

# Easy Detect—Web Application for Symptom Identification and Doctor Recommendation



**Abhinay Krishna Chitirala, Chunduri Sai Akhilesh, Garimella Venkasatyasai Phani Ramaaditya, Kollipara Bala Vineel, Krosuri Sri Chandan, and M. Rajesh**

**Abstract** It is not very safe to go to hospitals for regular check-ups as we are used to. Patients can get medical advice from the comfort of their homes from specialized medical professionals. Patient data can also be digitalized, so that it can be used at any hospital. Symptoms are taken as input, and our deep learning model will give the possible disease, which the person might be suffering from, and the doctor he/she needs to consult for further medication.

**Keywords** Deep learning · Medical · Patients · Doctors · Symptoms · Consultation · Medication

---

A. Krishna Chitirala · C. Sai Akhilesh · G. Venkasatyasai Phani Ramaaditya · K. Bala Vineel · K. Sri Chandan · M. Rajesh (✉)  
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India  
e-mail: [m\\_rajesh@blr.amrita.edu](mailto:m_rajesh@blr.amrita.edu)

A. Krishna Chitirala  
e-mail: [blenu4cse17001@bl.students.amrita.edu](mailto:blenu4cse17001@bl.students.amrita.edu)

C. Sai Akhilesh  
e-mail: [blenu4cse17019@bl.students.amrita.edu](mailto:blenu4cse17019@bl.students.amrita.edu)

G. Venkasatyasai Phani Ramaaditya  
e-mail: [blenu4cse17031@bl.students.amrita.edu](mailto:blenu4cse17031@bl.students.amrita.edu)

K. Bala Vineel  
e-mail: [blenu4cse17062@bl.students.amrita.edu](mailto:blenu4cse17062@bl.students.amrita.edu)

K. Sri Chandan  
e-mail: [blenu4cse17065@bl.students.amrita.edu](mailto:blenu4cse17065@bl.students.amrita.edu)

# 1 Introduction

## 1.1 *Motivation and Contributions of the Research*

COVID-19 has had an impact on all of us. The pandemic's implications and consequences, on the other hand, are felt differently in different sectors. This pandemic has had the most significant impact on the health care industry. As the number of cases of COVID increased, hospitals converted various wards into COVID units. This increase in COVID wards, and the need to avoid overcrowding make it more difficult for people with other diseases to see doctors and get to laboratories for consultation and testing. As a result, we need an app that provides an essential diagnosis based on the patient's symptoms and recommendations for which doctor to visit. The patient can even book an appointment in advance at a particular time so that he/she can avoid waiting in the hospitals in case of emergencies.

This study focuses on the design features of the prediction system for medical conditions to detect many typical diseases. Techniques like neural networks, decision making and logistic regression are used for the topic implementation. We have acquired the required data set. The algorithm also proposes physicians that are applicable for the pattern detected disease(s).

## 1.2 *Introduction to the Easy Detect Application*

We have developed a Web application that can detect diseases in patients based on their symptoms. They are connected with a specialized doctor for further consultation based on the application's results. We can schedule appointments ahead of time to avoid waiting in hospitals. Determine which specialized doctor to consult ahead of time. We cannot risk the patient's health during his regular check-ups in the given circumstances, where social distancing is critical. As a result, this application will keep track of its clients' health regularly and with the help of doctors.

So, we would like to propose an intelligent system trained based on past medical records (symptoms for specific diseases).

The proposed system is designed to support the decisions of doctors and is not designed for a patient without supervision by a medical practitioner for individual usage. The remaining part of this article has the following structure: The literature review conducted in the medical field is described in Sect. 2.

## 2 Literature Review

Intelligently analysed data becomes a corporate requirement to find effective and trustworthy detections of disease as quickly as possible to ensure the best possible treatment for patients. This detection has been conducted in recent decades by finding remarkable patterns in databases. The technique of retrieving information from the database is known as data mining. Finding these patterns, however, is a challenging process. This has led to the development of various artificial intelligence approaches, including machine learning as a tool for providing intelligent data processing. Medical data sets are usually multidimensional, on the other hand. The use of big data technology is necessary, in certain situations, when machine learning techniques fail. Deep learning has, therefore, developed into a subset of machine learning, which allows us to work with such data sets.

Caballé et al. [1] gave a comprehensive overview of smart data analysis tools from the medical area. They also include examples of algorithms used in different medical fields as well as an overview of probable trends depending on the objective, process employed and the application field. The benefits and cons of each approach were also overcome.

