Advances in Parallel Computing Algorithms, Tools and Paradigms
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC220035

An Improved Chatbot for Predicting Disease and Medicines Using Natural Language Processing with Fuzzy Logic

B. PhaniRaghava[a] and S. Ashok Kumar[b,1]
[a]*Research Scholar, Department of CSE, Saveetha School of Engineering,*
[b]*Project Guide, Department of CSE, Saveetha School of Engineering,*
Saveetha Institute of Medical and Technical Sciences, Saveetha University Tamil Nadu
Saveetha University, Chennai. Tamil Nadu. India

Abstract. The aim is to create an artificial conversation entity(chatbot) using python to predict disease and medicine for healthcare treatments. Two algorithms fuzzy support vector machine algorithms are compared with Decision tree algorithm sample size taken 28. G power of 81% and sample size is calculated using the G power tool. Performances of the score model validated test set accuracy with 95% confidence interval for fuzzy support vector machine algorithm with different sub-samples has 91.60% accuracy comparing with Decision tree which has 87.90% accuracy.Independent Sample T-test a significance difference in accuracy and loss is observed $p<0.005$.From the results it is concluded that proposed algorithm Fuzzy support vector machine will produce better results than the existing algorithm.

Keywords. Decision tree, Fuzzy logic, Accuracy, Innovative Artificial conversation entity, Machine learning.

1. Introduction

The main aim of my study is a medical artificial conversation entity(chatbot) to predict the user's(patient) symptoms and give appropriate medicine to the user after taking all the details of the user and symptoms. If the user symptoms are too complicated to handle for the Innovative artificial conversation entity(chatbot) then it gives suggestions to consult a doctor. In real time use of a healthcare artificial conversation entity(chatbot) is 1) Little time spent going to the doctor's office and 2) A small amount of money is spent on unnecessary treatments and tests. Applications of healthcare artificial conversation entity(chatbot)Health Tracking system Chatbot is also helpful in Mental Health patients and research studies Saurav Kumar Mishra, Dhirendra Bharti, Nidhi Mishra,"Dr.Vdoc: A Medical Chatbot that Acts as a Virtual Doctor", Journal of Medical Science.(Mishra, Bharti, and Mishra 2018)[1]

[1]S. Ashok Kumar,Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical andTechnical Sciences, Saveetha University Tamil Nadu

Healthcare artificial conversation entities(chatbots) are implemented by many researchers to help patients. Around 90 papers are published in IEEE and 30 papers in research study. A similar paper proposed by Rohit Binu Mathew,Sandra Varghese,Sera Elsa Joy,Swanthana Susan Alexartificial conversation entity(chatbot) for Disease Prediction and Treatment Recommendation using Machine Learning shows the implementation of concepts natural language processing and machine learning in medical artificial conversation entity(chatbot). [2](Magnusson 2006). A paper proposed by LekhaAthota,Vinod Kumar Shukla,Nitin Pandey,Ajay Rana (Chatbot for Healthcare System Using Artificial Intelligence) illustrates how Artificial Intelligence is being used to construct a medical chatbot that can identify diseases and offer basic information about them before contacting a doctor.(Maio et al. 2015)[3]. A paper Designing Disease Prediction Model Using Machine Learning Approach proposed by DhirajDahiwade,Gajanan Patle,EktaaMeshram shows the implementation of disease prediction model using knn and cnn machine learning algorithms for accurate prediction of disease (Kasthuri and Balaji 2021)[4]. A similar paper proposed by Anjan Nikhil Repaka,Sai Deepak Ravikanti,Ramya G Franklin(Design and Implementing Heart Disease Prediction Using Naives Bayesian) shows the implementation of heart disease prediction using naivesbayesian. The study focuses on the diagnosis of cardiac disease using past data and information. (Baby, Khan, and Swathi 2017)[5].

