iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77412

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

A Comprehensive Analysis of Disease prediction Using Machine Learning-Medi Consult

Dr. Devendra[1], Mr. Munish Kumar[2], Shorya Yadav[3]
CSE Department, IFTM University, Moradabad, UP, India

Abstract: This is a thorough and timely investigation of any health-related issue is essential for disease prevention and treatment. The normal way of diagnosis may not be sufficient in the event of a serious illness. Medi Consult is an Android app that predicts illnesses using machine learning algorithms based on symptoms supplied by patients or users. Using the K Nearest Neighbor (KNN) technique, the system calculates the disease's likelihood. Medi Consult's use of linear regression and decision tree algorithms to forecast illnesses such as Diabetes, Malaria, Jaundice, and Dengue assists in accurate and fast disease diagnosis. It provides people with a handy approach to monitor their health and gain early identification of potentially life-threatening disorders. The system's accuracy and quickness in forecasting illnesses aid in prompt medical intervention, resulting in better health outcomes. Furthermore, Medi Consult continually learns and improves its accuracy with each diagnosis, guaranteeing that it is up to speed with the most recent medical research and breakthroughs. Overall, Medi Consult is a helpful tool for both healthcare professionals and patients, enabling early illness identification and treatment and potentially saving lives.
Keywords: Disease prevention, K Nearest Neighbor technique, Accurate diagnosis, Life-threatening disorders

I. INTRODUCTION

Medicine and healthcare are two of the most important aspects of the economy and of human life. There is a significant difference between the world we live in now and the world that existed a few years ago. Everything has become ugly and erratic. In this situation, when everything has gone virtual, doctors and nurses are doing everything they can to save people's lives, even if it means putting their own lives in peril. Virtual doctors are board-certified physicians who choose to practise online through video and phone sessions rather than in-person appointments, albeit this is not practical in an emergency. Machines are always thought to be superior to people because, without human error, they can accomplish jobs more effectively and consistently. A disease predictor is a virtual doctor who can predict any patient's disease without any human error. In illnesses such as COVID-19 and EBOLA, a disease predictor may identify a human's ailment without any physical touch. Healthcare and medicine serve as fundamental cornerstones of human society, playing a vital role in preserving public wellness while supporting the economic and social foundations of countries worldwide. Over the centuries, communities have depended on organized medical systems to fight illnesses, extend life expectancy, and enhance general well-being. The input parameters dataset was supplied to each model while the data was being processed, and the disease was received as an output.

II. LITERATURE REVIEW

The rapid advancement of artificial intelligence (AI) and machine learning (ML) technologies has opened remarkable possibilities in medical diagnostics and predictive healthcare analytics. The primary goal of integrating ML into healthcare practice focuses on enhancing diagnostic accuracy, reducing dependence on manual interpretation, optimizing clinical decision-making processes, and supporting medical professionals through automated analytical systems. Many researchers have explored how machine learning models can detect diseases by examining patient data, clinical indicators, laboratory results, and medical imaging outcomes. ML technologies have transformed from supporting practitioners in routine clinical tasks to becoming autonomous decision-assistance tools that can exceed conventional diagnostic approaches in various specialized areas.

S. S. Mishra et al. [1], presented a method that is utilized to identify illnesses and this study analyzed multiple machine learning algorithms for disease prediction, including logistic regression, decision tree, random forest, k-nearest neighbor, and support vector machine. The dataset for the study was obtained from the National Institute of Diabetes and Digestive and Kidney Diseases. The random forest method outperformed all others in terms of accuracy, sensitivity, and specificity.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

N. Raj Kumar et al. [2] focused on utilising machine learning algorithms to detect cardiac disease in electronic health data. The authors employed logistic regression, random forest, and gradient boosting methods to predict heart disease. The random forest method outperformed the others in terms of accuracy, with an AUC of 0.901

Using electronic health information, A. Alaa et al. [3] suggested an ensemble learning strategy for diabetes prediction. Six machine learning techniques were utilized by the authors: logistic regression, decision tree, random forest, k-nearest neighbour, gradient boosting, and neural network. In terms of accuracy, sensitivity, and specificity, the results demonstrated that the ensemble learning strategy beat individual algorithms.

M. A. Siddiqui et al. [4] used machine learning techniques to predict Parkinson's disease. Four algorithms were utilised by the authors: decision tree, random forest, support vector machine, and artificial neural network. The study's dataset was received from the University of California, Irvine. The random forest method outperformed all others in terms of accuracy, with a score of 95.6%

A. Z. A. Al-Zubaidi et al. [5] examined several machine learning approaches for breast cancer identification using histopathology pictures. The authors explored the benefits and drawbacks of several algorithms, such as decision trees, support vector machines, artificial neural networks, and convolutional neural networks. Convolutional neural networks beat other methods in terms of accuracy and precision, according to the study.

