



# Disease Detection and Risk Prediction System Based Web Application Using Machine Learning

Raj Kumar<sup>(✉)</sup>, Ujjawal Singh, Soumya Sahoo, Ipsita Das, and Prashant Kr. Jha

Department of Computer Science and Engineering, C.V. Raman Global University,  
Bhubaneswar, Odisha, India  
ayerajkumar@gmail.com

**Abstract.** The purpose of the website We created is to provide consumers with health and disease predictions based on machine learning algorithms. The website is made with the user in mind, offering a straightforward and intuitive interface that makes it simple for users to get about the platform and access their health report and self-health evaluation.

The website's machine learning algorithms are made to assess user information, such as medical history, dietary preferences, and physical attributes, in order to produce a unique health report for each user. The health report will provide details on potential dangers and symptoms of various illnesses, as well as suggestions for safeguards and lifestyle modifications that might help lower such dangers.

The website also has a self-health assessment tool that lets users provide more details about their personal health and way of life. The machine learning algorithms will use this data to produce a more thorough health report that is customised to the user's requirements.

**Keywords:** Healthcare · Self-Assessment Tool · Diagnose · Machine Learning Algorithm · Web Application for Health and Disease Detection · Decision Tree · Django Backend

## 1 Introduction

Humans have long placed a high value on their health and well-being. Nonetheless, there has been a huge shift in how people view their health as a result of the development of technology and the rising demand for rapid gratification. Due to their capacity to correctly identify ailments and offer individualised healthcare solutions, machine learning algorithms have recently experienced tremendous growth in popularity in the healthcare sector.

The website we show in this study uses machine learning techniques to forecast ailments and give consumers a detailed health report. By giving users the resources, they need to take charge of their health and make wise decisions about their well-being, our website seeks to empower people. Users can detect potential health issues and take preventive action before it's too late thanks to our self-health assessment tool.

Our research article intends to offer a thorough description of the website construction process, algorithm selection method, and validation techniques. We think that by offering a straightforward and approachable platform for disease prediction and self-health evaluation, our website has the potential to transform the healthcare sector. One of the primary advantages of our website is that it enables users to take charge of their health. By offering simple access to individualised health reports and self-assessment tools, users are able to recognise potential health hazards and take preventative action before to the onset of serious illnesses. This can lower healthcare expenses and improve the quality of life overall. Our idea is unusual because it combines the most recent advancements in machine learning algorithms with a user-friendly website design to create an accessible and efficient healthcare solution. We think that by using the power of technology, our website has the ability to revolutionise the healthcare business by making it more personalised, convenient, and affordable.

The work done in this research aims to provide a solution for self-assessment and disease prediction using machine learning algorithms. The project has successfully designed and built a web application that uses machine learning models in the backend to process the data for prediction of disease based on trained models. Our web application offers two major services, first one is 'Self-Assessment' for the health examination and another one is 'Diagnosis Section' to test the blood report of patient corresponding to suffering disease to know whether they are in risk or safe.

## 2 Literature Survey

*Heart Health Prediction Using Web Application.* It uses machine learning, classification, and data analysis techniques to predict heart conditions based on various parameters such as glucose level, BP, BMI, etc. The model achieved an accuracy of 75-82% using a random forest classifier and is presented as a web application using Streamlit [1].

*Heart Disease Prediction Using Machine Learning and Data Mining* here the proposed method for detecting heart disease using machine learning algorithms such as K-Nearest Neighbors, Naïve Bayes, Decision Tree, Support Vector Machines, and Random Forest. The models are uploaded to a web server using Flask framework, and users can input their data. The authors conducted experiments and found that K-Nearest Neighbors provides the highest accuracy of 87%. They also discuss previous studies on heart disease detection using machine learning and data mining techniques [2].

*Disease Prediction Web App Using Machine Learning.* The study evaluates different models such as Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier, and uses Django to create a self-assessment system based on the best accuracy model. The objective is to improve healthcare through accurate and early disease detection. The proposed system, called HealthSure, allows users to register and check their symptoms to predict diseases [3].

*Heart Disease Prediction Using Machine Learning,* The proposal for a web application that aims to predict the occurrence of heart disease and suggest preventative measures. The project will use data mining techniques to extract hidden patterns from datasets and find a suitable machine learning technique for heart disease prediction. The objectives are

to develop an easy-to-use platform, require no human intervention, suggest preventative measures. The project aims to improve medical efficiency, reduce costs, and enhance the quality of clinical decisions [4].