In all shown fields of application, the author states that the categorization is the most usual action in the medical profession. In the realm of infectious diseases, regression, on the other hand, is a regular task. In illnesses like Alzheimer's or Parkinson's diseases, this duty is rare to be employed. In addition, the task of clustering in liver and cardiovascular diseases is briefly studied, but is used extensively in Alzheimer and Parkinson diseases. In the case of cancer, Alzheimer's, Parkinson's and renal disease studies, neural networks and other supervised algorithms are commonly employed in study into metabolism, hepatic, infectious and heart illness.

The author chose the technique based on the advantages and disadvantages of each tool in the specific application area and under his or her experimental conditions.

Traditional approaches can be used with large volumes of data and powerful hardware architectures to represent more complex statistical phenomena, while ML enables previously hidden patterns to be identified and trends extrapolated and the result to be predicted in the absence of trace problems as well. Currently, machine learning algorithms are employed in clinical practice in medical records, for instance, to forecast which patients would most likely be hospitalized or who are less susceptible to a prescription of therapies. Diagnostic, research, drug development and clinical trials have unlimited possibilities. Although there are large numbers of digital data, predictive medical record models are typically based on basic linear models and seldom take into account more than 20 or 30 parameters.

Dhomse Kanchan et al. [2] used SVM, Naive Bayes and decision tree with and without PCA on the data set to predict heart disease. The principal component analysis (PCA) approach is used to reduce the number of characteristics in a data set. When the data set size is decreased, SVM beats, Naive Bayes and decision tree. SVM

may potentially be used to forecast the start of cardiovascular illness. Their algorithms were developed using the WEKA data mining approach, which was utilized to evaluate algorithm accuracy after executing them in the output window.

These techniques evaluate classifier accuracy relying on properly identified examples, the time required to create a model, mean absolute error and ROC area. As a consequence, they concluded that, when compared to other methods, the maximum ROC area indicates outstanding prediction performance.

The methods are rated based on how long it takes to create a model, how many cases are properly categorized, the error rate and the ROC area. The algorithm's accuracy is displayed in Naive Bayes 34.8958 per cent correctly instances accuracy with a minimum Naive Bayes mean absolute error = 0.2841 and a maximum Naive Bayes ROC = 0.819 times needed to construct the model = 0.02 s. Based on the explorer interface data mining approach, we can infer that Naive Bayes has the greatest accuracy, the lowest error, the shortest time to develop and the maximum ROC.

Human illness diagnosis is a tough process that requires a high level of skill. Any attempt to develop a Web-based expert system for human illness diagnosis must overcome a number of obstacles.

This project's [3] objective is to develop a Web-based fuzzy expert system for detecting human illnesses. Fuzzy systems, which portray systems utilizing linguistic principles are currently being employed successfully in a growing variety of application domains. Hasan et al. [3] are investigating and developing a Web-based clinical tool to increase the quality of health information sharing between physicians and patients. This Web-based tool can also be used by practitioners to confirm diagnoses. To assess its performance, the proposed system is tested in a variety of scenarios. The proposed system achieves satisfactory results in all cases.

A control programme is created by gathering, encoding and storing knowledge. To diagnose the fuzzy expert system, a uniform structure was developed, and mathematical equivalence will be employed. The likelihood of illnesses was calculated using that equation, the value of which was determined via feedback during diagnosis. In this case, a catalytic factor is employed in the form of a question about prior results, which is also taken into consideration during the probability calculation.

The addition of catalyst after evaluation increases the accuracy of the system as past results play a significant role in illness prediction. The following system increases the accuracy and it works with real-time diagnosis. It was even found that the confidence level of this system after observing past pathological tests was far better than otherwise.

Laxmi et al. [4] The usage of Bayesian networks is presented in the creation of a system of clinical decision support. Infer network parameters, which offer the idea of learning were used to the Bayes ML technique. The study is unique in that, in addition to identifying diseases, it attempts to propose laboratory testing, infers diseases from laboratory test data and offer age-based therapeutic prescriptions for regularly occurring diseases in India. For simulating laboratory testing and medical prescriptions, a rule-based technique is employed.

Mohanty et al. [5] deal with the problem of the symptoms and seriousness of the most likely sickness in the physician. ANFIS benefits from the classic fuzzy models by being extremely flexible and easily learned. The patient and the diagnostic information will be the learning and testing of the system when the system is deployed to a clinic.