Cho, D., Kim, D., & Choi, S. [6] Various machine learning algorithms were utilised in this study to predict the risk of metabolic syndrome in Korean individuals. Deep learning models were shown to be the most accurate, followed by random forest and logistic regression models.

Ghosh, S., Mitra, S., & Saha, S. [7] The application of machine learning and deep learning algorithms for COVID-19 diagnosis utilising chest X-ray and CT scan images is discussed in this review paper. They discovered that these algorithms are capable of providing accurate and speedy diagnosis, which is especially beneficial during pandemics.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. [10] Explain Deep neural networks were utilised to achieve dermatologist-level categorization of skin cancer. The scientists discovered that their algorithm accurately identified skin cancer, which might possibly enhance early identification and treatment of this illness.

III. METHODOLOGY

To evaluate performance in the experiment, we first denote TP, TN, Fp and FNias true positive (the number of results correctly predicted as required), true negative (the number of results not necessary), falsely positive (the number of results incorrectly predicted as required), and false negative (the number of results incorrectly predicted as not required), respectively. As follows, we may obtain four measurements: recall, precision, accuracy, and F1:

Accuracy:-

$$\frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + False\,Positive + FalseNegative}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$F1\text{-}Measure = \frac{2 * precision * recall}{precision + recall}$$

Fig.1 Accuracy formula

In unambiguous border datasets between classes or when the data is well-separated into different clusters, KNN outperforms other ML algorithms in terms of accuracy. Furthermore, KNN is useful when the training data is little, as it does not require a big quantity of data to work effectively. Which is the most appropriate for our application environment and dataset.

A. System Architecture

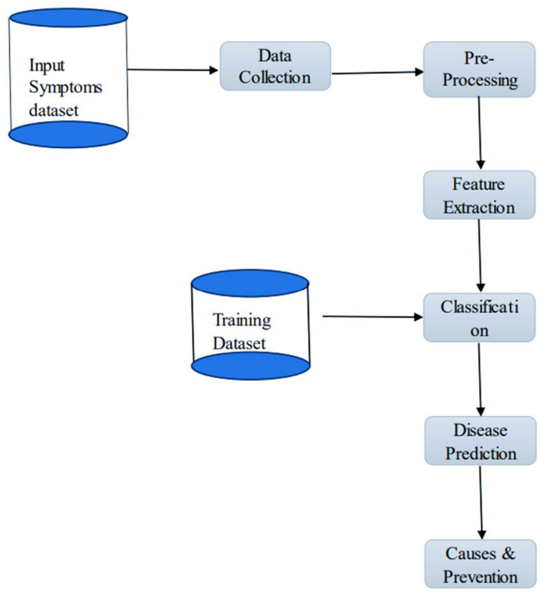


Fig.2 Proposed System Architecture

1) Input symptom dataset: The initial stage in the system design is to enter the symptom dataset, which comprises information on various symptoms and the diseases associated with them.
2) Data collection: The system then collects data on the symptoms and diseases linked with them from a variety of sources, including medical databases, research articles, and web sites.
3) Pre-processing: Once the data has been gathered, it must be cleaned and prepared for analysis. This process may include deleting duplicates, dealing with missing values, and translating the data into an analysis-ready format.
4) Following pre-processing, the system retrieves important characteristics from the symptom dataset. This may include characteristics such as symptom intensity, duration, and frequency.
5) Classification: The next stage is to employ KNN for classification, which entails categorising each symptom based on its characteristics. KNN works by identifying the k-closest neighbours to a particular symptom in a training dataset and then predicting the illness label based on the majority vote of its nearest neighbours' labels.
6) Training dataset: In order to execute classification using KNN, the system requires a training dataset that comprises labelled data on symptoms and illnesses. This dataset is used to train and fine-tune the KNN algorithm's parameters.
7) Disease prediction: Once trained, the KNN algorithm may be used to predict the illness associated with a given collection of symptoms. This forecast is based on the characteristics derived from the symptoms and the KNN classification model.
8) Finally, based on medical expertise and information in the symptom dataset, the system may give information on the causes and prevention of the anticipated condition. This information, together with the projected illness and any other relevant information, can be displayed to the user as part of the system output.

