Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

Multiple Disease Risk Analyzer System

Amit Kumar, Pulkit Raj Saxena, Mohit Kumar Mishra, Brahmdutt Gaur, Mohan Vishal Gupta, and Anushka Agarwal

CCSIT, Teerthanker Mahaveer University, Moradabad, India
Corresponding author e-mail: pulkitr.computers@tmu.ac.in

***Abstract*—Early detection of long-term and communicable diseases is important because timely treatment plays a major role in human life; if risks are reduced, the overall cost of healthcare will also decrease. In this study, we propose a multiple disease prediction system that works in a combined manner, in which machine learning and medical image analysis are applied to patient data. The model can predict Type-2 diabetes, anemia, and coronary artery disease from clinical records, and it can also detect pneumonia and tuberculosis from chest X-ray images. The system starts with basic preprocessing, which includes data cleaning and feature normalization. Different pipelines are then used according to the dependencies involved, and the system produces its output in probability form; according to this, it is easy to understand the model behavior. A supervised learning model is used to find the relationship between the dataset and the present health features, and a convolutional neural network is used to find patterns in medical images such as chest X-rays. Finally, the system predicts a risk score and short summaries, which help people with early screening and basic decision support. The system also includes extra features such as a login system, a profile section, and the ability to check previous reports and download results as needed; these features help the user maintain a health record with less effort. The system works with numerical data and medical images, which improves the results. The total setup of CureHelp+ is quite simple, and users can operate it easily without any extra training. This feature is mainly made for areas where large medical facilities are not available. It is used as a quick, basic screening tool, with no need to visit a specialist. It also helps doctors and staff in busy situations, so that they can attend to serious patients. Overall, this platform provides a simple and scalable way to support the healthcare system, and it brings machine learning methods into real clinical use in practical situations.**

***Index Terms*—CNN, healthcare system, logistic regression, ML, SVM.**

I. Introduction

Getting a medical check-up on time is a big problem in many developing areas. People in these areas always have to travel long distances in search of a multi-specialist hospital, whose charges are very high, and it often happens that doctors are not easily available; that is why the diagnosis process starts late [1]–[3]. That is why common health problems such as anemia, Type-2 diabetes, recurring fever, and heart disease are not detected in the early stages. They are felt only when the patient reaches a serious condition. At that time, treatment becomes difficult, and recovery is not easy. This problem mainly occurs in village areas and among middle-class families, where medical facilities are not easily available.

The healthcare system mostly depends on hospital visits, test reports, and doctor suggestions; this process is time-consuming and very costly. In many places, this process is very slow because facilities and doctors are not easily available [4]. Some digital apps are now available, but most apps focus on a single disease. Many times, their output results are complex, so normal people cannot understand them easily. That is why our system needs to be a simple system that detects diseases in their early stages, with low cost, less confusion, and ease of use, and that can give a basic idea of health.

CureHelp is a simple website made for basic health checkups. For the backend, we use Flask, and for the frontend, we use HTML, CSS, and JavaScript. Its main goal is that a person from a non-technical background can use it easily [5], [6]. The user can enter basic details such as age, gender, hemoglobin level, blood sugar, body temperature, and blood pressure. The system then processes that data and checks common health problems such as anemia, diabetes, fever, or heart disease to find the risk factor. The system simply works on IF-THEN rules taken from general guidelines. It is not a complex or advanced system. Its logic is simple and direct, which helps the user understand easily how the result is generated before looking at the final output [6]–[9]. This platform is designed especially for students, lower-resource communities, and first-level screening situations in which quick guidance is important. It gives a clear output such as low, medium, or high, and it also gives simple suggestions such as improving iron intake, monitoring glucose, or consulting the nearest doctor. Its modules work on a rule-based structure, so new diseases can be added according to different regions and modified, and the system can be expanded for offline use [10], [11]. To bridge the gap between symptoms and proper medical action, the primitive behavior of CureHelp plays a strong role. It improves timely doctor consultation and improves awareness at the community level, without replacing professional diagnosis or doctor-led treatment.