*Disease Prediction Using Machine Learning and Django and Online Consultation.* It is an online platform for users to receive quick medical advice based on their symptoms. The system uses data mining techniques to determine the most likely disease related to a patient's symptoms. Doctors can access patient information and diagnose them online. The system uses the Random Forest algorithm to predict diseases [5].

The authors conducted thorough analysis and normalization of models to understand the need for these predictions. The Voting Classifier of Decision Tree, Sigmoid SVC, and Adaboost achieved the highest accuracy of 88.57% for heart disease, while the voting classifier had an accuracy of 80.95% for diabetes. The study also suggests the potential extension of this methodology to predict a patient's immunity to COVID-19. Future work could involve developing a robust model with automated feature selection to analyze both diseases and their relationship to COVID-19 [6].

The research provided useful disease prediction based on symptoms. A script was developed to extract and format data as needed. Users can input symbols or select options for prediction. Users can also create medical profiles to contribute to the database and improve predictions over time. The analysis revealed similarities between diseases, but the training model would benefit from a larger database [7].

The proposed research had a positive impact on hospitals, catering to dynamic environments and benefiting both patients and organizations. It offers database filtering for monitoring disease outbreaks and ensures data security with individual hospital databases. Unique QR codes help identify patients and reduce wait times. The web app allows patients to access lab reports and medicine notes, eliminating the need for physical copies. Overall, the project improves patient experience and facilitates future consultations [8].

This research implemented a Hospital Management System using Django. The evaluation and assessment by respondents and end-users showed that the system effectively meets their needs and requirements. It was rated positively in terms of acceptability, effectiveness, quality, and productivity in automating hospital management. The conclusion is that the system is efficient, eliminates manual errors, improves hospital operations, enhances patient satisfaction, and overall improves the functioning of the hospital [9].

The research aims to provide a significant solution for icterus sufferers. The proposed rules engine framework is designed to identify parameters and administer appropriate medicines. Future plans involve implementing machine learning to enhance the rules engine's success [10].

### 3 Proposed System

Our project is a web-based application that is based on a medical and hospital ecosystem. This ecosystem is a comprehensive platform that is designed to provide patients with easy access to a variety of health care services and resources. This platform is a new strategy that has the potential to improve the healthcare business by providing patients with a variety of features that make the process more convenient and effective overall.

The capability of the platform to allow users to do self-evaluations and view health data online is one of the most important characteristics it possesses. Patients will be better able to monitor their own health and recognise any possible issues before they become more serious with the help of this tool. Patients are able to use this function by using the user-friendly interface that the platform provides. This interface makes it simple for patients to navigate and use the platform.

Comparatively with the existing system, we are able to achieve 89% accuracy in self-assessment system and in the other part, the diagnose section where the patient's health report is tested, there in heart disease testing we achieve 90.2% accuracy. Our project aims to build the whole health care environment to provide our users a complete health guide to know the disease - from symptoms to test their report data from blood sample, to know the potential risk of disease.

## 4 Analyzing Requirement

To build the project, the following requirements needed to be fulfilled:

- *Front-end Development:* The website's user interface was developed using HTML, CSS, and JavaScript.
- *Back-end Development:* Python Django framework was used for the back-end development, which handles user requests, retrieves and stores data, and interfaces with the machine learning algorithm.
- *Machine Learning Algorithm:* An appropriate machine learning algorithm was selected for disease prediction.
- *Self-assessment and Health Report Generation:* The website has a self-assessment module that allows users to input relevant information about their health, lifestyle, and medical history.

The tech stack used to build the project included HTML, CSS, and JavaScript for front-end development, Python Django framework for back-end development, and a machine learning algorithm for disease prediction.

### 4.1 Front-End Development

The project's front-end development included the design and creation of a user-friendly website interface utilising HTML, CSS, and JavaScript. HTML was utilised to structure the website's content and define text, graphics, and other page elements. Using font, colour schemes, and layout, CSS was utilised to style and make the website visually appealing.

