AI-Powered Chatbot for Disease Prediction: Advancing Patient Care through NLP and Machine Learning

[1]K. Nimisha,

M .Tech Student,

Dept.of Computer Science Engineering,

Vaagdevi College of Engineering,

Email:nimishakasarla28@gmail.com.

[2]Dr. M. Swapna,

Assistant Professor,

Dept. of Computer Science Engineering,

Vaagdevi College Of Engineering,

Email:swapna.manuparthi@gmail.com.

[3]Dr. E. Balakrishna,

Assistant Professor,

Dept. of Computer Science Engineering,

Vaagdevi College Of Engineering,

Email: balakrishnakits@gmail.com.

[4]Dr. N. Satyavathi,

Associate Professor ,Head of Department,

Dept. Of Computer Science Engineering

Vaagdevi College Of Engineering,

Email:satyanadendla15@gmail.com.

ABSTRACT

In recent decades, the relentless pace of modern life has often caused individuals to overlook regular health care, ultimately compromising their quality of life. However, advancements in Artificial Intelligence (AI) have opened avenues for providing accessible, affordable healthcare services at individuals' convenience. This paper focuses on utilizing AI-driven technologies to empower users with reliable health support, aiming to promote well-being and improve life quality. In today's technology-centric society, AI continues to thrive due to its vast applications, particularly in healthcare. Disease prediction, a focal point of research, leverages big data analytics to enhance the accuracy of health risk assessment, making early diagnosis more reliable and impactful. Yet, establishing robust e-healthcare facilities remains a challenge in developing regions due to limited resources and awareness.

Our study introduces a user-centric platform offering online medical support designed to assist healthcare professionals and streamline patient guidance. Users can access medical information, seek diagnostic insights, and gain awareness of various diseases and treatments. To improve user interaction, we incorporate a chatbot system using AI and Natural Language Processing (NLP) to facilitate disease prediction and offer responsive, human-like assistance. Leveraging the Support Vector Machine (SVM) algorithm for predictive analysis, our chatbot enhances accuracy and accessibility in health services. This proposed solution aims to bridge gaps in healthcare access and support informed health decisions among users worldwide.

Index-Terms: Artificial Intelligence (AI),Disease Prediction,E-Healthcare ,Natural Language Processing (NLP),Support Vector Machine (SVM) ,Big Data Analytics,Health Risk Assessment,Online Medical Services,Healthcare Accessibility ,Predictive Analysis,Patient Assistance System.

1.INTRODUCTION

Since the past few decades, humans have been tirelessly working day in and day out that they fail to prioritize their health on a regular basis. In the longer run, this problem leads to jeopardizing the quality of life. Nevertheless, with the aid of Artificial Intelligence, we can now provide health care services to individuals at their convenience at reasonable prices. One of the biggest blessings we possess is a healthy body. A healthy body and enhanced quality of life is something each one of us looks up to. The primary focus of this paper is to provide these services to fulfill the above mentioned purpose. It is difficult to imagine our lives without high tech gadgets because they have become an essential part of our lives. Therefore the field of Artificial Intelligence is prospering due to the various applications of it in the research field. Disease prediction is one of the main goals of the researchers based on the facts of big data analysis which in turn improves the accuracy of risk classification based on the data of a large volume. E-healthcare facilities in general, are a vital resource to developing countries but are often difficult to establish because of the lack of awareness and development of infrastructure. A number of internet users depend on the internet for clearing their healthcare based queries. We have designed a platform for providing online medical services to patients with a goal to provide assistance to healthcare professionals. The user can also seek medical guidance in an easier way and get exposure to various diseases and diagnosis available for it. In order to make communication more effective, we have implemented a chatbot for disease prediction. Chatbots are the human version of software that is based on AI and uses Natural language processing (NLP) to interpret and accordingly respond to the user. This study proposes the disease prediction chatbot using the concepts of NLP and machine learning algorithms. The prediction is carried out using the SVM algorithm.

