Ann. For. Res. 66(1): 1125-1137, 2023
ISSN: 18448135, 20652445

ANNALS OF FOREST RESEARCH
www.e-afr.org

A STUDY ON DISEASE PREDICTION USING MACHINE LEARNING

Kamal Narayan Kamlesh
Research Scholar, Magadh University Bodh Gaya
Kamal.kamleshjha@gmail.com
Teaching Associate, Dept. of CEA, GLA University, Mathura
Kamal.kamlesh@gla.ac.in

Dr Kumar Vishal
Assistant Professor, MU Bodh Gaya
kumarvishal.bhu@gmail.com

ABSTRACT:

Machine learning can be used in a variety of fields, including healthcare. Medical facilities should be much more advanced in order to provide the best possible treatment to patients. Additionally, a number of machine learning algorithms have been selected and implemented to achieve the best results for instance, Random Forest, Decision Tree Classifier, and K-Nearest Neighbour algorithms, among others. A high increase in patient satisfaction can be achieved by implementing machine learning in healthcare. An integrated health care system can be designed by integrating machine learning functions. If certain machine learning algorithms can be used to predict the disease beforehand instead of directly treating the patient, the entire treatment process will be much more efficient. The possibility of diseases being overlooked or not being detected early also exists. Predicting the patient's disease before starting treatment is therefore extremely important. Since prevention is better than cure, correctly predicting disease could lead to early detection and prevention.

Keywords: Healthcare, Logistic Regression and K-Nearest Neighbour.

INTRODUCTION

There has been a great deal of money made in the healthcare industry over the years. Data is generated and used by the healthcare industry to gather information about a patient's disease (Mareeswari, V. et. al. (2018)). Additionally, this information will be used to keep a patient's health as efficient and effective as possible. Using the informative data from healthcare sciences could also be helpful in improving this area. In order to extract information from the vast amount of data, we use data mining and machine learning techniques. According to a result of this study, it will be possible to predict the disease in advance so that it can be prevented at an early stage, so people's lives can be saved and treatment costs can be reduced (Verma, A., & Naaz, I. (2022)). Non-manual medical systems are most effective for improving individual health, so we should adopt them in India too. Machine learning will be used in healthcare to improve patient treatment using the concept of machine learning. The use of machine learning has already made the identification and

prediction of various diseases much easier (M.Young (1989)). By analyzing data and predicting diseases, machine learning algorithms can be used to treat patients efficiently. Data mining and machine learning techniques can also be used to predict diseases using machine learning algorithms. There have also been many studies using data mining techniques to predict some diseases based on pathological data. The purpose of these approaches was to predict the reoccurrence of certain diseases beforehand. As well as controlling the disease, some approaches attempted to predict it (Fotiadis, D. I. et. al. (2015)). The recent advances in deep learning in disparate areas of machine learning have allowed machine learning models to understand hierarchical representations of raw data. In recent years, the concept of big data technology has been giving more attention to the prediction of diseases.

PROPOSED SYSTEM

First, we use the concepts of the Sklearn library to build a simple classification model. Python uses Scikit-learn as one of its machine learning libraries. As well as supporting NumPy and SciPy, there is support for algorithms like support vector machines, random forests, and K-neighbors. Before we can compile the model, we must first define it. In order to test the performance of different algorithms, the different algorithms were compiled first and then their accuracy scores were determined. The selection of features was also an important step. With so many features in the dataset, selecting and retaining the most relevant ones is equally important. The algorithms used a variety of procedures for selecting features. By using the fit () function, we are able to fit our model on our respective data after we have created and compiled the model. Models are initially fitted to the original dataset, which has all of the features, but over fitting can happen. Several datasets with different numbers of features were fitted with this model in an attempt to address the problem. Different algorithms were selected based on their optimal features in order to achieve high accuracy. The model will be evaluated by examining parameters such as accuracy, precision, process speed, etc., as soon as we've completed the above step. The pros and cons of our model should be understood and corrected in an efficient manner. Due to this, the model must be analyzed critically and improved as necessary.