Based on the above citations, we have used a filter to reduce the number of features based on their importance in finding the result [6, 7]. The importance of each feature is determined using a coefficient matrix. We inferred that SVM can be more effective while dealing with cardiovascular illness but with an overall data set of large size, we concluded that Naive Bayes is better.

### 3 Design and Implementation

In this section, we are going to discuss the design and implementation of the modules used in the application. We have used different modules such as the data collection module, logistic regression module, decision tree module, neural network module and a disease prediction module.

#### 3.1 Design

This diagram depicts the operation of our application. The data will be separated into training, testing and model training once the deep learning model is pre-processed. The model will then be loaded into the Web application to forecast the ailment that the patient is suffering from (Fig. 1).

#### 3.2 Implementation

- a. **Data Collection Module:** The data collecting module is used to build a knowledge base for a medical illness prediction system. The collection of disease-related symptoms is the first step in the data collection process. 41 disorders and 132 symptoms were picked for the initial deployment. The symptoms considered were a wide range of common symptoms that a patient might experience. Later this data set is processed for feature extraction using the coefficient matrix as shown in the below figure. Among the coefficients, a 0.4 quantile of symptoms is removed which brings down the data set to make the data set more feasible for the model to be used. We use a pre-processing input function to pre-process the data, i.e. labelling the data and splitting it into a 70:30 ratios [8, 9].
- b. **Logistic Regression Module:** Training and testing are the two phases of the logistic regression module. The first phase is designing the model and training it

