Journal of
Computational and Theoretical Nanoscience
Vol. 17, 1–5, 2020

Smart Health Care ChatBot for Prognosis of Treatments and Disease Diagnosis Using Machine Learning

R. Vineeth*, R. Rithish, D. V. S. N. Sai Varma, and B. V. Ajay Prakash

Department of Computer Science and Engineering, SJB Institute of Technology, Bengaluru, Karnataka, India

In this present world there are various diseases for which treatments and remedies are available abundantly. It is impossible for human to remember all the precautions and remedies to cure the disease. There is no relevant platform that could exhibit all the diseases and their respective remedies. Health professionals are not always available to users on all the time. Hence, the necessity of health care chatbot plays a major role in this current world. In the proposed idea, we have created a HealthCare chatbot with Artificial Intelligence techniques which can process the text input and predict the diseases associated with the symptoms given by the user. The HealthCare chatbot implemented here is a user friendly platform which predicts the probable diseases and the home remedies, we can imply to cure based on the symptoms observed by the user in their knowledge.

Keywords: Machine Learning, Text Processing, Artificial Intelligence, Prognosis, Chatbot.

RESEARCH ARTICLE

1. INTRODUCTION

Artificial Intelligence is an approach to make a system think and respond how usually human does. Artificial Intelligence mainly emphasizes on how to improve knowledge on decision making, reasoning, learning, thinking, problem solving and perception. Artificial Intelligence has made chatbot seem more lifelike than ever before and they are becoming persuasive. Chatbot system acts as a substitute to human interaction, which is based on Artificial Intelligence, using this systems some advantages can be drawn to the medical field.

A Chatbot system [5] is a user interactive system or program which communicates with user using text processing. The main purpose of a chatbot is to create a conversational environment, which in our case it includes the input texts, symptoms given by user and the response as predicted from chatbot. The chatbot built in the proposed paper uses machine learning models such as random forest, decision tree, naive bayes to lend help to the user in diagnosing the symptoms faced by the user. Improvisation in the medical field makes a major impact in the society and hence the predicted prognosis and its corresponding remedies proposed in this paper plays a prominent role.

2. LITERATURE SURVEY

The health care chatbot study reviews medical self-diagnosis chatbot for accurate analysis of diseases using Artificial Intelligence [2].

In Madhu et al. [1], reviews the personalized medical assistance which heavily depends on AI and machine learning algorithms with the trained data as presented in the paper. The paper referred has implemented features such as—building a chatbot system which provides real time environment for interaction and it also provides the amount of dosage to be taken based on user's age. A chatbot system which is able to resolve all the problems regarding a medicine. In this review paper, we come across a few of the challenges such as—Research and implementation cost which are not yet addressed in the paper and also the accuracy of each of the algorithms used not being mentioned.

According to Divya et al. [2] survey regarding all the methodologies, implementation, results and issues of already proposed papers have been made. Gathering information regarding the papers would be quite helpful to find the research gap and the accuracies in each of those papers. In Saurav et al. [3], the chatbot being implemented creates a virtual environment for the user and makes possible for the user to interact with virtual doctor. In this referred paper, a chatbot named Vdoc has been designed which acts as a virtual doctor and provides the answer for several types of queries related to disease. The algorithms being implemented are Natural language processing and pattern matching algorithm using the Python Language. Based on the survey it is found that the accuracy given by the chatbot is 80%.

In Hameedullah et al. [4], describes a Unified Medical Language System (UMLS) based chatbot for medical students. The methods used for the development of

*Author to whom correspondence should be addressed.

this chatbot are Artificial Intelligence Modelling Language (AIML) pattern technique for pattern matching. As per the reference survey made in the paper [2] the accuracy given by the chatbot is 47%. However the natural dialog in chatbot has not been supported by the system.

In Jamilu Awwalu et al. [6], evaluates and determines the strength of the Artificial Neural Networks (ANN), Support Vector Machines (SVM), Naive Bayes and Fuzzy-Wuzzy Logic in diagnosing medical problems and depicts the required results. Based on the survey, It can be observed that accuracy of the ANN and SVM proposed better result for the model than compared to the other algorithms which is 90% for ANN and 86.4% for SVM. The implementation of Personalized Medicine relies heavily on Artificial Intelligence and Machine Learning algorithms.