LITERATURE REVIEW

An algorithm based on machine learning was proposed by Dahiwade et al. (2019) to predict common diseases. Many common diseases' symptoms were imported into the symptoms dataset from the UCI ML repository. The system predicts multiple diseases using CNN and KNEes classification techniques. Furthermore, the proposed solution was accompanied by additional information regarding the patient's lifestyle, which helped to determine the disease's risk level. Dahiwade et al (2019) compared KNN algorithms with CNN algorithms for accuracy and processing speed. 84.5% of CNN's accuracy and 11.1 seconds of processing time were achieved respectively. According to the statistics, the KNN algorithm does not perform as well as the CNN algorithm.

Compared to typically supervised algorithms such as KNN, NB, and DT, CNN outperformed them in the study by S Kolte et al (2019). Based on its ability to detect complex nonlinear relationships in the feature space, the authors concluded that the proposed model scored higher in terms of accuracy. Furthermore, CNN can accurately predict diseases with high complexity by detecting features with high importance that describe the disease better. There are empirical observations and statistical arguments that support this conclusion. However, the presented models lacked specifics, such as parameters related to NeuralNetworks, architecture types, learning rates, and back propagation algorithms (Schowengerdt, R. A. et. al. (1995)). Furthermore, the analysis of performance is only based on accuracy, which undermines the validity of the findings presented. A bias problem faced by the algorithms was also not considered by the authors.

Serek et al. (2012) evaluated classifiers for detecting chronic kidney disease based on the Kidney Function Test (KFT) dataset. In this study, RF, KNN, and NB classifiers are compared for accuracy, precision, and F-measure. As a result of analysis, RF performed better when it came to F-measure and accuracy, while NB performed better when it came to precision.

Vijayarani (2015) used SVM and NB to detect kidney diseases in consideration of this study. A classification system was developed to identify four types of kidney disease, including acute nephritis syndrome, acute kidney failure, chronic glomerulonephritis, and chronic kidney disease. In addition to determining the best algorithm based on accuracy and execution time, the research also aimed to determine which algorithm performed the best in terms of performance and accuracy. The accuracy of SVM was significantly higher than that of NB, making it the superior algorithm. In spite of this, NB classified data with the shortest possible execution time. In addition, there have been other empirical studies that have investigated how to recognize CKD; Charleonnan et al. (2017) and Kotturu et al. (2019) found that SVMs are best suited for kidney diseases since they handle semistructured and unstructured data well.

Due to its flexibility, SVM can handle larger feature spaces, which is why it is highly accurate at detecting complex kidney diseases. The conclusions obtained are supported by findings; however, the performance of ML algorithms could have been improved if different hyperparameters had been explored. ML algorithms can be optimized by exploring the hyperparameter space and generating different accuracy results. The prediction of heart disease was performed using supervised machine learning techniques in Marimuthu et al. (2018). In the study, the authors classified the data attributes based on gender, age, chest pain, gender, target, and slope. DT, KNN, LR, and NB are the applied machine learning algorithms. Analyses indicated that the LR algorithm performed the best compared to the other algorithms, achieving an accuracy rate of 86.89%.

With the use of a number of additional variables, such as resting blood pressure, serum cholesterol, and maximum heart rate, Dwivedi (2018) attempted to improve heart disease prediction. There are 130 positive samples and 140 negative samples in this dataset from the UCI ML laboratory. In this dataset from the UCI ML laboratory, 130 positive samples and 140 negative samples are included. The performance of ANNs, SVMs, KNNs, NBs, LRs, and Classification Trees was evaluated by

Dwivedi. As demonstrated by a tenfold cross validation, LR did an excellent job in detecting heart diseases due to its classification accuracy and sensitivity. Polaraju (2017) and Vahid et al. (2017) also support this conclusion, demonstrating that Logistic Regression is more efficient than ANN, SVM, and Adaboost. An extensive analysis of the MLmodels was conducted by the Studies. A variety of hyperparameters were tested for each ML algorithm in order to determine its accuracy and precision. In spite of this advantage, there remains a limitation to the ability of learning models to target diseases accurately and precisely due to the small size of imported datasets.