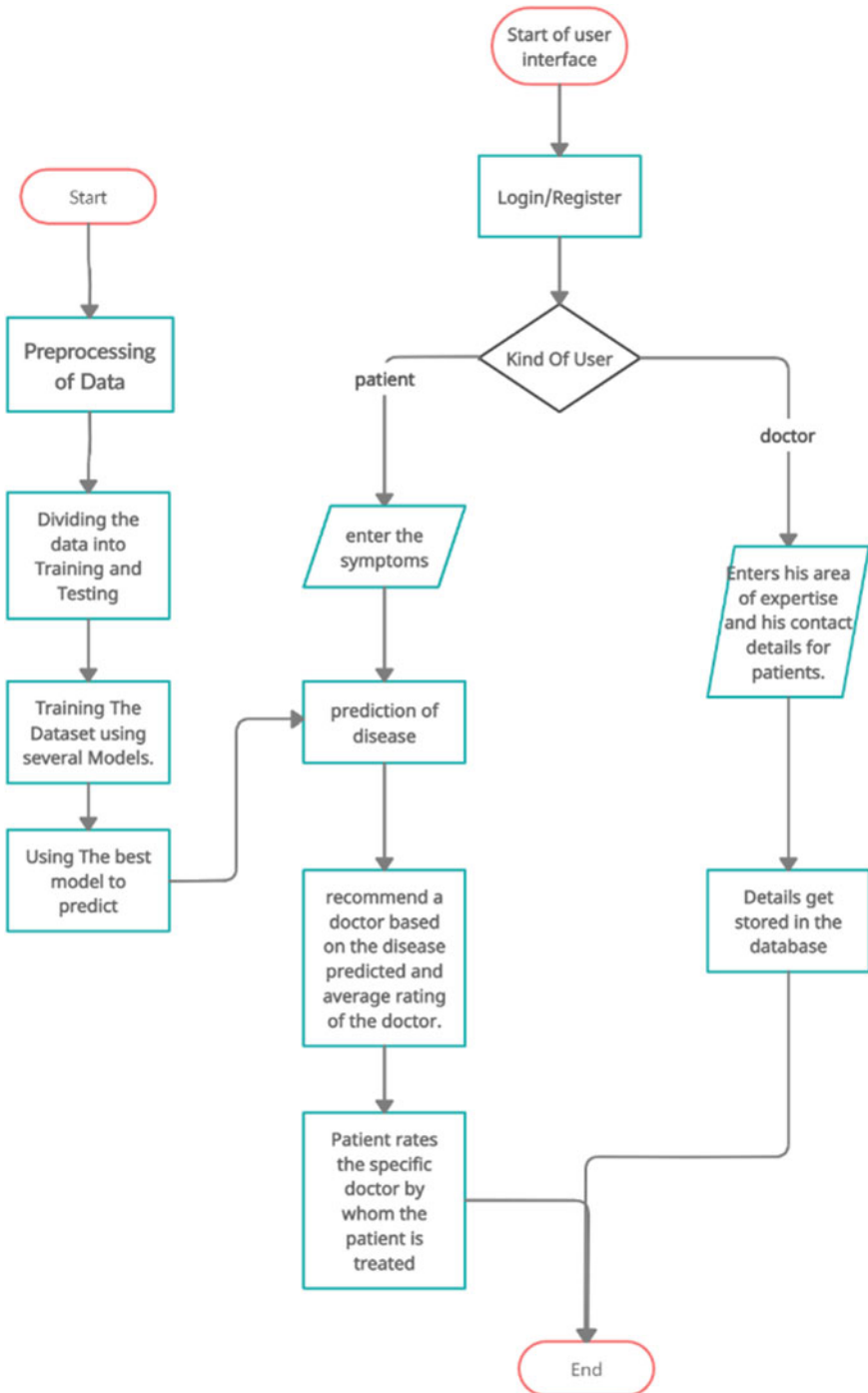


Fig. 1 Flow diagram of the application

with data gathered from the data collecting module; whereas, the second phase involves testing the model and finding accuracy [10].

- Logistic Regression Model Creation: The data set created in the above module is used to create logistic regression. The multinomial class detection and solver as limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) [11–13].
  - Logistic Regression Testing: Testing data is inputted into the trained model, which involves creation of probability for diseases using Gaussian algorithm [14, 15].
- c. **Decision Tree Module:** It is constructed with a data set from the data collection module. The module is separated into two phases: training and testing.
- Decision Tree Model Creation: The model is built using the data set from the data collection model. It uses the information gain algorithm to build the decision tree in which internal nodes represent the symptoms and leaf nodes represent diseases.
  - Decision Tree Testing: Testing data is inputted into the trained model, which involves traversing the tree through the symptoms to find the disease.
- d. **Neural Network Module:** The neural network model is a sequential model, which is built using different layers containing a different number of nodes or neurons. The model consists of three dense layers and three activation layers in the following order:
1. Dense layer (32-nodes)
  2. Activation layer (ReLU)
  3. Dense layer (16-nodes)
  4. Activation layer (ReLU)
  5. Dense layer (41-nodes): Output layer
  6. Activation layer (Softmax).

The training and testing phases of this model are divided into two parts:

- Neural Network Model Creation: The model is built using the data set from the data collection module. Each neuron has a weight associated with it. Activation functions are applied to a whole layer of neurons. These provide nonlinearity, without which the neural network reduces to a mere logistic regression model. After every epoch, the parameters and hyper-parameters of the model are modified such that the cost function is reduced till it reaches the point of global minima. The ReLU is the activation function employed here (rectified linear unit). The output layer, also known as the last layer, is made up of 41 neurons whose outputs are passed through the final activation layer containing Softmax activation function, which returns the probability of occurrence of the corresponding diseases. This model is compiled with categorical cross-entropy as loss (as we are performing a classification), validation accuracy as a metric and Adam as optimizer. An early stopping mechanism is also added with the patience of two epochs to prevent the

overfitting of the model on training data, i.e. if the validation accuracy is either decreasing or is constant, the training would end there ...

- **Neural Network Model Testing:** Testing data is passed to the model along with another data called validation data, which validates or verifies the performance of the model on testing data.
- **Disease Prediction Module:** The disease prediction module is designed on the trained model. The symptoms data is gathered from the UI provided for the user. The symptoms thus gathered are made into a NumPy list where the symptoms, which are marked by the user are given the value “1” and others have default value of “0” and this list is passed using the trained model to forecast the likelihood of each disease’s occurrence

### 4 Results and Analysis

A sample testing set of roughly 42 records was used to evaluate the decision tree approach for the current paper’s implementation.

- Accuracy for the Decision Tree Model is: 97.62
- Accuracy for the Logistic Regression is: 94.93
- Accuracy for the Neural Network is: 94.3.

Figure 2 illustrates different values of (accuracy and validation accuracy) versus (epochs), i.e. graph on left, the graph on the right illustrates the distribution of (loss and validation loss) versus (epochs). After the twentieth epoch, accuracy and validation accuracy remain almost constant.

Figure 3 shows the list of symptoms from which the patient can select particular symptoms, which he is suffering from and submit to generate a report.

