2022). Promise and provisos of artificial intelligence and machine learning in healthcare. Journal of Healthcare Leadership, 113-118.

[12] Khan, B., Naseem, R., Shah, M. A., Wakil, K., Khan, A., Uddin, M. I., & Mahmoud, M. (2021). Software defect prediction for healthcare big data: an empirical evaluation of machine learning techniques. Journal of Healthcare Engineering, 2021(1), 8899263.

[13] Amin, R., Al Ghamdi, M. A., Almotiri, S. H., &Alruily, M. (2021). Healthcare techniques through Machine learning: issues, challenges and opportunities. IEEE Access, 9, 98523-98541.

[14] Gangal, A., Kumar, P., Kumari, S., &Saini, A. (2021). Prediction models for healthcare using machine learning: A review. Handbook of research on disease prediction through data analytics and machine learning, 70-91.

[15] Tawhid, A., Teotia, T., &Elmiligi, H. (2021). Machine learning for optimizing healthcare resources. In Machine Learning, Big Data, and IoT for Medical Informatics (pp. 215-239). Academic Press.

[16] Gomathy, C. K., & Naidu, M. A. R. (2021). The prediction of Healthcare disease using machine learning. International Journal of Scientific Research in Engineering and Management (IJSREM), 5(10), 1-7.
[17] Kasula, B. Y. (2021). Machine Learning in Healthcare: Revolutionizing Disease Diagnosis and Treatment. International Journal of Creative Research In Computer Technology and Design, 3(3).
[18] Jindal, H., Agrawal, S., Khera, R., Jain, R., &Nagrath, P. (2021). Healthcare disease prediction using machine learning algorithms. In IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012072). IOP Publishing.
[19] Alazzam, M. B., Alassery, F., &Almulihi, A. (2021). [Retracted] A Novel Smart Healthcare Monitoring System Using Machine Learning and the Internet of Things. Wireless Communications and Mobile Computing, 2021(1), 5078799.
[20] Badawy, M., Ramadan, N., &Hefny, H. A. (2023). Healthcare predictive Techniques using machine learning and deep learning techniques: a survey. Journal of Electrical Systems and Information Technology, 10(1), 40.
[21] Gaurav, K., Kumar, A., Singh, P., Kumari, A., Kasar, M., &Suryawanshi, T. (2023). Healthcare disease prediction using machine learning techniques and real-life parameters. International Journal of Engineering, 36(6), 1092-1098.
[22] Bhatt, C. M., Patel, P., Ghetia, T., &Mazzeo, P. L. (2023). Effective Healthcare disease prediction using machine learning techniques. Algorithms, 16(2), 88.
[23] Chhetri, B., Goyal, L. M., & Mittal, M. (2023). How machine learning is used to Prediction techniques in digital healthcare: A systematic review. International Journal of Information Management Data Insights, 3(2), 100175.
[24] Sk, K. B., Roja, D., Priya, S. S., Dalavi, L., Vellela, S. S., & Reddy, V. (2023, March). Healthcare Disease Prediction and Classification using Hybrid Machine Learning Algorithms. In 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) (pp. 1-7). IEEE.
[25] Chaithra, Chilkaragi Shankar, ShivarudraswamySiddesha, VN ManjunathAradhya, and ShanmukharadhyaKeragoduNiranjan. "A Review of Machine Learning Techniques Used in the Prediction of Healthcare Disease." Revue d'IntelligenceArtificielle 38, no. 1 (2023): 201.