Sengar et. al. (2020). Investigated various ML algorithms, including RF, Bayesian Networks, and SVM, for detecting breast cancer. A Wisconsin original breast cancer dataset was provided to the UCI Repository for the purpose of comparing learning models for accuracy, recall, precision, and AUC. In order to validate the classifiers, the K-fold validation method was used, where K is equal to 10. Based on simulation results, SVM has proven to have excellent recall, accuracy, and precision. The ROC graph indicated that RF had a higher chance of correctly classifying the tumor.

To determine the best algorithm for predicting breast cancer, Yao (2013) experimented with RF and SVM algorithms. A Random Forest algorithm scored 95.85% accuracy, 95.95% sensitivity, and 95.53% specificity, respectively, with an accuracy rate of 96.28%, 95.74%, and 93.47%, respectively. SVM scored 96.15% accuracy, 95.95% sensitivity, and 95.53% specificity. According to Yao, RF algorithms perform better than SVM algorithms since they provide better estimates of information gained in each attribute. Aside from scalability and less likelihood of causing variances and over fitting, RF is the best way to classify breast diseases. A number of performance metrics were presented in the studies that provided evidence to support the underlined argument. Even so, the preprocessing stage is not beneficial to ML models since it prepares raw data for training. When parts of the data are omitted, the ML algorithm performs worse, according to Yao.

Fuzzy K-Nearest Neighbors (FKNN) are effective tools for the diagnosis of Parkinson's disease (PD) developed by Chen et al. (2013). SVM-based and FKNN-based approaches were compared in the study. PCA was used for the development of optimal FKNN models using the most discriminated features. In the UCI depository, 31 people, 24 of whom have Parkinson's disease, have been measured for biomedical voice measurements. Experimental results show that the FKNN approach performs more sensitivity, accuracy, and specificity than the SVM methodology.

According to Behroozi (2016), a new classification framework can improve the accuracy of diagnosing PD by 15% using a filter-based feature selection algorithm. In spite of the loss of valuable information, independent classifiers were used for each subset of the dataset to classify the framework. Discriminant Analysis, KNN, SVM, and NB were selected as classifiers. According to the results, SVM performed the best across all performance metrics.

As an alternative to SVM, Eskidere (2012) also discussed LS-SVM, General Regression Neural Networks (GRNN), and Multilayer Perceptron Neural Networks (MLPNN). According to the findings, LS-SVM was found to be the most effective model. In order to support this conclusion,

the optimal performance metric used for decoding has been adequately compared to decoders (N. Lavesson (2006)). Various ML algorithms optimize various performance metrics, such as squared error with neural networks and accuracy with KNN and SVM, according to Lavesson (2006). The authors excel at presenting frameworks with detailed explanations. The kernel and regularization value of SVMs, for example, were discussed thoroughly. A calibration of ML models was not performed prior to evaluating performance. A few learning models, like NB, SVM, and RF, can be significantly improved with calibration, according to Caruana (2006).

In A.A. Al-Aidaroos, K.M Bakar, and Z Othman studied the best medical diagnosis mining technique. A comparison of Nave Baeyes with five other classifiers was conducted: KStar (K*), Decision Tree (DT), Neural Network (NN), and a basic rule-based algorithm (ZeroR). UCI's machine learning repository was utilized to evaluate 15 real-world medical problems (Asuncion and Newman, 2007). A Naive Bayes prediction was more accurate than those of other algorithms for eight out of fifteen data sets. It has been proven that treating chronic illness globally is neither time-effective nor cost-effective according to Darcy A. In order to predict potential disease risks, the authors conducted this study. Based on the patient's medical history and ICD-9-CM codes, the CARE program was used to perform this analysis. In order to predict the greatest disease risk, CARE employs collective filtering approaches with clustering based on the medical history of a patient and that of similar patients. Additionally, the authors have defined an ICARE version that incorporates ensemble principles to increase efficiency. A wide range of medical conditions can be predicted quickly with these cutting-edge systems that require no advanced knowledge. As a result of ICARE's remarkable potential risk coverage, early warnings can be issued several years in advance for thousands of illnesses. CARE extends investigation to a wider range of diseases, raises previously unanswered questions, and facilitates early detection and prevention discussions to the maximum extent possible.