Our wide portfolio in research has translated into publications in numerous interdisciplinary projects. [6-10](Johnson et al. 2020; Keerthana and Thenmozhi 2016; Thejeswar and Thenmozhi 2015; Krishna and Babu 2016;Jain, Kumar, and Manjula 2014; Nandhini, Babu, and Mohanraj 2018; Kannan and Thenmozhi 2016). Now we are focussing on this topic.

The limitations of an Innovative artificial conversation entity(chatbot) is user privacy. The users may hesitate to share personal information with the software and it can handle the first level questions, they may not be able to solve complicated (complex) questions if it occurs, the Innovative artificial conversation entity(chatbot) suggests consulting a doctor. Previously, our team has extensive expertise working on a variety of research projects across a variety of areas. We decided to explore this project because of the rising trend in machine learning.(“Disease Prediction Based on Symptoms Using Machine Learning” n.d.). Our lab has had rich experience collaborating with various researchers across multiple disciplines[11-15] (Govindaraju, Neelakantan, and Gutmann 2017; Danda et al. 2010; Kumar et al. 2006; Neelakantan et al. 2011; Mathew et al. 2020;Samuel, Acharya, and Rao 2020; Paramasivam, VijayashreePriyadharsini, and Raghunandhakumar 2020; Ezhilarasan, Sokal, and Najimi 2018) The primary goal of the research is to develop a chatbot using python to predict disease and suggest medicine for healthcare treatments.

2. Materials and Methods

The Study setting was done in our university Saveetha school of engineering institute of computer science, department of Big data and network security. In this study 2 sample groups were taken. The group 1 was Fuzzy support vector machine algorithm and group 2 was Decision tree algorithm. pre-test and sample size are calculated using the G-power tool.Sample size is 28. Confidence interval of 95% with pre-test power of 80%.The dataset is taken from the kaggle website. The dataset contains symptoms description.csv, symptom-precation.csv, symptom-severity.csv etc. The dataset

contains disease description and precaution to take in csv files.Some good dataset sources for future comes may be found at r/datasets, UCI Machine Learning Repository, or Kaggle. The larger the dataset, the additional data the model can get to learn from. The statistical software used for our study is the IBM SPSS tool.

The decision surface is learned by a support vector machine (SVM) from two different classes of input points. Each input point may not be fully assigned to one of these two classes in many applications. A fuzzy support vector machine applies fuzzy membership to each input point and reformulates SVMs so that various input points can contribute differently to decision surface learning. (Novak 2017). Table 1 shows the pseudocode for fuzzy support vector machine algorithms. A decision tree is a tree with internal nodes that may be used as tests (on data patterns) and leaf nodes that can be used as categories (of these patterns). To acquire the proper output to the input pattern, these checks are filtered down through the tree.(Maheswari and Pitchai 2019). Below Table 2 shows the pseudocode for the Decision tree algorithm.

Table 1. Pseudocode for Fuzzy Support Vector Machine Algorithm

```
Start
Construct set of rules
Fuzzy set input
Obtain fuzzy set value
Extracting fuzzy rules from SVM:
bo ← b
j ← 1
FORi= 1 TO I
IF αi> 0
zj ← xi
bj ← xiαi
  j ← j+1
END IF
END FOR
m ← j-1
//chatbot code
bot = ChatBot(
storage_adapter="chatterbot.storage.SQLStorageAdapter",
  trainer='chatterbot.trainers.ChatterBotCorpusTrainer',
input_adapter="chatterbot.input.TerminalAdapter",
output_adapter="chatterbot.output.TerminalAdapter"
)
bot.train("chatterbot.corpus.english")
CONVERSATION_ID = bot.storage.create_conversation()
def get_feedback():
from chatterbot.utils import input_function
text = input_function()
if 'yes' in text.lower():
return False
elif 'no' in text.lower():
return True
else:
print('Please type either "Yes" or "No"')
return get_feedback()
END
```