### 4.2 Back-End Development

The project's backend was developed using the Python Django web framework. Django is a robust and flexible framework that offers a comprehensive solution for web development. It is renowned for its usability, scalability, and safety.

The website's backend was responsible for processing user requests, retrieving and storing data, and communicating with the machine learning algorithm. The backend was developed utilising the Model-View-Controller (MVC) architectural paradigm of Django. This pattern divides the programme into Model, View, and Controller components.

### 4.3 Machine Learning Algorithm

We are developing symptom checker tools as part of our ongoing work on machine learning, which will be used in the healthcare project we are working on. This will allow patients to perform their own self-assessment check-ups. These tools are intended to assist patients in evaluating their symptoms and providing recommendations regarding the next steps that should be taken.

In this part of the project, machine learning is also being used to develop predictive models in order to assist our healthcare project in predicting the health risks that patients may face. These models are able to identify patients who are at a high risk of developing certain conditions and provide personalised recommendations on how to manage the patients' health by analysing large sets of data pertaining to individual patients.

**Decision tree** is a method for machine learning that builds a tree-like representation of decisions and their possible outcomes. It is useful for classification and regression issues, and is ideally suited to data with both categorical and continuous characteristics. Decision trees are simple to read, and the resulting model may be represented graphically and explained to non-specialists.

**Logistic regression** is an approach for supervised learning used to solve classification problems. It models the link between a dependent variable and one or more independent factors and estimates the likelihood that the dependent variable belongs to a specific class. It is frequently used in medical research to estimate a patient's likelihood of developing a particular condition.

**Random forest classifier** is a technique for ensemble learning that mixes numerous decision trees to enhance accuracy and avoid overfitting. At each split, it randomly selects a subset of the features and trains each tree using a separate random sample of the data. It is frequently employed for high-dimensional datasets with numerous features.

**Naïve Bayes** algorithm is a probabilistic classification system. Bayes' theorem states that the probability of a hypothesis is updated based on fresh data. The Naïve Bayes assumption is that the features are conditionally independent of the class label, hence the term "Naïve." It excels at text categorization tasks, including spam detection and sentiment analysis.

### 4.4 Data Management

Our training dataset contains a total of 4920 rows and 133 columns. Out of these columns, 132 represent different symptoms that a person can experience, and the final column represents the prognosis or target disease.

We have a total of 41 diseases in our dataset, and each disease is represented by a set of symptoms that have been mapped to it. Each disease is present in 120 rows of our

dataset. This means that for each disease, we have 120 instances of symptom data that can be used to train our machine learning model.

During the training phase, we used the symptom data from these 41 diseases dataset to train our machine learning model. We divided the patient dataset into two files: training.csv and testing.csv. one file was used to train the model, while the other file was used to validate the model's performance during the training process (Table 1).

**Table 1.** Training Dataset Statistics

Type	Count
rows	4920
columns	133
symptoms	132
disease	41
rows-disease	120

## 5 Methodology

We've been working on extending the project to create a web application that incorporates a machine learning model [11, 12]. The web application provides an interface that allows users to input their symptoms and receive a prediction (ref. Fig. 3) of the most likely disease they might be suffering from, along with the confidence score of having that disease [13–15].

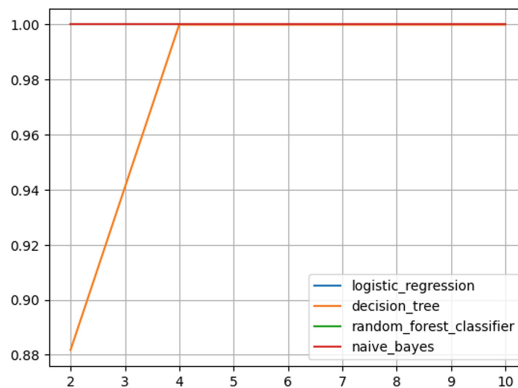
Our project aimed to predict diseases using medical data through a machine learning system employing multiple supervised learning approaches, such as logistic regression, random forest classifier, Naïve Bayes, and decision trees. During the testing phase, we evaluated the performance of these algorithms using k-fold cross-validation, a typical strategy for model selection that involves dividing the data into k subsets, training the model on k-1 subsets, and assessing the model's performance on the final subset. We performed this procedure with k values of 2, 4, 6, 8, and 10 and calculated the ultimate score by averaging the results, represented in Table 2 and Fig. 1.