In Doina et al. [7], conveys regarding the expert systems for medical predictions of Hepatitis C through Artificial Neural Networks is implemented in MATLAB 7.0, and provides a very attractive environment with high performance. In Chantarotwong [8] conveys how to train the chatbot to mimic the personality of a human. The basic question and answering text responses has been portrayed to convey the idea of learning chatbot.

The paper proposed here can be further enhanced by inculcating the text to speech processing using Natural Language processing as proposed in expert systems by du Preez et al. [9] which provides a speech to text and text to speech for a user friendly platform to the clients.

RESEARCH ARTICLE

3. PROPOSED METHODOLOGY

From the literature survey, there is need of more accurate prognosis system for health care. In this concern, diagnosis of patient problem and predicting the treatments using machine learning is proposed. The proposed methodology is as shown in Figure 1.

3.1. User

Initially, User interacts with the healthcare chatbot for medical assistance and to provide information regarding the symptoms they would be facing. The input is given in the form of symptoms with general text processing. The user would expect diagnosis of the symptoms to be provided from the healthcare chatbot as a response.

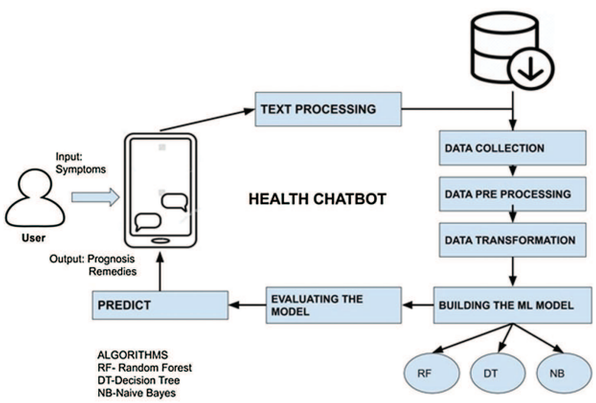

Fig. 1. Proposed architecture for predicting the disease and remedies.

3.2. HealthCare Chatbot User Interface

HealthCare Chatbot interface is the application providing the platform for conversation between the user and the bot, through which the user gives the inputs and the chatbot responds to the corresponding queries.

3.3. Text Processing

This is the basic preprocessing that is done to the user input using natural Language processing and FuzzyWuzzy logic. The FuzzyWuzzy logic is a python library has been implemented for string matching. It uses the logic of matching the user input strings with the question and answering texts trained to the healthcare chatbot. The FuzzyWuzzy logic is mainly based on the Levenshtein Distance to calculate the differences between sequences. The user text is then processed for data abstraction and the data being collected is stored in the form of lists.

3.4. Data Collection

Data Collection is done by retrieval of data from the user and the data comprises of various symptoms, respective diseases and remedies given by the dataset. This Dataset [10] as shown in Tables I and II, comprises of both structured and unstructured data which is further processed, to transform into the supported algorithms. The trained dataset used comprises of 4921 instances with 132 attributes of symptoms. The Dataset [10] referred can be accessed publicly from Kaggle. The Table I presented below comprises of the first five entries whereas the Table II comprises of the last five entries of the dataset.

3.5. Data Pre Processing

The data being collected from the trained dataset and the data from text processing is then pre processed to fit and train the model based on the input data received. The symptoms provided by the user would play a major role in data pre processing based on which the model is being trained.

3.6. Data Transformation

The trained dataset sample is split into 80% training and 20% for testing based on which the model is being tested. A random number is generated to test the predictions for the trained model. The tested results on the built data model is seen to have 100% reliable, which helps for more accurate predictions to be done by the Machine Learning models.

Table I. First five entries of sample dataset.

	Itching	Skin_rash	Nodal_skin_eruption	Continous_sneezing	shivering	chills	Joint_pain	Stomach_pain	Acidity
0	0	0	0	0	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0
2	1	0	1	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0
4	1	1	1	0	0	0	0	0	0

Table II. Last five entries of sample dataset.