II. Literature Review

A study showed that deep learning techniques can improve the accuracy of cardiac MRI segmentation. However, models trained on single-center data need better generalization. The M&Ms Challenge introduced multi-center datasets, but there is still a need for generalization across different imaging conditions [1], [2].

An investigation by G. Parimala *et al.* (2023) applied machine learning algorithms such as KNN and ensemble classification for predicting and diagnosing diabetes. This study showed that, by using big data in the field of healthcare, diseases can be identified in a more efficient manner. Still, the performance of the model depends on the quality of the data [3], [4]. In a study by M. Kavitha *et al.* (2019), a hybrid

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

machine learning model based on Random Forest and Decision Tree algorithms was suggested for predicting heart disease. The predictive model showed 88.7% accuracy on the Cleveland dataset. However, this model is limited by its dependence on a particular dataset and may not perform well on the different kinds of data that are available in practice [4], [5].

In the study by A. Mohammad *et al.* (2024), the authors proposed a hybrid deep learning approach using both CNN and RNN models for pneumonia detection from chest X-ray images. The authors suggested integrating the spatial and temporal features of the model, which resulted in 98.5% accuracy using the CNN-LSTM model. However, the models are computationally expensive and sometimes have difficulty generalizing [6], [7]. In a study by Josef Yayan *et al.* (2020), the authors reviewed various approaches for the early detection of tuberculosis, including deep learning-based X-ray analysis, interferon-gamma release assays, and nucleic acid amplification tests. The study highlighted the significance of early screening and the application of advanced detection tools to improve the detection rate of the disease. However, the approaches reviewed involved the use of multiple detection tools and lacked a single automated system for detection [8].

TABLE I
SUMMARY OF LITERATURE REVIEW

S. No.	Author Name	Year	Methodology Used	Findings	Limitations	Accuracy
1	Valter M. Campello	2021	CNN for cardiac MRI segmentation (deep learning-based medical image analysis)	Effective segmentation of cardiac structures for heart disease assessment	Models trained on single-center datasets show poor generalization across different scanners and imaging conditions	~85–95% (varies by challenge results)
2	G. Parimala *et al.*	2023	K-Nearest Neighbors (KNN) and ensemble learning-based classification (supervised ML)	Diabetes prediction using structured clinical data	Performance depends heavily on dataset quality and feature selection; lacks real-time clinical validation	~80–90%
3	M. Kavitha *et al.*	2019–2023	Hybrid machine learning model (Random Forest + Decision Tree ensemble)	Improved prediction of heart disease using ensemble techniques	Limited interpretability and dependency on dataset quality; lacks deployment validation	88.7%
4	A. Mohammad	2024	Hybrid deep learning model combining CNN (feature extraction) and LSTM (sequence learning)	High-performance pneumonia detection from medical imaging data	Requires large labeled datasets and high computational resources; risk of overfitting and limited generalization	98.5%
5	Josef Yayan *et al.*	2020	Systematic review of deep learning-based image analysis and clinical diagnostic techniques (IGRA, NAAT)	Comprehensive analysis of tuberculosis detection methods	Lack of a unified automated system; high cost and dependency on multiple diagnostic methods	Not fixed (review-based study)
6	Dheeru Dua and Casey Graff	2017	Logistic Regression on diabetes dataset (UCI ML Repository)	Baseline diabetes prediction using structured health data	Limited dataset diversity; not fully representative of real-world populations	~76–82%
7	Robert Detrano *et al.*	1989	Statistical machine learning methods on heart disease dataset	Early prediction of heart disease using clinical parameters	Outdated dataset with limited features; lacks modern validation techniques	~80–85%
8	Pranav Rajpurkar	2017	Deep learning (DenseNet-based CNN model, CheXNet) for X-ray analysis	State-of-the-art automated pneumonia detection from chest X-rays	High computational cost and large dataset requirements	~90–93%

III. METHODOLOGY

In the proposed work, a multi-disease patient dataset is analyzed with proper data processing. For training and prediction, we use multiple machine learning algorithms, such as Decision Tree, Random Forest, SVM, and Logistic Regression. The model is designed so that it can predict multiple diseases at a time; for each disease, we can use the same model. To handle data bias, the system is already trained on a diverse dataset that includes underrepresented groups, so that it can improve its prediction accuracy for different types of people. Algorithms that avoid overfitting and give better results on new data are used in this model, with techniques such as regularization and cross-validation. The system is designed in such a way that a normal person can easily understand its prediction; that is why we use explainable AI, in order to know the real reason behind a result. Through this, the system becomes trusted and easily accepted by doctors and patients.