1.1.OBJECTIVE

The objective of the healthcare chatbot project is to provide users with a convenient and efficient way to access healthcare information and assistance through natural language conversation. By leveraging artificial intelligence and natural language processing technologies, the chatbot aims to offer personalized responses to user queries regarding symptoms, medical conditions, treatment options, medication reminders, appointment scheduling, and general

healthcare guidance. The ultimate goal is to improve access to healthcare services, enhance patient engagement, and alleviate the burden on healthcare providers by offering timely and accurate support to users anytime, anywhere.

1.2 PROBLEM STATEMENT

The problem statement for the healthcare chatbot project is to address the challenges faced by individuals in accessing timely and reliable healthcare information and services. This includes difficulties in scheduling appointments, managing medications, obtaining accurate health-related information, and navigating the complex healthcare system. Additionally, there is a need to enhance patient engagement and support proactive health management practices. The healthcare chatbot aims to mitigate these challenges by providing users with a convenient and personalized platform for accessing healthcare assistance, guidance, and support in real-time, thereby improving overall healthcare accessibility, efficiency, and patient outcomes.

2.LITERATURE SURVEY

Conversational AI agents, specifically chatbots or "Pregbots," have become an effective tool in addressing health care needs, especially for women's obstetric and mental health. Chung et al. developed "Dr. Joy," a Pregbot designed to provide perinatal women and their partners with obstetric and mental health support through KakaoTalk, a mobile instant messaging platform. The Pregbot uses a Q&A knowledge database to respond to user queries and was evaluated through contextual usability testing. Participants engaged with Dr. Joy over a seven-day period, completing daily tasks to interact with the Pregbot and provide feedback. The study highlighted that users found the Pregbot beneficial, offering a positive user experience and an efficient way to access vital health information during different stages of pregnancycoping review on the role of Artificial Intelligence in pregnancy, Romero-Ternero et al. analyzed research on AI and Affective Computing for managing pregnancy health. The study reviewed literature spanning 12 years to identify AI methodologies and techniques applied to pregnancy-related health and well-being. Notably, it emphasized the impact of emotional factors—such as stress, anxiety, and depression—on pregnancy. Despite the acknowledged importance of emotional well-being, the review found limited studies addressing automatic emotion analysis,

suggesting a strong need for future research on AI-driven emotional monitoring for pregnant women .

Chaud. conducted a comparative study of medical Pregbots, examining their current capabilities and limitations. Pregbots in the medical field are increasingly used to handle tasks such as appointment scheduling, patient management, and reminders. However, these systems are not yet advanced enough to handle comprehensive medical assistance duties. In response, the authors proposed "HEALTHBOT," a multilingual Pregbot that can communicate in English and Marathi to assist with patient symptoms, test reports, and basic medical guidance. The Pregbot also intends to streamline the work of medical practitioners by functioning as a personal assistant, bridging the medical treatment gap .

Outside the med, AI chatbots have also been deployed in educational settings. Daswani et al. introduced "CollegeBot," a conversational AI tool designed to assist university students in navigating college information. The bot was implemented with a knowledge base collected from university resources and evaluated using semantic similarity and Sequence-to-Sequence learning models. CollegeBot demonstrated effective context switching and intent recognition, enabling users to access college-related information more efficiently. This approach showcased how AI-powered bots could enhance user engagement and facilitate access to complex organizational information .

For the foundational prord segmentation in conversational AI, various methodologies have been explored. Javed (2015) described a character-space-based segmentation method that calculates average gap distances between characters to tokenize sentences. Lee (2017) introduced the use of the Natural Language Toolkit (NLTK), a Python-based NLP package with inbuilt tokenizers that efficiently segment text without requiring additional algorithmic implementation. Finally, Jaing (2011) discussed a Conditional Random Fields (CRF) model for segmenting sentences based on character spacing thresholds. Of these methods, NLTK stands out for its ease of use, speed, and versatility, making it a preferred choice for tokenization in chatbot development .