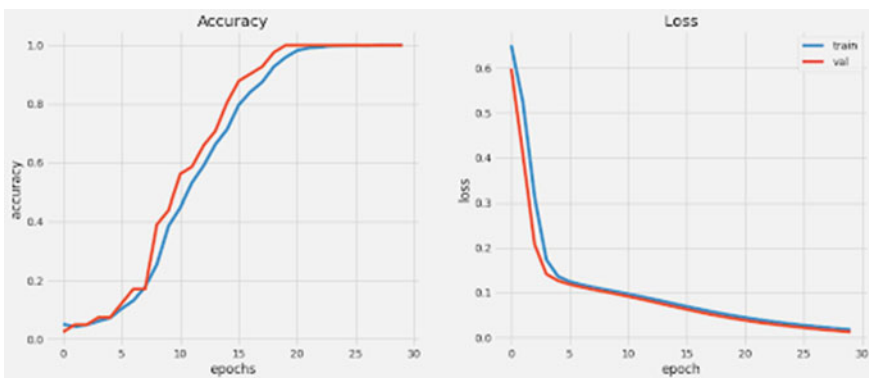


Fig. 2 Values of accuracy and loss for validation and training versus epochs



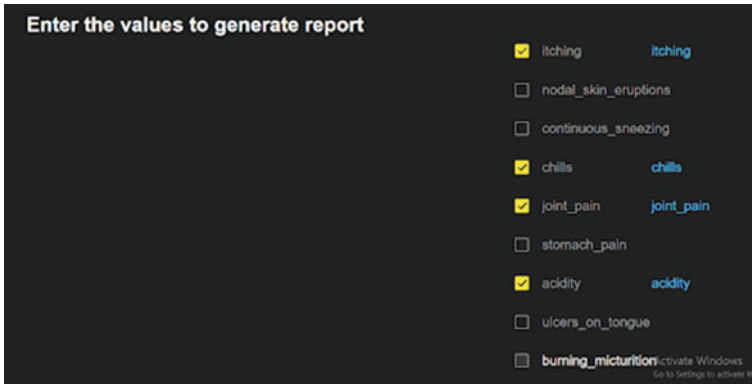


Fig. 3 User interface for patients to enter data

Figure 4 depicts the output generated by the model. This output consists of different values, which range from 0 to 1 multiplied by 100, that represent the probability of occurrence of the list of diseases, and the result of the record illustrated by figure informs us that there is a high chance the patient or the details related to a person is suffering from “urinary tract infection”.

Figure 5 shows the details of the doctors recommended for the respective diseases.

In this page, the user can decide on a doctor and move ahead as they will be redirected to booking appointments.

## 5 Conclusion

A linear regression model to predict a most likely disease from a particular set of symptoms is developed. As a result, the number of symptoms reduced from 133 to 79 symptoms using a coefficient matrix and took 0.4 quantile out of it, and trained the model which gave the accuracy of 95.93%. A decision tree model to predict disease using all the symptoms which gave us an accuracy of 97.6% is also developed. A neural networks model is developed with two hidden layers and an output layer with 21 epochs, which gives the accuracy of 95.3%. From these results, it can conclude that decision tree is the best model for the given data set. It is able to provide a user interface for the disease prediction, mapped the respective diseases with a specialization, so that a doctor with the required specialization can be recommended to the patient. Provided an option of rating the doctor after the respective appointment based on which the doctors are recommended later on.

There is a possibility of advancement in the machine learning part, where we can improve or add new models such as neural networks with different activation functions. We can also try out different models such as SVM, and also, we can include feature reduction methods like PCA. Regarding the Web application, we can

```
'Allergy': 0.29797658841897195,  
'Arthritis': 0.10202953011841367,  
'Bronchial Asthma': 0.08577025658186797,  
'Cervical spondylosis': 0.374924267671638,  
'Chicken pox': 0.031507094918759394,  
'Chronic cholestasis': 0.0594407229662809,  
'Common Cold': 0.021243515902439482,  
'Dengue': 0.015267691518277605,  
'Diabetes ': 0.08082729040338064,  
'Dimorphic hemmorhoids(piles)': 0.08282832122938537,  
'Drug Reaction': 2.348780216413582,  
'Fungal infection': 0.12382000725163839,  
'GERD': 0.07026928262853761,  
'Gastroenteritis': 0.15664442618135044,  
'Heart attack': 0.15147591201697141,  
'Hepatitis B': 0.02217275379773512,  
'Hepatitis C': 0.03593907272777648,  
'Hepatitis D': 0.02494600689683914,  
'Hepatitis E': 0.02287860782924121,  
'Hypertension ': 0.17533155907786818,  
'Hyperthyroidism': 0.0539436252361747,  
'Hypoglycemia': 0.025092467487966483,  
'Hypothyroidism': 0.043682263052353856,  
'Impetigo': 0.1441766106480116,  
'Jaundice': 0.055863802668430385,  
'Malaria': 0.04184641012666852,  
'Migraine': 0.045926798979937095,  
'Osteoarthritis': 0.05848399011768744,  
'Paralysis (brain hemorrhage)': 0.41466098140482055,  
'Peptic ulcer disease': 0.10328803030986862,  
'Pneumonia': 0.024423020142999012,  
'Psoriasis': 0.08096796127646837,  
'Tuberculosis': 0.021387069013669272,  
'Typhoid': 0.029587648713441043,  
'Urinary tract infection': 90.42662461719962,  
'Varicose veins': 0.3651929313327923,  
'hepatitis A': 0.030051155196802487}
```