Based on our evaluation (ref. Table 2), we found that the decision tree algorithm at k-value: 2 was the most effective method for predicting diseases using medical data. This is because decision trees are simple to read and can be displayed and explained to non-experts, making them a valuable tool for medical practitioners and patients. Additionally, decision trees can accommodate both categorical and continuous variables, which are prevalent in medical datasets. We found that the decision tree algorithm accurately classified 89% of the test cases, a high rate of accuracy. In contrast, we observed that some methods, such as logistic regression, random forest classifier, and Naïve Bayes, appeared overfit, with training scores above testing scores. Overfitting happens when a model is overly complicated and has learned the noise in the training data, resulting

in inadequate generalization to new data. Therefore, the decision tree approach was the most appropriate for predicting diseases using medical data. So, we choose to build our web application backend using the decision tree model to process and predict the health.

**Table 2.** Self-Assessment, Testing Score using K-fold cross-validation a different k value.

Algorithm	k = 2	k = 4	k = 6	k = 8	k = 10
logistic regression	100%	100%	100%	100%	100%
decision tree	89%	100%	100%	100%	100%
random forest	100%	100%	100%	100%	100%
naïve bayes	100%	100%	100%	100%	100%



**Fig. 1.** Self-Assessment, testing score plot

The web application has another service (ref. Fig. 4) for health report generation where the user inputs data from a blood report to determine whether they are at risk or safe [16].

We have performed model preparation on heart disease. To achieve this, four machine learning models were trained on the available data: random forest classifier, K-nearest neighbors (KNN), logistic regression, and naive Bayes. The performance of each model was evaluated using their training and testing scores, and the results are as follows (Table 3):

In general, the higher the accuracy score, the better the model performs. From the results above, we can see that the KNN and logistic regression models performed the best, achieving the highest testing scores of 90.2%. However, between these two models, logistic regression had a slightly better training score, suggesting that it might be more reliable when it comes to predicting new cases [17–21].

Overall, the logistic regression model is likely the best choice for this web application as it demonstrated the highest testing score and a relatively high training score.

**Table 3.** Report Testing, Heart Disease Accuracy table

Algorithm	Training Score	Testing Score
Random Forest	81%	86.9%
KNN	83.05%	90.2%
Logistic Regression	85.12%	90.2%
Naïve Bayes	84.29%	86.9%

The web application service has a total of six diagnoses (Fig. 2) in addition to heart disease (Fig. 4). These include liver disorder, tuberculosis, kidney disorder, blood sugar test, and diabetes test.

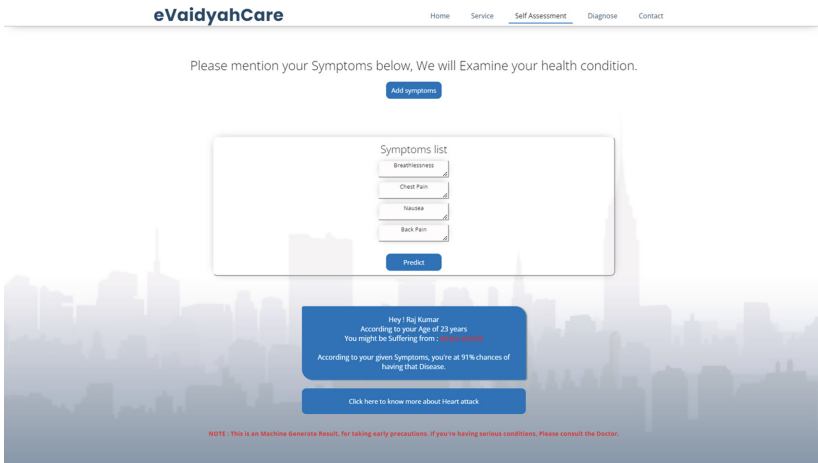
**Fig. 2.** Report Testing, Diagnose Section

The web application is designed to be user-friendly and accessible to anyone. Users can input their symptoms in a simple and straightforward way, using dropdown menus to select. The web application sends the symptom data to the backend, where our machine learning model processes it and predicts the most likely disease based on the symptom data. The prediction is then displayed to the user along with the confidence score [22–24].