	Itching	Skin_rash	Nodal_skin_eruption	Continous_sneezing	Shivering	Chills	Joint_pain
4915	0	0	0	0	0	0	0
4916	0	1	0	0	0	0	0
4917	0	0	0	0	0	0	0
4918	0	1	0	0	0	0	1
4919	0	1	0	0	0	0	0

3.7. Building the Machine Learning Model

3.7.1. Random Forest

Random Forest Algorithm also known as Random Decision Forest is a supervised classification model are used in constructing decision trees for training and to get the mean prediction of all involved decision trees. Unlike Decision trees which uses information of gain index for calculating root node, here in random forest process of finding root node and splitting featured nodes will happen randomly.

$$\hat{f} = \frac{1}{B}\prod_{b=1}^{B} f_b(x') \tag{1}$$

3.7.2. Decision Tree

Decision Tree algorithm is used for solving problems on regression and classification. There are many algorithms to build a decision tree which may include-CART, C4.5 and ID3. This paper is based on developing a decision tree using ID3 (Iterative Dichotomiser 3). In the proposed idea decision tree is used to create a training model with symptoms that can be used to predict a disease. The decision tree algorithm solves the problem by using tree representation. The internal node is an attribute of symptoms provided by the user and the leaf node refers to the symptoms label.

At the beginning, the whole training dataset is considered as the root node of the decision tree. We start comparing the values of root attributes with the record's attribute and continue comparing the record's attributes with other internal nodes until we reach a leaf node with predicted class value of prognosis.

3.7.3. Naive Bayes

Naive Bayes algorithm is an algorithm for predictive modelling based on Bayes Theorem. In the proposed paper the dataset used has the features with row of symptoms. Prognosis and its corresponding remedies would be the response vector. The response vector depicts the output of the class.

Bayes' theorem is based on the probability of an event occurring with an event already taken place. The theorem can be stated with the equation:

$$P(A/B) = \frac{P(A)\cdot P(B/A)}{P(B)} \tag{2}$$

Where, A and B are two events.

ALGORITHM (TRAINING PHASE).

Input: Dataset consisting of symptoms and prognosis

Output: Prediction of disease and it's respective remedy based on the trained dataset

Method:

1. Read the training. csv file for training phase
2. Text processing applied to clean data
3. Label the symptoms from the user data text into a list.
4. Split the preprocessed data for training(Ratio 80:20, Random = 0)
5. Create a Machine Learning model for analyzing the prognosis and remedies
6. Random forest, decision tree, naive bayes algorithms are applied to the proposed models
7. Evaluate the machine learning models for obtaining the performance in terms of precision, recall, accuracy and F1 score.

ALGORITHM: Training phase ends

ALGORITHM: Predicting Phase

Input: User text input of symptom.

Output: Prognosis and corresponding remedies are predicted based on user input of symptoms

Method:

1. Read the testing. csv for predicting prognosis.
2. Text processing to clean data from the user input text.

RESEARCH ARTICLE

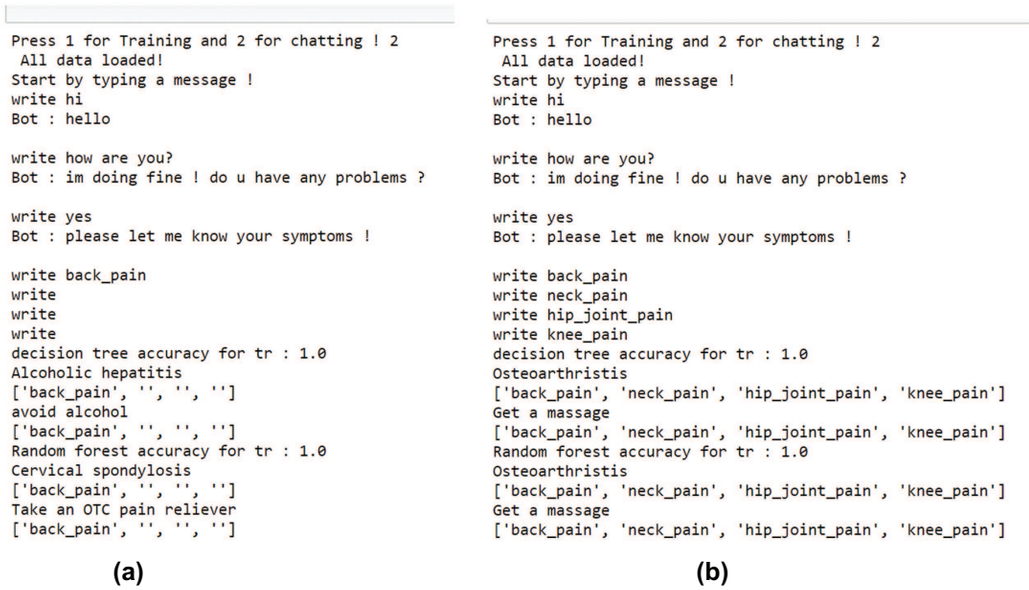

Fig. 2. (a) User and health chatbot interaction with one symptom. (b) User and health chatbot interaction with more than three symptoms.

3. The cleaned data of symptoms is concatenated into the list.
4. Apply the built machine learning model such as random forest, decision tree, naive bayes to predict the prognosis and its corresponding remedies.

ALGORITHM: Predicting phase ends.

3.8. Evaluating the Model

The built machine learning models such as random forest, decision tree and naive bayes proposed in the paper can be evaluated using numerous measures like accuracy, precision, recall and F1 score. In the proposed models the accuracy can be used as one of the metrics for evaluating classification models. Informally, it can be stated that accuracy is the fraction of predictions our model got right. The accuracy of the built training models was found out to be 100%.

3.9. Predict

This is the final stage of the model where the required prediction of disease is done and the corresponding remedy is shown as a chat via chatbot to user. The prediction done by the models mentioned is of cent percent accurate and is evaluated and shown through the measures mentioned.

4. IMPLEMENTATION AND RESULTS

For the implementation of the proposed machine learning models we have used Python programming language. We started with setting up the basic text processing using Fuzzy Wuzzy logic which is written in python language code. The text processing is done for filtering the text given by the user and the text to be replied is trained with various assumed phrases given to the program as file inputs.

The main operation starts when the bot encounters in a chat, where the user is keen to know the disease and remedies to be given to them for assistance. As shown in Figure 1 chat is encountered by the bot, it asks the user to type in the symptoms. These Input symptoms are then appended to the new list. During the training phase the model is trained according to the trained dataset. The user inputted symptoms are then applied to that of the machine learning which in return predicts the prognosis and remedies going through the model as shown in Figure 2. Finally, the predicted result is given to user on the next chat. When the user enters only one symptom as shown in Figure 2(a), the health chatbot provides the probable diseases, whereas when the user enters more than three symptoms as shown in Figure 2(b), the disease predicted will be the same for all the proposed machine

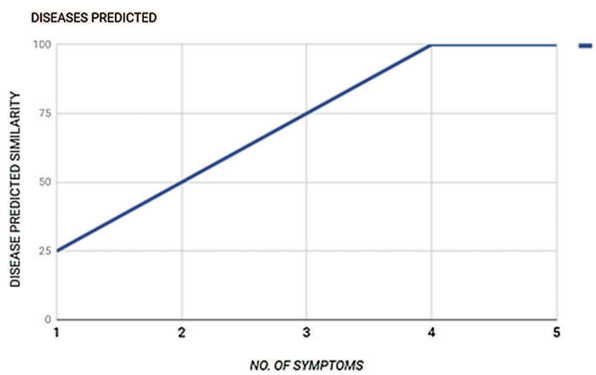

Fig. 3. Graph depicting the disease similarity.