An extensive review of data mining techniques used in today's medical research, especially for predicting heart disease, is provided by Jyoti Soni in this study. In a series of experiments on the same dataset, predictive data mining techniques were compared, and the results showed that Decision Trees are sometimes more accurate than Bayesian classifications. Other predictive approaches, including KNN, Neural Networks, and Cluster-based classification, did not perform well. Using Decision Tree Algorithms, Asma Parveen and Shadab Adam Pattekari investigated how consumers' data can be compared with a qualified set of values in order to predict heart disease (Harahap, C. B. et. al. (2018)).

According to this study, heart disease was highly likely based on patient information. In association rule mining, B Gomathy and M.A.NisharaBanu applied grouping, clustering, and association rule mining. Techniques to analyze the various types of heart-related problems (Parajuli, A. (2016)). Any possible outcome of a decision can be shown in a decision tree. The best result is achieved by devising a variety of rules. Age, gender, smoking, overweight, alcohol consumption, blood pressure, glucose level, and heart rate were the criteria used in the study. A risk level is assigned to various parameters with an id between 1 and 100. Prediction levels less

than 1 indicate standard levels of prediction, while risk levels greater than 1 indicate higher levels of prediction. In order to analyze the dataset, the K-means clustering method is used. There are k groups of data in the algorithm. Each dataset point is assigned a closed cluster. As an average of the points of each cluster center, the center is recalculated.

Shankhdhar, et. al (2021, December). In December 2019, a new virus was discovered in the Covid family that caused the Covid-19 infection. There are similar effects to an ordinary flu, making it difficult to perceive as an irresistible illness that primarily affects the lungs area of the body. Since the time it began, it has spread quickly across the globe, posing dangerous challenges. Testing arrangements should not only be technically sound, but should also be achievable and user-friendly as nations seek to expand testing. Recent CT scans and X-rays have revealed remarkable highlights that highlight Covid's seriousness. For quick detection of possible COVID-19-prompted lung contaminations, radiographs, such as X-rays and CT scans, can be used since they are commonly available in general health offices, emergency clinics, trauma centers and even rural settings. With advanced AI, it is now possible to deal with vast amounts of information with precise and quick results in clinical images so as to analyze sicknesses all the more precisely and efficiently in remote regions with additional assistance. A CT-scan x-ray picture is used in this study to analyze Covid-19 using deep learning.

Yadav, A. (2021, July). It is no longer a pandemic, but rather an endemic disease which has killed a number of people around the world. It is unavoidable to live with COVID-19 and its consequences at this time, and there is no precise treatment or remedy for it. One can reduce the financial and administrative burden on healthcare systems by quickly and efficiently screening for COVID-19. A number of variables have been used to predict the likelihood of infection based on research. Health care systems in these countries are strained largely because of inadequate healthcare systems around the world. There are alternative options that might alleviate both the burden on healthcare systems and the economy of COVID-19, despite the fact there are no proven antiviral medication methods or licensed vaccines. Among the most promising approaches for use outside of a clinical setting are non-clinical approaches like machine learning, data mining, deep learning, and other forms of artificial intelligence. You can use these options to diagnose and prognosis patients with 2019-NCoV. A COVID-19 dataset also validates artificial intelligence systems, including decision trees, support vector machines, neural networks, and nave Bayesian models. Correlation coefficients were calculated to establish the degree of interdependence between dependent characteristics and independent variables. At the same time, the model was tested for 20% of the time during preparation. A Random Forest model had 94.99% precision based on the success evaluation.

Bhadana, et. al. (2020, December). Every field uses machine learning. A machine learning (ML) forecasting system can facilitate decision-making on the potential course of action based on perioperative effects. Models in machine learning have been used for defining and prioritizing adverse threat variables for many years. Prediction approaches are widely used to address forecasting challenges. According to this study, ML models are capable of estimating the number

of upcoming COVID-19-affected patients. This study compares five machine learning standard models for forecasting COVID-19's threatening variables, such as linear regression (LR), decision trees, least absolute shrinkage and selection operators (LASSO), random forests, and support vector machines (SVMs). Three forecasts are made by each model, namely the total number of active cases, the total number of deaths, and the total number of recoveries over the next five days. Based on the findings presented in the paper, implementing these techniques for the current COVID-19 pandemic scenario is a promising option. In order to improve accuracy, we used a polynomial of six degrees. Based on the results of the experiment, poly LR and poly LASSO give the best results, followed by LR, LASSO, random forest, and decision trees. In the prediction of COVID-19, SVM shows poor results.