Table 2. Pseudocode for Decision Tree Algorithm

```
FUNCTION DecisionTree():
from sklearn import tree
clf3 = tree.DecisionTreeClassifier()   # empty model of the decision tree
ENDIF
clf3 = clf3.fit(X,y)
psymptoms=[Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]
for k in range(0,len(l1)):
# OUTPUT (k,)
for z in psymptoms:
IF(z==l1[k]):
l2[k]=1
ENDIF
ENDFOR
ENDFOR
inputtest<- [l2]
predict <- clf3.predict(inputtest)
predicted=predict[0]
h='no'
for a in range(0,len(disease)):
IF(predicted = a):
h='yes'
break
ENDIF
ENDFOR
IF (h=='yes'):
t1.delete("1.0", END)
t1.insert(END, disease[a])
ELSE:
t1.delete("1.0", END)
t1.insert(END, "Not Found")
```

2.1 Statistical Analysis

The statistical SPSS software is used to test the procedures to check the validity of the system. The dataset is split into two parts as training and testing setup which after splitting the data set will fix the algorithm into two training and test sets. Training the datasets for dependent and independent have been incorporated.

3. Results

The proposed algorithm Fuzzy support vector machines will give more accurate results for the selected datasets than the Decision Tree. Table 3 shows the accuracy values of Fuzzy support vector machines and Decision tree algorithm. In Table 4 group statistics results are shown Fuzzy support vector machine algorithm and Decision tree algorithm accuracy are 91.60% and 87.90% respectively.Std.error mean for Decision tree algorithmis 0.64031 is less compared with Fuzzy support vector system learning algorithm 0.70238. Figure 1 shows the Bar graph comparison of Fuzzy support vector machine algorithm which has mean accuracy 91.60% compared with Naive bayes which has 87.90% with +/-1 SD.

Table 3. Results of Decision tree and Fuzzy support vector system algorithm (Fuzzy support vector systemalgorithm mean accuracy = 91.60% Decision tree algorithm mean accuracy = 87.90%)

S.no	Fuzzy support vector system Accuracy (%)	Decision tree algorithm Accuracy (%)
1	94	88
2	89	87
3	90	86
4	88	82
5	87	85
6	90	87
7	93	87
8	92	91
9	93	90
10	89	88

Table 4. T-Test Group statistics Results (std.error mean for Decision tree algorithm is 0.64031 is less compared with Fuzzy support vector system learning algorithm 0.70238)

Algorithm	N	mean	Std.Deviation	Std.Error Mean
Fuzzy SVM accuracy	10	91.60	2.22111	.70238
Decision tree accuracy	10	87.90	2.02485	.64031
Fuzzy SVM loss	10	8.40	2.22111	.70238
Decision tree loss	10	12.10	2.02485	.64031

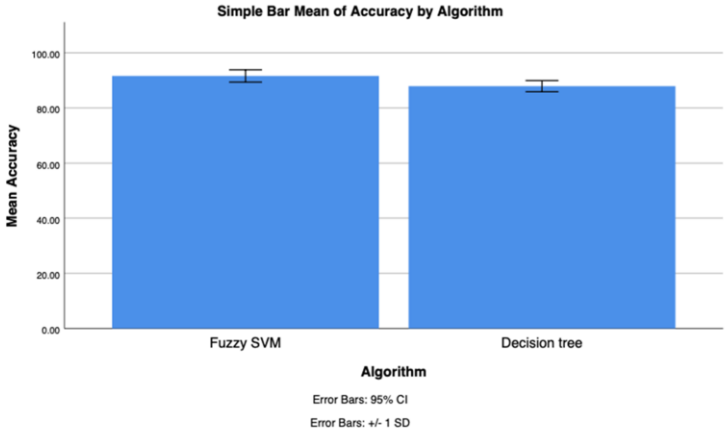

Figure 1. Bar Graph comparison of Fuzzy support vector system which has mean accuracy 91.60% compared with Decision tree algorithm which has 87.90%. The x-axis represents the two groups included and the y-axis represents the algorithms used.We can infer that the value or accuracy of Fuzzy support vector system is more than the Decision tree algorithm accuracy displaying the error bars with a mean accuracy of detection +/-1 SD

4. Discussion

In the previous proposed work is a Decision Tree algorithm which gives mean accuracy of 87.90%. The proposed algorithm is a Fuzzy support vector machine algorithm which gives mean accuracy of 91.60%.