One of the benefits of our platform is that it allows individuals to self-assess (Fig. 3) their health status without requiring a doctor's visit. This can be particularly useful for people who live in remote areas, or who may not have access to medical professionals due to cost or other reasons. Our platform can also help people become more aware of their health status, and encourage them to seek medical attention if necessary.

Suppose you are feeling “breathlessness, chest pain, nausea and back pain” in this case you have to visit our ‘self-assessment section’ (Fig. 3) to know about the disease that may risks in you.





**Fig. 3.** Self-Assessment Section

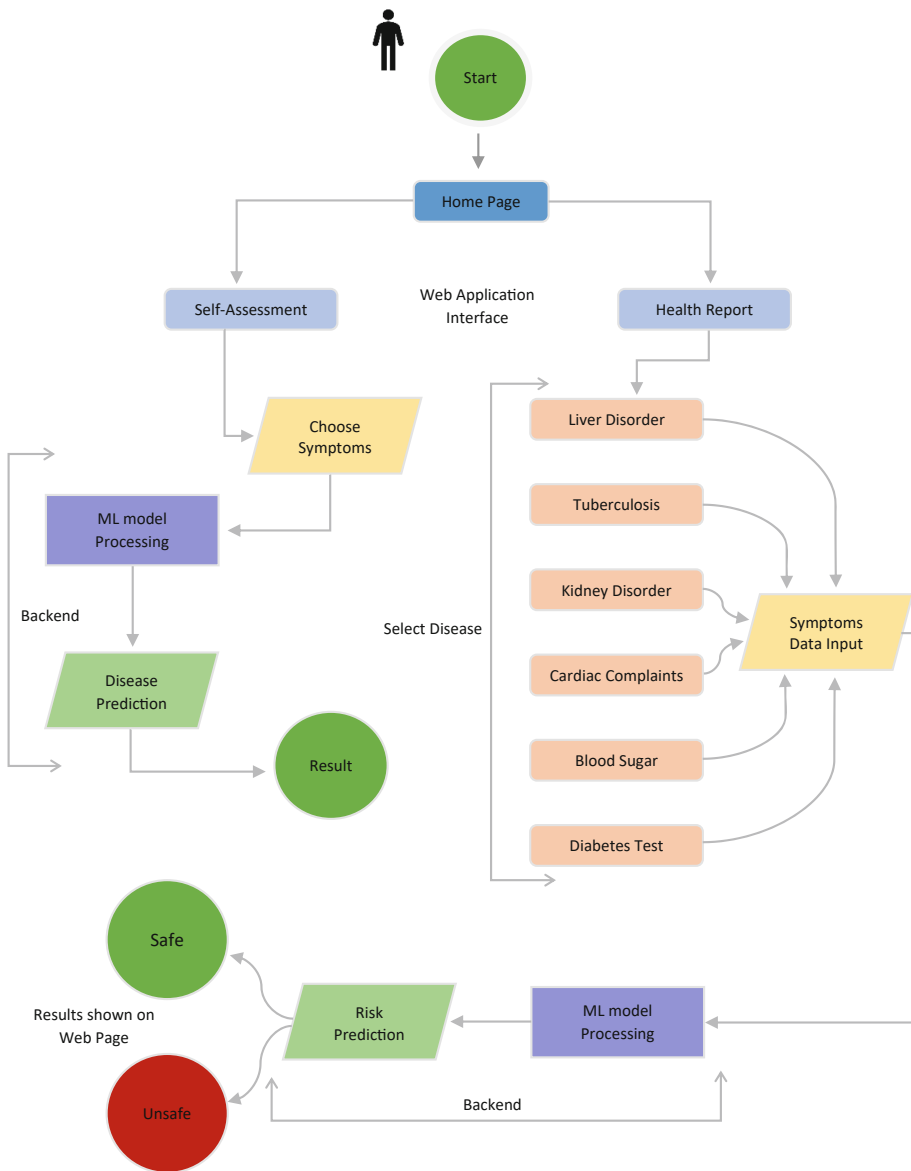
Above in Fig. 3 Our system predicts the potential health disease from the give symptoms. You may risk of 'Heart Attack'. Now you have to visit testing laboratory to fill the following data in 'diagnosis section' (Fig. 4) of heart diagnose, which will examine the report from machine learning trained model to give the health report whether you're in risk or safe of heart attack.