TECHNOLOGY AND METHODS

Our dataset adapts to the model through the use of machine learning algorithms. A machine learning algorithm specifies how each step of the transformation will be learned by the model as data is transformed from input to output. Non-parametric KNNs are also lazy learning algorithms. Data from a database pointing to several classes is used to predict how a new sample will be classified. The term nonparametric refers to a technique that makes no assumptions about the data before use. Therefore, the structure can only be determined by the data. In order to find the best algorithm, we carried out a comparative study of algorithms to understand and improve the process of disease prediction. The data and its trends were visualized in addition to comparing accuracy scores of algorithms. Among the 135 columns in the dataset, there are 131 different symptoms experienced by patients with a variety of ailments. This dataset contains 36 diseases.

When the dependent variables are binary, logistic regression is used to study the relationship between the two variables. An analysis of logistic regression uses maximum estimation as a method of approximating a categorical variable or target, like all regression analyses. Decision trees are supervised learning algorithms that solve classification problems by using predefined target variables. Based on the most significant input variables, we will categorize the samples into some homogeneous groups. It can work with categorical and continuous input variables. It is possible to learn various processes supporting machine learning through the use of supervised algorithms and techniques, such as Linear Discriminant Analysis and Normal Discriminant Analysis.

In ensemble learning, random decision trees are used to classify data. Training sets are corrected for overfitting using these methods. Separating two or more classes is done using these for analyzing differences among groups. This is used to emphasize that they are used as a means of studying differences between groups, i.e. to separate classes. In order to highlight features in a lower dimension of space, features in a higher dimension of space are highlighted.

The LDA is presented in a simple manner that everyone can understand. Every class of our model consists of statistical properties calculated from our data. With multiple variables, the Gaussian has the same properties and is calculated in the same way. The use of these problems is therefore ess

Diseases predicted (Table:1)

Breast cancer	HyperTension	Cervical Spondylosis	Heart Attack
Allergy	Hyperthyroidism	Migraine	Alcohol Hepatitis
Kidney diseases	Drug Reaction	AIDS	Hepatitis D
Common Cold	Gastroenteritis	Chronic cholestasis	Hepatitis B
Diabetes	Varicose Veins	Asthma	Hepatitis E
Monkeypox	Peptic ulcer disease	Paralysis	Hepatitis C
Dengue	Tuberculosis	Typhoid	Hepatitis A
Impetigo	Osteoarthritis	Chicken Pox	Pneumonia
Jaundice	Drug Reaction	Hypoglycemia	Arthritis

Table: 2

Supervised		Unsupervised	
Regression	Classification	Clustering	Dimensionality Reduction
Linear regressions	Support Vector Machine	mean shifts	Feature Selection
Logistic regression	The k-nearest neighbors	K medoids	Linear discriminant analysis

Algorithms for classification

Table: 3

Algorithms	Accuracy Scores	Standard Deviation	No. Of Features
LOGISTIC REGRESSION	0.978687	0.002904	123
CART	0.958263	0.005879	62
KNN	0.970101	0.007956	51
SVM	0.962310	0.007131	53
LDA	0.940268	0.005987	76
RANDOM FOREST	0.8156271	0.043129	51

Results of Feature Selection Using Embedded Methods

An inadequate dataset is always a challenge for machine learning. The handling of missing data is one of the most important steps in ensuring that machine learning algorithms and models produce accurate results. A category object must be created, an encoder fitted to the 'prognosis' column of the system, and then the categories must be transformed into final integers using this encoder. The columns with fewer than 50 values also had missing values. In the next step, we must identify the input and target variables present after we have processed our dataset to a considerable extent. All columns except the 'prognosis' column will be input, since that is the variable we are attempting to predict. The training and test sets are seeded with five each for reproducibility of the results

RESULT AND DISCUSSION

To determine the best algorithm for disease prediction, we examined and improved the process of disease prediction throughout this study. The data was also visualized to allow an in-depth understanding of the data, in addition to comparing accuracy scores between algorithms.