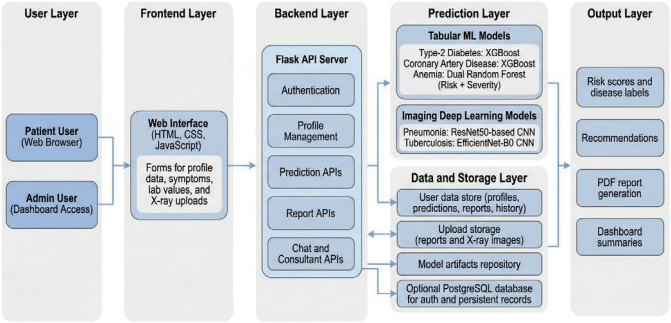


Fig. 1. Simplified architecture of the CureHelp+ multi-disease risk analyzer system.

A. Machine Learning Algorithms Used

TABLE II
MACHINE LEARNING ALGORITHMS USED

S. No.	Input Type	Disease Name	ML Algorithm	Purpose
1	Tabular (Clinical)	Type-2 Diabetes	Gradient Boosting – XGB Classifier	To estimate diabetes risk from demographic and metabolic indicators for early screening.
2	Tabular (Clinical)	Coronary Artery Disease	Gradient Boosting – XGB Classifier	To predict cardiovascular risk using key clinical and hemodynamic features.
3	Tabular (Clinical)	Anemia	Two Random Forest classifiers (Risk + Severity)	To compute anemia risk and categorize severity/type from hematological parameters.
4	Imaging (X-ray)	Tuberculosis	EfficientNet-B0 CNN	To detect tuberculosis-related abnormalities in chest X-ray images.
5	Imaging (X-ray)	Pneumonia	ResNet50-based CNN	To identify pneumonia patterns in chest X-ray images for rapid triage support.

B. Dataset

TABLE III
TABULAR DATASETS USED TO TRAIN, TEST, AND VALIDATE THE MODELS

S. No.	Disease	Size of Dataset
1	Type-2 Diabetes	5733 Rows
2	Coronary Artery Disease	4556 Rows
3	Anemia	5660 Rows

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

TABLE IV
IMAGING (CHEST X-RAY) DATASETS USED TO TRAIN, TEST, AND VALIDATE THE MODELS

S. No.	Disease	Size of Dataset
1	Pneumonia	8009 Images
2	Tuberculosis	4200 Images

IV. RESULTS

The CureHelp+ system gives stable and reliable performance in its disease prediction models, and it can provide quick results in normal usage. Its unified interface gives the user the ability to check multiple conditions in one place, with no need to check each condition separately. The system has strong validation, namely its input validation, in which it checks all of the user's data to determine whether they are correct or incorrect. If any value is wrong or missing, the system shows a red flag on it, through which errors are reduced and system reliability improves. When we give input to the system, it generates a clear and easy output. Overall, this framework is user-friendly and well suited to early-stage health risk assessment, which will help us make better and timely decisions.