3.SYSTEM ANALYSIS

3.1 EXISTING SYSTEM:

In order to facilitate communication between patients and healthcare providers, the

scheme responds quickly to questions submitted by patients. They prefer to spend time on the internet rather than be concerned about their own well-being. As a result, they stop going to the doctor for minor ailments that could turn into more serious ones. Question and answer platforms are now an easy way to find information rather than searching the internet for a list of documents that may be of interest. Current implementations are plagued by issues such as patients not receiving timely responses or having to wait a long time for specialists to acknowledge their concerns. Patients may be charged a fee if they are able to communicate with doctors online via live chat or phone. Now, there is no system or software that can provide the best answer to the most common questions from patients. We went through a lot of books during the subject discussion, but we couldn't find any that were relevant to the strategy

3.1.1 DISADVANTAGES:

- It is used for general conversation not for the specific task.

.

3.2PROPOSED SYSTEM:

User interface is user-friendly like chat window or chat room. Natural Language Processing The chatbot should be trained with vast knowledge of medical information such as symptoms, conditions, treatments, drug information and precautions. The user authentication is required Chatbot can provide preliminary diagnose. Chatbot can connect users to telemedicine services, allowing than to virtual consultation. Chatbot support multiple languages to make the chatbot accessible to wider user base. Ensure the chatbot compiles with relevant healthcare regulations and standards

3.2.1 ADVANTAGE:

1.Users can ask multiple questions as well as symptoms he is going through regarding the medical health issues, and can save much time without consulting a doctor physically.

2.The medical chatbot give output related health issues and to prevent the disease in appropriate manner to the user.

3.Using this web application we can create awareness through the user and the user can get proper medical solutions to find out the exact disease

4.Medical chatbots are designed to assist patients and avoid issues that may arises during daily life and normal business hours, and with 24/7 accessability patient have

immediate access to medical chatbot assistance whenever they need.

4.SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE:

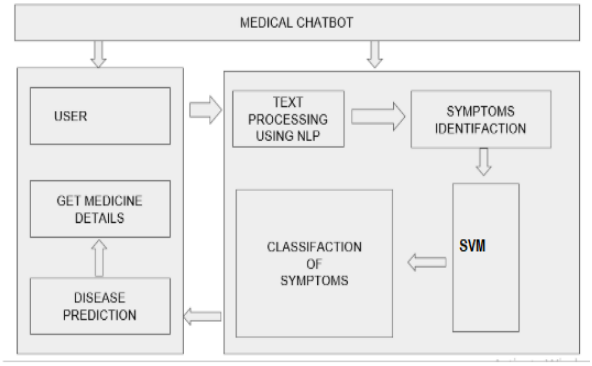


Figure 1. System Architecture

5.IMPLEMENTATION

MODULES:

A medical chatbot using AI typically consists of several modules that work together to provide a comprehensive and effective healthcare assistance experience. Here are some key modules commonly found in such systems:

NATURAL LANGUAGE PROCESSING :

NLP is the parsing and semantic interpretation of human-generated text, allowing machines to learn, analyze, and understand the context. By using NLP algorithms, valuable insights regarding the contextual, behavioral, and sentiment segmentation of the data stream can be achieved. The input and output of an NLP system can be voice, text, and image. There are various techniques that are heavily used during the NLP process, including but not limited to: grammar induction [2], which is used to produce a formal grammar with no given context; lemmatization [3] to identify word's lemma according to the meaning within the context; morphological segmentation [4] task to split words into individual elements and recognize their class; part-of-speech tagging [5] technique that can identify words with similar grammatical properties; "bag of words" [6] to tokenize-vectorize words after split from sentences; and word embedding [7] algorithm to extract features of words with the same meaning based on semantic lookalike relationships and same vector space distance. The most important module of NLP is natural language understanding (NLU).

NATURAL LANGUAGE UNDERSTANDING

NLU is the subset of the understanding and comprehension part of NLP. It

elucidates the concept of the input string and transforms the unstructured data into classified data assigning them to the appropriate intents. In order to distinguish the meaning, classify and conclude the correct intent from the provided input, specific techniques are used like sentiment and content analysis. The systems that heavily rely on NLP are the AI ChatBots.