It has been found that the CART model, also known as the decision tree model, was the best performing algorithm when predicting disease based on hospital data when compared to logistic regression, KNN, SVM, Random Forest, and LDA. While only two models were selected using the RFE method, it made a significant impact. Pearson Correlation had the least impact compared to the embedded method, which was the second preferred feature selection method. In order to obtain a certain level of accuracy, the features must also be reduced to some degree. Researcher can reduce the number of embedded methods from 123 to 51 while still retaining relatively good accuracy for all algorithms when we consider embedded methods from this perspective. In comparison to the other two methods with higher precision, Pearson correlation performs less well for feature selection.

Using hospital data, a comparison of different algorithms was performed for the prediction of

diseases. The best results were achieved by Logistic Regression, followed by CART, or simply decision trees, LDA, SVM, Random Forest and KNN. Despite being used on only two models, the RFE method of feature selection had a significant impact. It was found that embedding methods were preferable, but Pearson Correlations were least effective. While analyzing the results, it is important to also keep in mind how much information has been reduced. Embed methods maintained relatively good accuracy across all algorithms by reducing features from 123 to 51. The Pearson Correlation method did not perform as well as the other two methods because it selects features less precisely.

CONCLUSION

The proposed system achieves a higher degree of accuracy compared to existing systems. To provide patients with the best medical care, researchers, physicians, or doctors utilize this information. When applied to healthcare, machine learning can therefore result in an effective treatment as well as well-being for the patient. Our paper demonstrates how machine learning in healthcare can be implemented into our system in some ways. Healthcare can become smarter and more efficient if machine learning algorithms are implemented to predict a disease instead of direct diagnosis. Based on our dataset and the expected results, the Logistic Regression algorithm and KNN algorithm have the best accuracy, while the LDA algorithm has the lowest accuracy. Methods, processes, and techniques can be applied to a variety of medical domains using Machine Learning (ML). Today, ML is used in clinical research for predicting and analyzing outcomes. Furthermore, machine learning is used to detect data errors and correct them, such as in data analysis. As data is transformed from input to output, the machine learning algorithm determines how each step of the transformation will be learned by the model. Non-parametric KNNs are also lazy learning algorithms. New samples are classified based on information in a database that points to several classes. The term nonparametric refers to a technique that does not make any assumptions about the data before it is used. Therefore, the structure can only be determined by the data. In order to find the best algorithm, we carried out a comparative study of algorithms to understand and improve the process of disease prediction. And also try to find out the best algorithm based on accuracy and execution time, the researcher will also determine which algorithm performed the best in terms of performance and accuracy. And which algorithm will perform best compared to the other algorithms for achieving a good accuracy rate. When machine learning algorithms are correctly implemented and used in healthcare, they can be an important source of assistance in integrating computer systems, improving doctors' work, and ultimately improving the quality and efficiency of medicine. It is a debatable topic.

References

1. Dinesh, K. G., Arumugaraj, K., Santhosh, K. D., & Mareeswari, V. (2018, March). Prediction of cardiovascular disease using machine learning algorithms. In 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-7). IEEE.
2. Hartz, A. J., Krakauer, H., Kuhn, E. M., Young, M., Jacobsen, S. J., Gay, G., & Rimm, A.

A. (1989). Hospital characteristics and mortality rates. New England journal of medicine, 321(25), 1720-1725.

3. Verma, A., & Naaz, I. (2022). Prospects and Difficulties of Artificial Intelligence (AI) Implementations in Naturopathy. In Artificial Intelligence for Innovative Healthcare Informatics (pp. 309-327). Springer, Cham.

4. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.

5. Dahiwade, D., Patle, G., & Meshram, E. (2019, March). Designing disease prediction model using machine learning approach. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1211-1215). IEEE.

6. Jadhav, S., Kasar, R., Lade, N., Patil, M., & Kolte, S. (2019). Disease prediction by machine learning from healthcare communities. International Journal of Scientific Research in Science and Technology, 29-35.

7. Paola, J. D., & Schowengerdt, R. A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. International Journal of remote sensing, 16(16), 3033-3058

8. Prange, A. N. S., Bartsch, M., Meiners, J., Serek, M., & Winkelmann, T. (2012). Interspecific somatic hybrids between Cyclamen persicum and C. coum, two sexually incompatible species. Plant cell reports, 31(4), 723-735.