RESEARCH ARTICLE

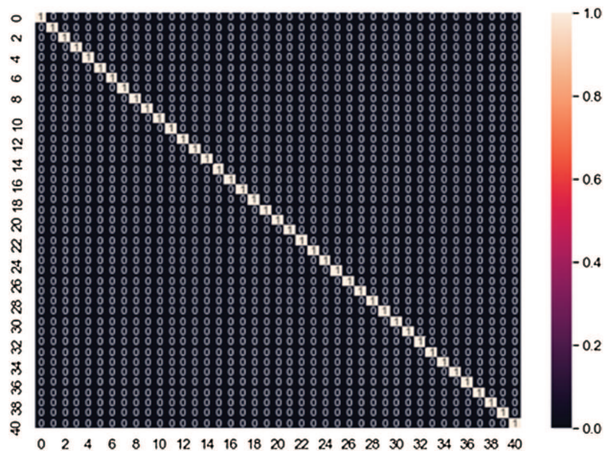

Fig. 4. Confusion matrix depicting the actual and predicted prognosis results.

learning models. The concept is depicted through a graph in Figure 3. A confusion matrix has been used to depict the relationship between the actual and predicted prognosis results as shown in Figure 4.

5. CONCLUSION

Chatbot in recent times have evolved to a great extent, which has made it easy for humans to get their queries replied with ease and in a fraction of time, using his own natural way of questioning. Using this facility, Health Care ChatBot is being designed to help the user to get the diagnosis of the disease that is being predicted with the help of symptoms provided and get a general home remedy that they can follow to get the cure. This model is helpful, not only to the user who are suffering from prognosis but also helpful for one who is keen to get the knowledge of various diseases and remedies and get them forestalled. In the proposed models, we have gone through various symptoms and corresponding diseases that it could cause obtained by that of the datasets of hospitals. The built machine learning models such as Random Forest model, Decision Tree model and Naive bayes model are explored for the forecast. The accuracy obtained from our preprocessed data using these models are of cent percent which guarantees that prediction obtained is complete trustworthy. Since, it is very critical in predicting the prognosis and its corresponding home remedy in the medical field, our model has been proposed with cent accuracy for accurate results.

References

1. Divya Madhu, Neeraj Jain, C.J., Elmy Sebastain, Shinoy Shaji and Anandhu Ajayakumar, **2017**. A Novel Approach for Medical Assistance Using Trained Chatbot. *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, IEEE. pp.243–246.
2. Divya, S., Indumathi, V., Ishwarya, S., Priya Shankari, M. and Kalpana Devi, S., **2018**. Survey on medical self diagnosis chatbot for accurate analysis using artificial intelligence. *International Journal of Trend in Research and Development*.
3. Mishra, Saurav Kumar, Dhirendra Bharti and Nidhi Mishra, **2018**. Dr. Vdoc: A medical chatbot that acts as a virtual doctor. *Research & Reviews: Journal of Medical Science and Technology, 6*(3), pp.16–20.
4. Kazi, H., Chowdhry, B.S. and Zeesha Memon, **2012**. Medchatbot: An umls based chatbot for medical students. *International Journal of Computer Applications, 55*(17).
5. Bayu Setiaji and Wibowo, F.W., **2016**. Chatbot using a knowledge in database: Human-to-Machine Conversation Modeling. *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, IEEE. pp.72–77.
6. Jamilu Awwalu, Ali Garba, Anahita Ghazvini, Rose Atuah, **2015**. *Artificial Intelligence in Personalized Medicine Application of AI Algorithms in Solving Personalized Medicine Problems*, December.
7. Drăgulescu, D. and Albu, A., **2007**. *Medical Predictions System*. Politehnica University of Timisoara, Vol. 4.
8. Chantarotwong, B., **2006**. The learning chatbot. *Final Year Project*. [Online]: http://courses.ischool.berkeley.edu/i256/f06/projects/bonniejc.pdf.
9. du Preez, S.J., Lall, M. and Sinha, S., **2009**. An Intelligent Web-Based Voice Chat bot. *IEEE EUROCON*, IEEE. pp.386–391.
10. Web resources from Kaggle, **2019**. [Online]: https://www.kaggle.com/neelima98/disease-prediction-using-machine-learning.

Received: 17 September 2019. Accepted: 10 October 2019.

RESEARCH ARTICLE