IV. ALGORITHM TECHNIQUE

K Nearest Neighbour (KNN) is a very basic, easy-to-understand, adaptable, and one of the best machine learning algorithms. The illness will be predicted by the user in the Healthcare System. In this approach, the user may anticipate whether or not the sickness will be detected.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

In the proposed method, disease is classified into numerous groups that indicate which disease will occur based on symptoms. For each regression and classification problem, the KNN rule is employed. The KNN method is based on the feature similarity technique. It is the finest option for tackling various classification-related issues. By specifying the nearest neighbour class, the K-nearest neighbour classifier method predicts the target label of a new instance. Distance metrics such as Euclidean distance will be used to determine the closest class. If K = 1, the instance is simply allocated to the category of its closest neighbour.

The value of 'k' must be given by the user, and the optimal option is determined by the data. A higher 'k' number minimizes noise in the categorization. If the new feature, in this example the symptom, must be classified, the distance is computed and the class of feature closest to the newer occurrence is chosen. The Hamming distance must be employed when dealing with categorical data. It also raises the issue of standardisation of numerical variables between zero and one when the dataset contains both numerical and category variables.

1) Using a distance measure such as Euclidean distance or Manhattan distance, compute the distance between the new sample and each sample in the training set.
2) Sort the training data samples in increasing order by distance to the new sample.
3) Based on the distance ranking, choose the K nearest neighbours to the new sample.
4) Assign the new sample to the class with the highest frequency of occurrence among its K closest neighbours.

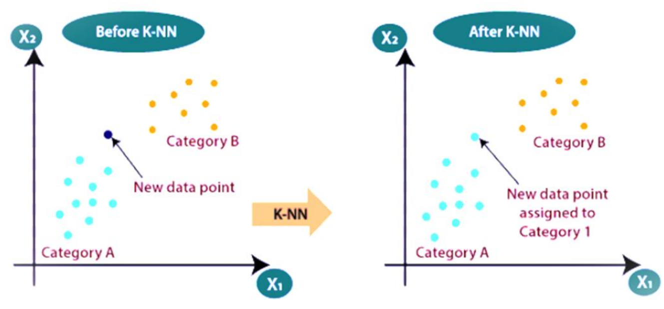


Fig.3 Data processing using KNN

T = (si1, si2,....sin, di) is a training dataset consisting of n symptoms and their accompanying illnesses.

Using a distance metric such as Euclidean distance or Manhattan distance, compute the distance between the new symptoms s and all symptoms in the training dataset T. The distance formula is as follows:

$$d(s, si) = sqrt(sum((si1 - s1)^2 + (si2 - s2)^2 + ... + (sin – sn)^2))$$

The k-nearest neighbors to the new symptoms are then selected based on the calculated distances, and the majority class among the k-nearest neighbors is assigned as the predicted disease.

ML Technique	Accuracy	Precision	Recall	F1-score	AUC-ROC
Logistic Regression	0.85	0.87	0.81	0.84	0.90
Decision Tree	0.81	0.79	0.85	0.82	0.83
Random Forest	088	0.89	0.86	0.87	0.92
K-Nearest Neighbors	0.90	0.92	0.88	0.90	0.94

Table1 Comparison of KNN with another ML algorithms

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

V. RESULTS

Different machine learning models were employed to investigate illness prediction using accessible input datasets. The Weighted KNN model achieved the maximum accuracy of 93.5%. The accuracy is high since the weighted KNN was high because the value of K changed in this model. This number varies depending on our dataset, for example, it was tiny and huge for the training set. Because of this variance, it proved to be the most accurate model among the other ML methods. We gathered raw data and classified it according to gender, age group, and symptoms. As the application output shows in fig 5
it will predict the disease and gives the Causes and prevention.