TABLE V
COMMON TERMS USED FOR DISEASE PREDICTION

S. No.	Name of Attributes	Use	Range
1	Age	General risk factor across Diabetes/CAD	1–120
2	Gender / Sex	Biological risk stratification and model encoding	Male / Female
3	Body Mass Index (BMI)	Obesity and metabolic-risk indicator	10–70
4	Glucose	Blood sugar status for diabetes risk	40–400
5	Blood Pressure	Metabolic and vascular stress indicator (diabetes module)	60–250
6	Pregnancies	Female diabetes-risk context	0–20
7	Skin Thickness	Body fat/insulin resistance proxy	5–80
8	Insulin	Glucose regulation marker	15–900
9	Diabetes Pedigree Function	Family-history diabetes risk score	0–3
10	Resting Blood Pressure (Resting BP)	Baseline cardiac pressure indicator (CAD)	80–220
11	Cholesterol	Lipid-based heart risk marker	100–600
12	Chest Pain Type	CAD symptom category	1 / 2 / 3 / 4
13	Fasting Blood Sugar > 120	Hyperglycemia flag for CAD	Yes / No (encoded 1/0)
14	Resting ECG	Rest-state electrical heart pattern	0 / 1 / 2
15	Max Heart Rate	Cardiac exertion capacity marker	60–220
16	Exercise Angina	Exertion-induced chest pain marker	Yes / No (encoded 1/0)
17	ST Depression	Ischemia-related ECG feature	0–10
18	Slope of ST	Stress ECG trend category	1 / 2 / 3
19	Major Vessels (ca)	Number of affected major vessels	0–3
20	Thal	Perfusion/thal category in CAD model	3 / 6 / 7
21	RBC	Red blood cell count marker	2–8
22	Hemoglobin	Oxygen-carrying capacity marker	6–20
23	Hematocrit	RBC volume proportion marker	20–60
24	MCV	Average RBC size marker	60–110
25	MCH	Hemoglobin per RBC marker	15–40
26	MCHC	Hemoglobin concentration in RBC marker	25–38
27	WBC	Infection/immune activity marker	2–30
28	Platelets	Clotting function marker	50–1000
29	RDW	RBC size-variation marker	10–20
30	PDW	Platelet size-variation marker	5–25
31	PCT (Plateletcrit)	Total platelet mass fraction marker	0.05–0.6
32	Lymphocytes	Adaptive immune-response marker	5–60
33	Neutrophils (%)	Infection-defense proportion marker	20–80
34	Neutrophils (#)	Absolute neutrophil count marker	1–10
35	Chest X-ray Image	Input for Pneumonia and TB models	JPG/JPEG/PNG, up to 10 MB (processed to 224 × 224)

A. Models Summary

1) Classification Report

This classification report shows that the models perform strongly in both classes, and it shows that they easily detect pneumonia cases. The models maintain better precision and recall.

TABLE VI
CLASSIFICATION REPORT FOR THE DISEASES IN THE PREDICTION SYSTEM

S. No.	Disease Name	Input Size	Class	Precision	Recall	F1-Score	Accuracy
1	Type-2 Diabetes	1 × 8	0	97	95	96	95.20
			1	93	96	94	
2	Coronary Artery Disease	1 × 13	0	95	93	94	93.68
			1	93	95	94	
3	Tuberculosis	224 × 224 × 3	0	94	98	96	94
			1	92	80	82	
4	Pneumonia	224 × 224 × 3	0	94	86	90	90
			1	89	95	92	
5	Anemia	1 × 13	0	94	98	96	94.71
			1	95	88	91	

V. CONCLUSION

Due to changing lifestyle and environmental conditions, people face multiple health problems. It is very important to detect any disease in its early stage, before it becomes severe, but the diagnosis of multiple diseases is very difficult and time-consuming for doctors. Machine learning provides an effective way to solve this problem because it can analyze large amounts of data and identify patterns that are not easily visible; this supports early diagnosis and increases the chances of better treatment. As a result, patients will get better and more accurate care. However, in healthcare, there are several challenges with machine learning, such as the need for unbiased and good-quality datasets; ongoing research and the use of new technology will continue to improve this field. In the coming era, machine learning in healthcare will become an important part of multi-disease prediction, which will help doctors in decision-making and make it easy to adjust the treatment of every patient.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

REFERENCES

[1] G. Parimala, R. Kayalvizhi, and S. Nithiya, "Diabetes prediction using machine learning," in *Proc. 2023 Int. Conf. Comput. Commun. Informat. (ICCCI)*, Coimbatore, India, 2023, pp. 1–6.