AI CHATBOTS

AI ChatBot's (hereafter called "ChatBots" for simplicity) objective is to use any applicable technology in order to mimic the conversation among human beings, achieved by the NLP algorithms.

1. **Natural Language Processing (NLP) Module**: This module is responsible for processing user inputs, understanding their intent, and extracting relevant information. NLP techniques are used to analyze text data and identify keywords, entities, and intents within user queries.
2. **Knowledge Base Module**: The knowledge base module contains a repository of medical information, including symptoms, diseases, treatments, medications, and healthcare guidelines. It serves as a reference source for the chatbot to provide accurate and up-to-date information to users.
3. **Diagnostic Module**: The diagnostic module uses AI algorithms to analyze user symptoms and medical history to provide preliminary diagnoses or recommendations. It may employ decision trees, rule-based systems, or machine learning models to assist in medical decision-making.
4. **Treatment Recommendation Module**: This module suggests appropriate treatment options based on the user's diagnosis, medical history, and preferences. It may provide information on medications, therapies, lifestyle modifications, and preventive measures tailored to the user's condition.
5. **Appointment Scheduling Module**: The appointment scheduling module facilitates the booking of appointments with healthcare providers. It integrates with scheduling systems to check availability, book appointments, send

reminders, and manage user calendars.

6. **Medication Management Module**: This module helps users manage their medications by providing reminders for dosage schedules, medication refills, and potential drug interactions. It may also offer information on medication usage, side effects, and precautions.
7. **User Profile Management Module**: The user profile management module stores and manages user information, including personal details, medical history, preferences, and past interactions with the chatbot. It ensures personalized and tailored assistance for each user.

6. ALGORITHM

Algorithms: Supporting vector machines (SVMs, support vector networks too) are supervised models with related learning algorithms that evaluate data used to assess classification and regression. In view of a collection of instances of training, each of which is marked one or the other of two categories, the SVM training algorithm creates a model that assigns fresh examples to each of the categories, making it a non-probabilistic binary linear classifier.An SVM model shows the instances as spatial points, mapped so as to split the examples of the different categories into a distinct break that is as broad as feasible. New instances will subsequently be mapped to the same space and will belong to a class depending on their side of the gap. Besides the linear classification, SVMs may conduct an effective non-linear classification utilising what is known as the kernel technique to map their inputs to high dimension function spaces. Software Requirement: In a software-based solution running on a Linux or window system with a 2.4-GHz dual core CPU is presented Purpose: The major objective of the system is to construct a linguistic gap between the user and healthcare professionals by providing fast answers to the user's questions. People nowadays are more likely to be addicted to the internet, yet they are unconcerned about their own health. They avoid coming to the hospital for minor issues that might develop into a serious sickness in the future. Creating question-andanswer forums is becoming an easy approach to answer those inquiries rather than sifting through a list of possibly relevant online pages. Many present systems have drawbacks, such as the fact that patients do not receive an immediate answer; instead, they must wait for

specialists to recognise their request, which might take a lengthy time. Some processes may demand a fee for online live chat or telephonic connection with doctors. The goal of this technology is to duplicate a person's conversation

Support Vector Machine (SVM) is a powerful machine learning algorithm used for linear or nonlinear classification, regression, and even outlier detection tasks. SVMs can be used for a variety of tasks, such as text classification, image classification, spam detection, handwriting identification, gene expression analysis, face detection, and anomaly detection. SVMs are adaptable and efficient in a variety of applications because they can manage high-dimensional data and nonlinear relationships.

SVM algorithms are very effective as we try to find the maximum separating hyperplane between the different classes available in the target feature.

Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well it's best suited for classification. The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space. The hyperplane tries that the margin between the closest points of different classes should be as maximum as possible. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

Let's consider two independent variables x_1, x_2, and one dependent variable which is either a blue circle or a red circle.