**Fig. 4.** Heart Diagnosis Section

In summary, our web application provides a user-friendly and accessible platform for health assessment and disease detection, refer Fig. 5 for Data Flow Diagram of our Web Application where we have shown the services that our application offers in convenient way.

## 6 System Architecture



**Fig. 5.** Data Flow Diagram

## 7 Future Work

In the future, we plan to develop a new feature for our web application by using recommender system model to calculate the risk percentage of a person's daily or regular food consumption.

This model will be very beneficial for people who want to take control of their health by understanding whether their diet is healthy or unhealthy for them.

To use this model, individuals will need to input their daily diet items and provide information about any existing health conditions they have. Based on this data, the model will provide personalized recommendations on what foods to avoid to help reduce the risk of exacerbating any existing health conditions.

Overall, this model has the potential to significantly improve people's health outcomes by providing tailored and accurate recommendations for their diet.

## 8 Conclusion

In conclusion, the work done in this research aims to provide a solution for self-assessment and disease prediction using machine learning algorithms. The project has successfully designed and built a web application that uses machine learning models in the backend to process the data for prediction of disease based on trained models.

The dataset used in this project has 133 columns, with 132 of these columns being symptoms that a person experiences, and the last column being the prognosis. These symptoms are mapped to 41 diseases that the project aims to classify. The project has trained its models on a training dataset and tested it on a testing dataset.

Decision tree classifier model was found to give the best accuracy among other models for disease prediction. The project has designed and built a user-friendly web application that allows users to perform self-assessment and generate health reports about their risk or safety from specific diseases.

Overall, this project has successfully achieved its goal of building a machine learning-based solution for disease prediction and self-assessment. It provides an easy-to-use tool that can assist patients in identifying their disease risk, which ultimately leads to early detection and treatment, thereby improving the overall healthcare system.

**Acknowledgment.** We would like to express our gratitude to all those who have contributed to the successful completion of this research project. First and foremost, we would like to thank our supervisor for providing us with invaluable guidance and support throughout the project.

Once again, we extend our heartfelt thanks to all those who have played a role in making this project a success.

## References

1. Kalshetty, J.N., Achyutha Prasad, N., Mirani, D., Kumar, H., Dhingra, H.: Heart health prediction using web application. *Int. J. Health Sci.* (2022)
2. Srivastava, K., Choubey, D.K.: Heart disease prediction using machine learning and data mining. *Int. J. Recent Technol. Eng.* (2020)