A paper proposed by Rian Budi Lukmanto, Suharjito,AriadiNugroho,Habibullah Akbar. Early Detection of Diabetes Mellitus using Feature Selection and Fuzzy Support Vector Machine(Lukmanto et al. 2019)[16]. In this paper a fuzzy support vector machine is used for detection of diabetes. A paper proposed by MehrbakhshNilashi, Hossein Ahmadi, Azizah Abdul Manaf, Tarik A. Rashid, Sarminah Samad, Leila Shahmoradi, NahlaAljojo&Elnaz Akbari. Coronary Heart Disease Diagnosis Through Self-Organizing Map and Fuzzy Support Vector Machine with Incremental Updates(Lukmanto et al. 2019; Nilashi et al. 2020). In this paper a self-organizing map and fuzzy support vector machine is used for coronary heart disease diagnosis.A paper proposed by Y.Deepthi,K.Pavan Kalyan,Mukul Vyas,K.Radhika, D.KishoreBabu,N.V.KrishnaRao.Disease Prediction Based on Symptoms Using Machine Learning(Lukmanto et al. 2019; Nilashi et al. 2020; "Disease Prediction Based on Symptoms Using Machine Learning" n.d.)(Novák, n.d.). shows the implementation of machine learning algorithms.A paper proposed by SubburajMaheswari,RamuPitchai. Heart Disease Prediction System Using Decision Tree and Naive Bayes Algorithm[17][18](Maheswari and Pitchai 2019).Shows the implementation of decision tree in heart disease predictions. There are no opposing findings for the proposed algorithm.From the literature survey, artificial conversation entity(chatbot)for Disease predictionsFuzzy support vector machine algorithm gives better performance.The limitations of an artificial conversation entity(chatbot) is user privacy,effectiveness and it can handle the first level questions they may not be able to solve complicated (complex) questions. If it occurs the artificial conversation entity(chatbot)suggests consulting a doctor.

In the future, the use of chatbots will increase a lot of companies spending a large amount of money in R&D into chatbots to give better customer support. by using advanced frameworks like keras,tensorflow,pytorch and adding more entity data sets can significantly improve artificial conversation entity(chatbot) performance.

5. Conclusion

In the proposed work fuzzy support vector machine algorithm is showing the better result with accuracy 91.60% compared with Decision Tree algorithm.

References

[1] 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): 16–20.

[2] Magnusson, Martin. 2006. "Natural Language Understanding Using Temporal Action Logic." Proceedings of the Workshop KRAQ'06 on Knowledge and Reasoning for Language Processing - KRAQ '06. https://doi.org/10.3115/1641493.1641499

[3] Maio, Carmen De, Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, and Mimmo Parente. 2015. "Natural Language Query Processing Framework for Biomedical Literature." Proceedings of the 2015

Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology. https://doi.org/10.2991/ifsa-eusflat-15.2015.232

[4] Kasthuri, E., and S. Balaji. 2021. "Natural Language Processing and Deep Learning Chatbot Using Long Short Term Memory Algorithm." Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.04.154

[5] Baby, Cyril Joe, Faizan Ayyub Khan, and J. N. Swathi. 2017. "Home Automation Using IoT and a Chatbot Using Natural Language Processing." 2017 Innovations in Power and Advanced Computing Technologies (i-PACT). https://doi.org/10.1109/ipact.2017.8245185.

[6] Johnson, Jayapriya, Ganesh Lakshmanan, Biruntha M, Vidhyavathi R M, KohilaKalimuthu, and DurairajSekar. 2020. "Computational Identification of MiRNA-7110 from Pulmonary Arterial Hypertension (PAH) ESTs: A New microRNA That Links Diabetes and PAH." Hypertension Research: Official Journal of the Japanese Society of Hypertension 43 (4): 360–62.