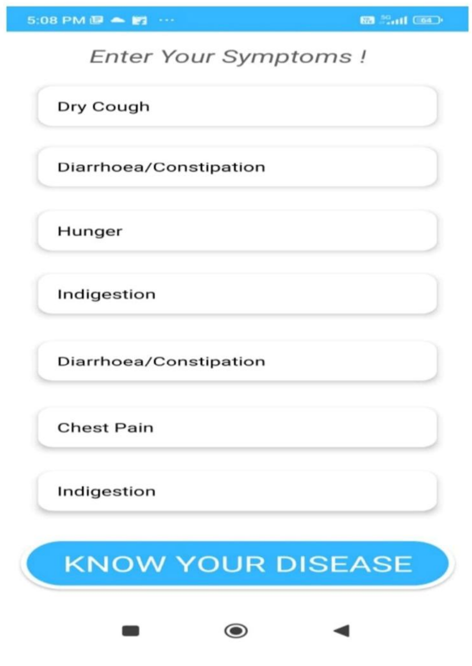


Fig.4 Input Symptoms

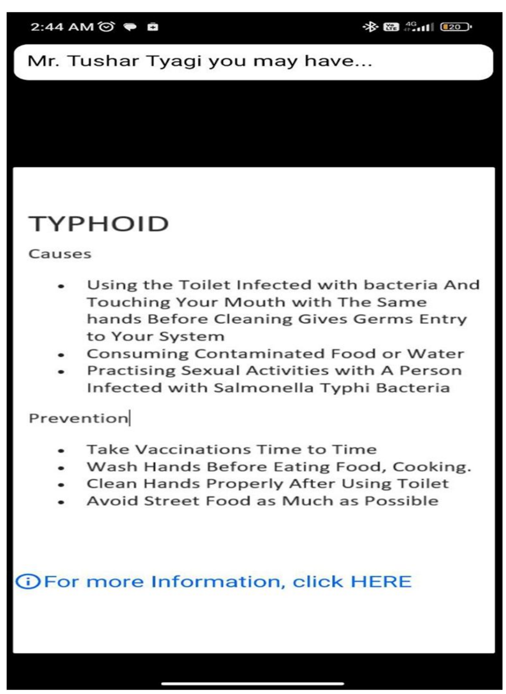


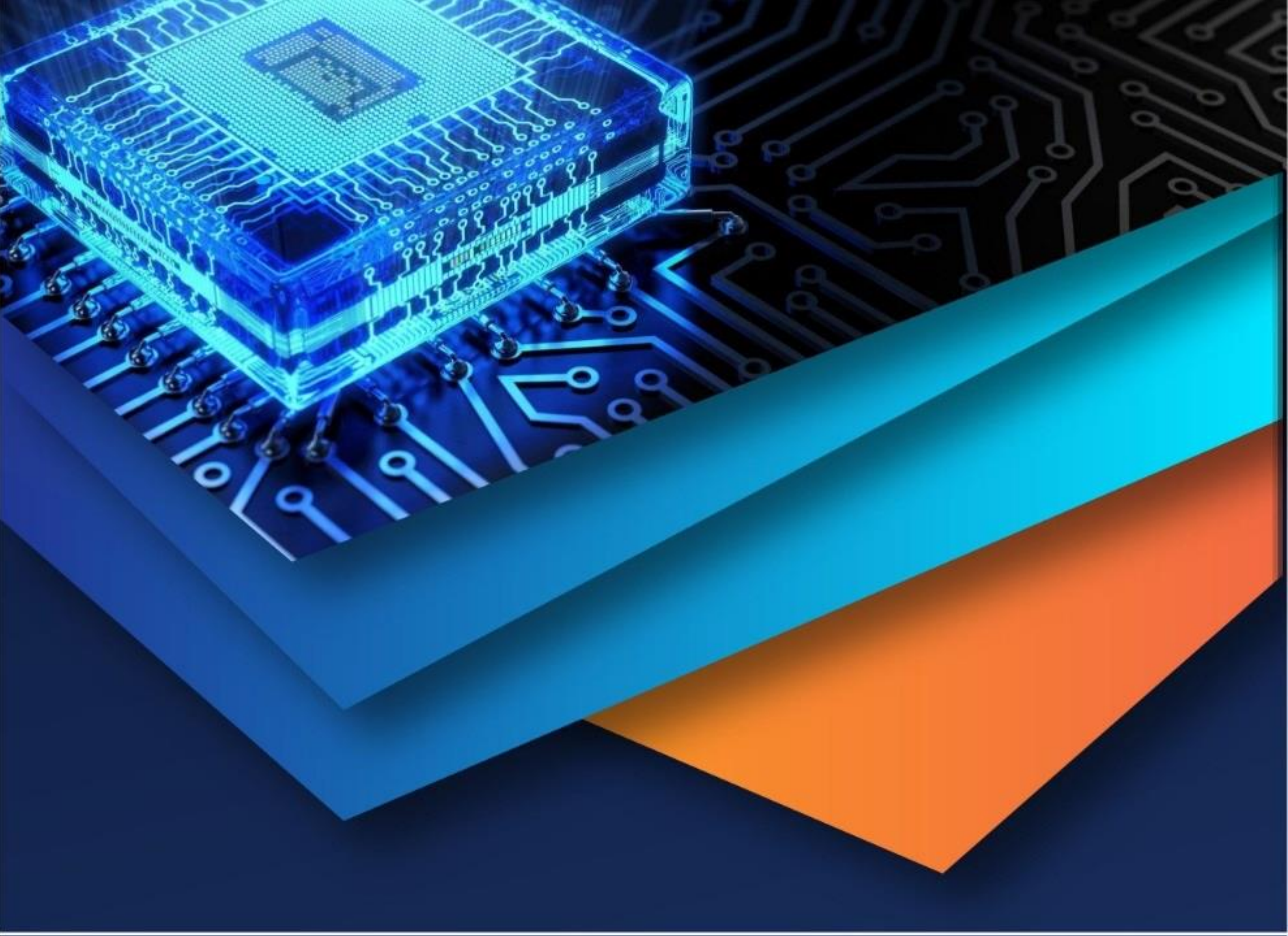
Fig.5 Output

VI. CONCLUSION

The publication described a strategy for predicting illness based on a patient's symptoms, age, and gender. The Weighted KNN model produced .The above-mentioned parameters resulted in the maximum accuracy of 93.5% for disease prediction. Almost all of the ML models produced good accuracy results. Because some models were parameter dependant, they were unable to predict the illness and had a poor accuracy rate. We could simply manage the medical resources necessary for therapy once the sickness was predicted. This concept would assist to reduce the expense of treating with the sickness while also improving the healing process.

REFERENCES

[1] "A Comparative Study of Machine Learning Algorithms for Disease Prediction" by S. S. Mishra (2021)J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[2] "Machine learning-based prediction of heart disease using electronic health records" by N. Rajkomar (2018)
[3] "An Ensemble Learning Approach for Diabetes Prediction Using Electronic Health Records" by A. Alaa (2018)
[4] "Prediction of Parkinson's disease using machine learning algorithms" by M. A. Siddiqui et al. (2020)
[5] "Machine Learning Techniques for Breast Cancer Diagnosis Using Histopathology Images: A Review" by A. Z. A. Al-Zubaidi et al. (2021)
[6] .Cho, D., Kim, D., & Choi, S. (2020). Predicting the risk of metabolic syndrome using machine learning and artificial intelligence. Journal of Personalized Medicine, 10(3), 97
[7] Ghosh, S., Mitra, S., & Saha, S. (2021). Machine learning and deep learning-based COVID-19 diagnosis using chest X-ray and CT scan images: A review. Journal of X-Ray Science and Technology, 29(3), 483-495.