3. Nawab, M.A., Malpani, T., Kaundal, T., Soni, M.D.: Disease prediction web app using machine learning. *Int. Res. J. Mod. Eng. Technol. Sci.* (2022)
4. Jagtap, A., Malewadkar, P., Baswat, O., Rambade, H.: Heart disease prediction using machine learning. *Int. J. Res. Eng. Sci. Manag.* (2019)
5. Tambe, Y., Awhad, S., Keny, A.: Disease prediction using machine learning and Django and online consultation (2022)
6. Dhande, B., Bamble, K., Chavan, S., Maktum, T.: Diabetes & heart disease prediction using machine learning. *ITM Web Conf.* **44**, 03057 (2022)
7. Srivastava, M., Yadav, V., Singh, S.: Implementation of web application for disease prediction using AI. *BOHR Int. J. Data Min. Big Data* (2020)
8. Ramees, A.R., Akhil, P.S., Amrutha, M.A., Fathima, J., Elia, N.: Hospital managing QR code web application using Django and Python. *Int. J. Creat. Res. Thoughts (IJCRT)* (2020)
9. Gawande, M.V., Pisey, P., Shinde, A.A., Ghagre, P., Bhusari, K.: Hospital management system in Django. *Int. Res. J. Mod. Eng. Technol. Sci.* (2022)
10. Roghit, K.K., Jayachandran, G.: Assessment for hyperbilirubinemia health care using Django. *Int. Res. J. Eng. Technol. (IRJET)* (2020)
11. Mukherjee, D., Raj, I., Mishra, S.: Song recommendation using mood detection with Xception model. In: Mallick, P.K., Bhoi, A.K., Barsocchi, P., de Albuquerque, V.H.C. (eds.) *Cognitive Informatics and Soft Computing*. LNNS, vol. 375, pp. 491–501. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-16-8763-1\\_40](https://doi.org/10.1007/978-981-16-8763-1_40)
12. Abhishek, Tripathy, H.K., Mishra, S.: A succinct analytical study of the usability of encryption methods in healthcare data security. In: Tripathy, B.K., Lingras, P., Kar, A.K., Chowdhary, C.L. (eds.) *Next Generation Healthcare Informatics*. SCI, vol. 1039, pp. 105–120. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-19-2416-3\\_7](https://doi.org/10.1007/978-981-19-2416-3_7)
13. Sinha, K., Miranda, A.O., Mishra, S.: Real-time sign language translator. In: Mallick, P.K., Bhoi, A.K., Barsocchi, P., de Albuquerque, V.H.C. (eds.) *Cognitive Informatics and Soft Computing*. LNNS, vol. 375, pp. 477–489. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-16-8763-1\\_39](https://doi.org/10.1007/978-981-16-8763-1_39)
14. Mishra, Y., Mishra, S., Mallick, P.K.: A regression approach towards climate forecasting analysis in India. In: Mallick, P.K., Bhoi, A.K., Barsocchi, P., de Albuquerque, V.H.C. (eds.) *Cognitive Informatics and Soft Computing*. LNNS, vol. 375, pp. 457–465. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-16-8763-1\\_37](https://doi.org/10.1007/978-981-16-8763-1_37)
15. Patnaik, M., Mishra, S.: Indoor positioning system assisted big data analytics in smart healthcare. In: Mishra, S., González-Briones, A., Bhoi, A.K., Mallick, P.K., Corchado, J.M. (eds.) *Connected e-Health*. SCI, vol. 1021, pp. 393–415. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-97929-4\\_18](https://doi.org/10.1007/978-3-030-97929-4_18)
16. Periwal, S., Swain, T., Mishra, S.: Integrated machine learning models for enhanced security of healthcare data. In: Mishra, S., Tripathy, H.K., Mallick, P., Shaalan, K. (eds.) *Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis*. SCI, vol. 1024, pp. 355–369. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-19-1076-0\\_18](https://doi.org/10.1007/978-981-19-1076-0_18)
17. De, A., Mishra, S.: Augmented intelligence in mental health care: sentiment analysis and emotion detection with health care perspective. In: Mishra, S., Tripathy, H.K., Mallick, P., Shaalan, K. (eds.) *Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis*. SCI, vol. 1024, pp. 205–235. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-19-1076-0\\_12](https://doi.org/10.1007/978-981-19-1076-0_12)
18. Dutta, P., Mishra, S.: A comprehensive review analysis of Alzheimer’s disorder using machine learning approach. In: Mishra, S., Tripathy, H.K., Mallick, P., Shaalan, K. (eds.) *Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis*. SCI, vol. 1024, pp. 63–76. Springer, Singapore (2022). [https://doi.org/10.1007/978-981-19-1076-0\\_4](https://doi.org/10.1007/978-981-19-1076-0_4)

19. Banerjee, D., Kukreja, V., Hariharan, S., Sharma, V.: Fast and accurate multi-classification of kiwi fruit disease in leaves using deep learning approach. In: 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, pp. 131–137 (2023). <https://doi.org/10.1109/ICIDCA56705.2023.10099755>
20. Raj, A., Sharma, V., Shanu, A.K.: Comparative analysis of security and privacy technique for federated learning in IoT based devices. In: 2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM), pp. 1–5. IEEE, November 2022
21. Manikandan, N., Ruby, D., Murali, S., Sharma, V.: Performance analysis of DGA-driven botnets using artificial neural networks. In: 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1–6. IEEE, October 2022
22. Ali, M.A., Balamurugan, B., Sharma, V.: IoT and blockchain based intelligence security system for human detection using an improved ACO and heap algorithm. In: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 1792–1795. IEEE, April 2022
23. Juyal, V., Sagggar, R., Pandey, N.: An optimized trusted-cluster-based routing in disruption-tolerant network using experiential learning model. *Int. J. Commun. Syst.* **33**(1), e4196 (2020)
24. Juyal, V., Pandey, N., Sagggar, R.: A heuristic light weight security algorithm for resource constrained DTN routing. In: 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), pp. 1–4. IEEE, December 2016