[2] M. Kavitha, G. Ganeswar, R. Dinesh, Y. Rohith Sai, and R. Sai Suraj, "Heart disease prediction using hybrid machine learning model," in *Proc. IEEE Conf.*, 2019, pp. 1–5.

[3] A. Mohammad, G. Suryakala Eswari, P. Ratna Kumari, A. Lakshmanarao, and D. Chandra Mouli, "A hybrid CNN-RNN model for enhanced pneumonia detection using X-ray imaging," in *Proc. 2024 Int. Conf. Softw., Syst. Inf. Technol. (SSITCON)*, 2024, pp. 1–6.

[4] J. Yayan, K.-J. Franke, M. Berger, W. Windisch, and K. Rasche, "Early detection of tuberculosis: A systematic review," *Springer Nature*, 2020.

[5] D. Dua and C. Graff, "UCI machine learning repository: Pima Indians diabetes dataset," Univ. California, Irvine, CA, USA, 2017.

[6] R. Detrano, A. Janosi, W. Steinbrunn *et al.*, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *Amer. J. Cardiol.*, 1989.

[7] D. S. Kermany, M. Goldbaum, W. Cai *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, 2018.

[8] P. Rajpurkar, J. Irvin, K. Zhu *et al.*, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.

[10] D. J. Park, M. W. Park, H. Lee, Y.-J. Kim, Y. Kim, and Y. H. Park, "Development of machine learning model for diagnostic disease prediction based on laboratory tests," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021.

[11] Y. Perwej, "The bidirectional long-short-term memory neural network based word retrieval for Arabic documents," *Trans. Mach. Learn. Artif. Intell. (TMLAI)*, Soc. Sci. Educ., U.K., vol. 3, no. 1, pp. 16–27, 2015, ISSN 2054-7390, doi: 10.14738/tmlai.31.863.

[12] Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "A machine learning approach to predicting diabetes complications," *Healthcare*, vol. 9, no. 12, Art. no. 12, Dec. 2021, doi: 10.3390/healthcare9121712.

[13] M. Anila and G. Pradeepini, "A review on Parkinson's disease diagnosis using machine learning techniques," *Int. J. Eng. Res. Technol.*, vol. 9, no. 6, Jun. 2020.

[14] Y. Perwej, "Unsupervised feature learning for text pattern analysis with emotional data collection: A novel system for big data analytics," in *Proc. IEEE Int. Conf. Adv. Comput. Technol. Appl. (ICACTA)*, Coimbatore, India, 2022, ISBN 978-1-6654-9515-8, doi: 10.1109/ICACTA54488.2022.9753501.

[15] J. Cao, M. Wang, Y. Li, and Q. Zhang, "Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment," *PLoS One*, vol. 14, no. 4, 2019.

[16] Y. Perwej, S. A. Hann, and N. Akhtar, "The state-of-the-art handwritten recognition of Arabic script using simplified fuzzy ARTMAP and hidden Markov models," *Int. J. Comput. Sci. Telecommun. (IJCST)*, no. 8, pp. 26–32, 2014, ISSN 2047-3338.

[17] H. Hamidi and A. Daraee, "Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases," *Int. J. Eng., Trans. B, Appl.*, vol. 29, no. 7, pp. 921–930, 2016.

[18] M. Maurano *et al.*, "Systematic localization of common disease-associated variation in regulatory DNA," *Science*, vol. 337, pp. 1190–1195, 2012, doi: 10.1126/science.1222794.

[19] A. Kumar, "Disease prediction and doctor recommendation system using machine learning approaches," *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)*, 2021, doi: 10.22214/IJRASET.2021.36234.

[20] Y. Perwej, F. Parwej, and N. Akhtar, "An intelligent cardiac ailment prediction using efficient ROCK algorithm and K-means & C4.5 algorithm," *Eur. J. Eng. Res. Sci. (EJERS)*, vol. 3, no. 12, pp. 126–134, 2018, doi: 10.24018/ejers.2018.3.12.989.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)