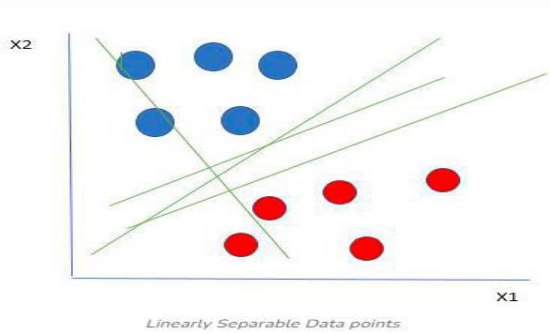


Figure.2 Support Vector Machine

From the figure above it's very clear that there are multiple lines (our hyperplane here is a line because we are considering only two input features x_1, x_2) that segregate our data points or do a

classification between red and blue circles. So how do we choose the best line or in general the best hyperplane that segregates our data points?

1. **Hyperplane:** Hyperplane is the decision boundary that is used to separate the data points of different classes in a feature space. In the case of linear classifications, it will be a linear equation i.e. $wx+b = 0$.
2. **Support Vectors:** Support vectors are the closest data points to the hyperplane, which makes a critical role in deciding the hyperplane and margin.
3. **Margin**: Margin is the distance between the support vector and hyperplane. The main objective of the support vector machine algorithm is to maximize the margin. The wider margin indicates better classification performance.
4. **Kernel**: Kernel is the mathematical function, which is used in SVM to map the original input data points into high-dimensional feature spaces, so, that the hyperplane can be easily found out even if the data points are not linearly separable in the original input space. Some of the common kernel functions are linear, polynomial, radial basis function(RBF), and sigmoid.
5. **Hard Margin:** The maximum-margin hyperplane or the hard margin hyperplane is a hyperplane that properly separates the data points of different categories without any misclassifications.
6. **Soft Margin:** When the data is not perfectly separable or contains outliers, SVM permits a soft margin technique. Each data point has a slack variable introduced by the soft-margin SVM formulation, which softens the strict margin requirement and permits certain misclassifications or violations. It discovers a compromise between increasing the margin and reducing violations.
7. **C:** Margin maximisation and misclassification fines are balanced by the regularisation parameter C in SVM. The penalty for going over the margin or misclassifying data items is decided by it. A stricter penalty is imposed with a greater value of C,

which results in a smaller margin and perhaps fewer misclassifications.

8. **Hinge Loss:** A typical loss function in SVMs is hinge loss. It punishes incorrect classifications or margin violations. The objective function in SVM is frequently formed by combining it with the regularisation term.
9. **Dual Problem:** A dual Problem of the optimisation problem that requires locating the Lagrange multipliers related to the support vectors can be used to solve SVM. The dual formulation enables the use of kernel tricks and more effective computing.

7.RESULTS

Figure.3 Home Page

Figure.4 Chat Bot

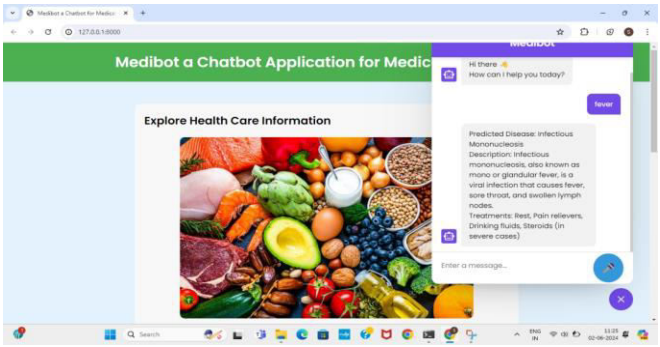
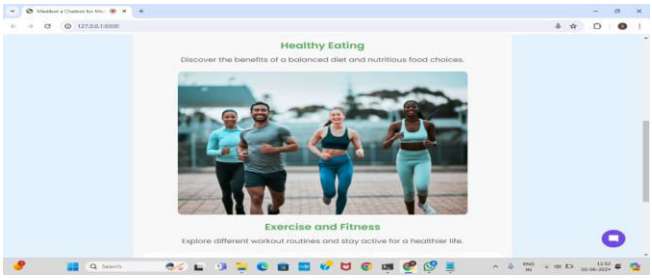