[7] Keerthana, B., and M. S. Thenmozhi. 2016. "Occurrence of Foramen of Huschke and Its Clinical Significance." Journal of Advanced Pharmaceutical Technology & Research 9 (11): 1835.

[8] Krishna, R. Nivesh, and K. YuvarajBabu. 2016. "Estimation of Stature from Physiognomic Facial Length and Morphological Facial Length." Journal of Advanced Pharmaceutical Technology & Research 9 (11): 2071.

[9] Jain, Ravindra Kumar, Sridhar Prem Kumar, and W. S. Manjula. 2014. "Comparison of Intrusion Effects on Maxillary Incisors among Mini Implant Anchorage, J-Hook Headgear and Utility Arch." Journal of Clinical and Diagnostic Research: JCDR 8 (7): ZC21–24

[10] Nandhini, J. S. Thaslima, K. YuvarajBabu, and Karthik Ganesh Mohanraj. 2018. "Size, Shape, Prominence and Localization of Gerdy's Tubercle in Dry Human Tibial Bones." Journal of Advanced Pharmaceutical Technology & Research 11 (8): 3604

[11] Govindaraju, Lokhasudhan, Prasanna Neelakantan, and James L. Gutmann. 2017. "Effect of Root Canal Irrigating Solutions on the Compressive Strength of Tricalcium Silicate Cements." Clinical Oral Investigations 21 (2): 567–71

[12] Danda, Anil Kumar, Murali Krishna Tatiparthi, Vinod Narayanan, and Avinash Siddareddi. 2010. "Influence of Primary and Secondary Closure of Surgical Wound after Impacted Mandibular Third Molar Removal on Postoperative Pain and Swelling--a Comparative and Split Mouth Study." Journal of Oral and Maxillofacial Surgery: Official Journal of the American Association of Oral and Maxillofacial Surgeons 68 (2): 309–12."Disease Prediction Based on Symptoms Using Machine Learning." n.d. Accessed May 19, 2021. https://www.springerprofessional.de/en/disease-prediction-based-on-symptoms-using-machine-learning/18334718

[13] Kumar, Muthusamy Senthil, Gelli Vamsi, Ramasamy Sripriya, and Praveen Kumar Sehgal. 2006. "Expression of Matrix Metalloproteinases (MMP-8 and -9) in Chronic Periodontitis Patients with and without Diabetes Mellitus." Journal of Periodontology 77 (11): 1803–8.

[14] Mathew, Mebin George, S. R. Samuel, Ashu Jagdish Soni, and Korishettar Basavaraj Roopa. 2020. "Evaluation of Adhesion of Streptococcus Mutans, Plaque Accumulation on Zirconia and Stainless Steel Crowns, and Surrounding Gingival Inflammation in Primary Molars: Randomized Controlled Trial." Clinical Oral Investigations 24 (9): 3275–80

[15] Ezhilarasan, Devaraj, Etienne Sokal, and Mustapha Najimi. 2018. "Hepatic Fibrosis: It Is Time to Go with Hepatic Stellate Cell-Specific Therapeutic Targets." Hepatobiliary & Pancreatic Diseases International: HBPD INT 17 (3): 192–97

[16] Lukmanto, Rian Budi, Suharjito, Ariadi Nugroho, and Habibullah Akbar. 2019. "Early Detection of Diabetes Mellitus Using Feature Selection and Fuzzy Support Vector Machine." Procedia Computer Science. https://doi.org/10.1016/j.procs.2019.08.140.

[17] Maheswari, S., and R. Pitchai. 2019. "Heart Disease Prediction System Using Decision Tree and Naive Bayes Algorithm." Current Medical Imaging Reviews 15 (8). https://doi.org/10.2174/1573405614666180322141259

[18] Nandhini, J. S. Thaslima, K. YuvarajBabu, and Karthik Ganesh Mohanraj. 2018. "Size, Shape, Prominence and Localization of Gerdy's Tubercle in Dry Human Tibial Bones." Journal of Advanced Pharmaceutical Technology & Research 11 (8): 3604.