9. Vijayarani, S., & Dhayanand, S. (2015). Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), 816-820.

10. Gunarathne, W. H. S. D., Perera, K. D. M., & Kahandawaarachchi, K. A. D. C. P. (2017, October). Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD). In 2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE) (pp. 291-296). IEEE.

11. Kotturu, P. K., & Kumar, A. (2020, June). Data mining visualization with the impact of nature inspired algorithms in big data. In 2020 4th international conference on trends in electronics and informatics (ICOEI)(48184) (pp. 664-668). IEEE.

12. Pc, J., Marimuthu, T., Devadoss, P., & Kumar, S. M. (2018). Prevalence and measurement of anterior loop of the mandibular canal using CBCT: A cross sectional study. Clinical implant dentistry and related research, 20(4), 531-534.

13. Dwivedi, S. K., & Vishwakarma, M. (2018). Hydrogen embrittlement in different materials: A review. International Journal of Hydrogen Energy, 43(46), 21603-21616.

14. Polaraju, K., & Prasad, D. D. (2017). Prediction of heart disease using multiple linear regression model. International Journal of Engineering Development and Research Development, 5(4), 1419-1425.

15. Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H., & Gutierrez, J. (2017,

July). A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In 2017 IEEE symposium on computers and communications (ISCC) (pp. 204-207). IEEE.

16. Jain, A., Chaurasia, R., Sengar, N. S., Singh, M., Mahor, S., & Narain, S. (2020). Analysis of vitamin D level among asymptomatic and critically ill COVID-19 patients and its correlation with inflammatory markers. Scientific reports, 10(1), 1-8

17. Wu, Z., Chen, X., Zhu, S., Zhou, Z., Yao, Y., Quan, W., & Liu, B. (2013). Enhanced sensitivity of ammonia sensor using graphene/polyaniline nanocomposite. Sensors and Actuators B: Chemical, 178, 485-493.

18. Chen, X., & Yan, G. Y. (2013). Novel human lncRNA–disease association inference based on lncRNA expression profiles. Bioinformatics, 29(20), 2617-2624.

19. Rodríguez-Puebla, A., Behroozi, P., Primack, J., Klypin, A., Lee, C., & Hellinger, D. (2016). Halo and subhalo demographics with Planck cosmological parameters: Bolshoi–Planck and MultiDark–Planck simulations. Monthly Notices of the Royal Astronomical Society, 462(1), 893-916.

20. Eskidere, Ö., Ertaş, F., & Hanilçi, C. (2012). A comparison of regression methods for remote tracking of Parkinson's disease progression. Expert Systems with Applications, 39(5), 5523-5528.

21. Lavesson, N., & Davidsson, P. (2006, July). Quantifying the impact of learning algorithm parameter tuning. In AAAI (Vol. 6, pp. 395-400).

22. Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning (pp. 161-168).

23. Al-Aidaroos, K. M., Bakar, A. A., & Othman, Z. (2010, March). Naive Bayes variants in classification learning. In 2010 international conference on information retrieval & knowledge management (CAMP) (pp. 276-281). IEEE.

24. Asuncion, A., & Newman, D. (2007). UCI machine learning repository.

25. 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.

26. Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B. (2018, August). Implementation of Naïve Bayes classification method for predicting purchase. In 2018 6th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-5). IEEE

27. Parajuli, A. (2016). DISEASE PREDICTOR (Doctoral dissertation, Ph. D. diss., Tribhuvan University).

28. Shankhdhar, A., Agrawal, N. K., & Srivastava, A. (2021, December). COVID-19 Detection System using Chest X-rays or CT Scans. In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART) (pp. 428-432). IEEE.

29. Yadav, A. (2021, July). Predicting Covid-19 using Random Forest Machine Learning

Algorithm. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
30. Bhadana, V., Jalal, A. S., & Pathak, P. (2020, December). A comparative study of machine learning models for COVID-19 prediction in India. In 2020 IEEE 4th conference on information & communication technology (CICT) (pp. 1-7). IEEE.