Figure.5 Chatbot Replay



8.CONCLUSION

A Chatbot is a great tool for conversation. Here the application is developed to provide quality of answers in a short period of time. It removes the burden from the answer provider by directly delivering the answer to the user using an expert system. The project is developed for the user to save the user their time in consulting the doctors or experts for the healthcare solution. Serve as 24/7 companions, monitor health status in real-time, and automatically call for assistance in case of an emergency. Help manage chronic - conditions, mental health issues, and behavioral and psychological disorders. Proactively identify symptoms, crosscheck them against medical history, suggest the next steps, and improve the

treatment success rate in cases where early diagnosis can play a critical role. Make self-care easier by acting as a virtual assistant and providing timely medical advice. Thus, we can conclude that this system giving the accurate result. As we are using large dataset which will ensures the better performance. Thus we build up a system which is useful for people to detect the disease by typing symptomsThus, we can conclude that this system giving the accurate result. As we are using large dataset which will ensures the better performance. Thus we build up a system which is useful for people to detect the disease by typing symptomsThus, we can conclude that this system giving the accurate result. As we are using large dataset which will ensures the better performance. Thus we build up a system which is useful for people to detect the disease by typing symptomsThus, we can conclude that this system giving the accurate result. As we are using large dataset which will ensures the better performance. Thus we build up a system which is useful for people to detect the disease by typing symptomsTop of Form Bottom of Form

9.FUTURE ENHANCEMENTS

Chat bots are a phenomenon of the future that has yet to realise its full potential, but with their growing popularity and enthusiasm among businesses, they are certain to stay for a long time. Machine learning has altered the way businesses communicate with their consumers. With the introduction of new platforms for building various sorts of chat bots, it is exciting to see the emergence of a new domain in technology while exceeding the previous barrier

10. REFERENCES

[1] Lokman, A.S.; Ameedeen, M.A. Modern Chatbot Systems: A Technical Review. In Proceedings of the Future Technologies Conference (FTC) 2018; Arai, K., Bhatia, R., Kapoor, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 1012–1023.

[2] Saurav Kumar Mishra, DhirendraBharti, Nidhi Mishra," Dr.Vdoc: A Medical Chatbot that Acts as a Virtual Doctor", Journal of Medical Science and Technolog, Volume: 6, Issue 3,2017. [3] Siddique, Sarkar & Chow, James. (2021). Machine Learning in Healthcare Communication. Encyclopedia. 1. 220-239. 10.3390/encyclopedia1010021.

[4] https://onlim.com/en/the-history-of-chatbots/

[5] A. K. Tripathy, R. Carvalho, K. Pawaskar, S. Yadav and V. Yadav, "Mobile based healthcare management using artificial intelligence," 2015 International Conference on Technologies for Sustainable Development (ICTSD), 2015, pp. 1-6, doi: 10.1109/ICTSD.2015.7095895.

[6] Dharwadkar, R., Deshpande, N.A.: A medical chatbot. Int. J. Comput. Trends Technol. (IJCTT) 60, 41–45 (2018)

[7] Hoermann S, McCabe KL, Milne DN, et al. Application of synchronous text-based dialogue systems in mental health interventions: Systematic review. J Med Internet Res 2017; 19(8): e267. [PMC free article][PubMed] [Google Scholar]

[8] Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles:) the al-chatbot approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, ACM, 2017, pp.261–265.

[9] Comendador BEV, Francisco BMB, Medenilla JS, et al. Pharmabot: A pediatric generic medicine consultant chatbot. J Automat Control Eng 2015; 3(2): 137--140. [Google Scholar]

[10] Divya Madhu, Neeraj Jain C. J, ElmySebastain, ShinoyShaji, AnandhuAjayakumar," A Novel Approach for Medical Assistance Using Trained Chatbot", International Conference on Inventive Communication and Computational Technologies (ICICCT 2017.