Fig. 4 Probabilities of each disease as predicted by the model

include the exact time limit for the appointment booking. The location of the doctor can be known to the patient using Google Maps API. We can also include payment methods like UPI, credit card billing, etc. We can also provide an electronic health record (EHR) facility for large type organizations.

NAME	SPECIALISATION	RATING	PHONE NUMBER	ADDRESS
idbys1	Dermatologist	4.0	02466786	indore

NAME	SPECIALISATION	RATING	PHONE NUMBER	ADDRESS
idbys5	Allergist	3.75	02466786	indore

Fig. 5 Recommended doctors

## References

1. Caballé, N. C., Castillo-Sequera, J. L., Gómez-Pulido, J. A., Gómez-Pulido, J. M., & Polo-Luque, M. L. Machine learning applied to diagnosis of human diseases: A systematic review. *Journal of Applied Sciences*.
2. Dhomse Kanchan, B., & Mahale Kishor, M. Study of machine learning algorithms for special disease prediction using principal of component analysis. In *2016 international conference on global trends in signal processing, information computing and communication*.
3. Hasan, M. A., Sher-E-Alam, K. M., & Chowdhury, A. R. (2010). Human disease diagnosis using a fuzzy ExpertSystem. *Journal of Computing*, 2(6).
4. Laxmi, P., Gupta, D., Radhakrishnan, G., Amudha, J., & Sharma, K. (2021). Automatic multi-disease diagnosis and prescription system using bayesian network approach for clinical decision making. In *Advances in artificial intelligence and data engineering* (Vol. 1133, pp. 393–409). Springer.
5. Mohanty, A., Parida, S., Nayak, S. C., Pati, B., Panigrahi, C. R. (2022). *Study and impact analysis of machine learning approaches for smart healthcare in predicting mellitus diabetes on clinical data. Smart healthcare analytics: State of the art. Intelligent systems reference library* (Vol. 213). Springer. [https://doi.org/10.1007/978-981-16-5304-9\\_7](https://doi.org/10.1007/978-981-16-5304-9_7)
6. Sasikala, T., Rajesh, M., & Sreevidya, B. (2020). Prediction of academic performance of alcoholic students using data mining techniques. In *Cognitive informatics and soft computing, advances in intelligent systems and computing* (Vol. 1040). Springer.
7. Sreevidya, B., Rajesh, M., & Sasikala, T. (2019). Performance analysis of various anonymization techniques for privacy preservation of sensitive data. In *International conference on intelligent data communication technologies and internet of things (ICICI) 2018, lecture notes on data engineering and communications technologies* (Vol. 26). Springer.
8. Das Mohapatra, S., Nayak, S. C., Parida, S., Panigrahi, C. R., Pati, B. (2021). *COVTrac: Covid-19 tracker and social distancing app*. In C. R. Panigrahi, B. Pati, B. K. Pattanayak, S. Amic & K. C. Li (Eds.), *Progress in advanced computing and intelligent engineering. Advances in intelligent systems and computing* (Vol. 1299). Springer. [https://doi.org/10.1007/978-981-33-4299-6\\_50](https://doi.org/10.1007/978-981-33-4299-6_50)
9. Krishna, C. S., & Sasikala, T. (2019). Healthcare monitoring system based on IoT using AMQP protocol. In *International conference on computer networks and communication technologies, lecture notes on data engineering and communications technologies* (Vol. 15). Springer.
10. Zadeh, L. A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems On Elsevier*, 11(1–3), 197–198.
11. National Research Council Canada. Available <http://www.nrcnrc.gc.ca/eng/index.html>

12. Yourdiagnosis, an online diagnosis tool. Available <http://www.yourdiagnosis.com/>
13. Easydiagnosis, an online diagnosis tool. Available <http://easydiagnosis.com/>
14. Symptoms of different diseases. Available <http://www.wrongdiagnosis.com/>
15. Symptoms of different diseases. Available <http://www.webmd.com/>