[8] Nwankwo, E. O., Adebesin, O. A., Adetunji, A. E., Ogunmefun, O. T., & Ojo, S. A. (2021). Machine learning models for malaria prediction: A review. Heliyon, 7(7), e07459.
[9] Zame, W. R., & Javaheri, A. (2020). Machine learning in healthcare: Overview, trends, and future directions. Journal of Medical Systems, 44(8), 1-18.
[10] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.
[11] Galloway, M., & Hamilton, J. (2019). Machine learning in medical imaging. Biomedical Physics & Engineering Express, 5(1), 1-20.
[12] Jang, S. M., Song, I., Cho, J. W., Kim, B. H., & Kim, H. (2020). Diagnosis of Alzheimer's disease using machine learning algorithms on MRI images: A systematic review. Healthcare, 8(1), 37
[13] Chen, C., Huang, J., Cheng, Z., Lu, X., & Yang, J. (2021). Applications of machine learning in the prediction of mental health disorders: A review. Journal of Medical Systems, 45(6), 1-11.
[14] Gomathi, N., & Sumathi, P. (2021). An overview of machine learning algorithms for predicting disease outbreak. Health and Technology, 11(3), 757-768.
[15] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
[16] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. Seminars in Cancer Biology, 52, 228-243.
[17] Liao, K., Xie, Y., Liu, Y., Chen, Y., & Huang, Y. (2020). Deep learning-based disease diagnosis with genetic data. IEEE Transactions on Biomedical Engineering, 67(1), 141-153.
[18] Zhou, X., Zeng, X., Yang, J., Zhang, X., Wang, J., Liu, Z., ... & Gao, X. (2019). Deep learning: A review from a structure perspective. Neurocomputing, 335, 103-111.
[19] Zhou, Y., Liu, Y., Liu, L., & Wu, J. (2021). A systematic review of machine learning-based diagnosis models for COVID-19. Frontiers in Public Health, 9, 1-10.
[20] Chen, Z., Chu, C., Zheng, W., Zhang, Y., Han, Y., & Cai, Y. (2020). Early warning system for sepsis using machine learning: A systematic review. Journal of Healthcare Engineering, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY