

An Intelligent Conversational Healthcare System for Symptom-Based Disease Prediction and Location-Aware Hospital Recommendation Using Machine Learning

¹ Megha Thakre, ² Dr.Sudhir Mohod

¹Student, ²Professor

^{1,2} Department of computer science & Engineering
Bapurao Deshmukh College of Engineering
Wardha , Maharashtra ,India 442104

Abstract : The rapid advancement of artificial intelligence has significantly enhanced the accessibility and efficiency of healthcare services. In today's fast-paced lifestyle, individuals often neglect early symptoms, leading to delayed diagnosis and severe health complications. To address this challenge, this study proposes an intelligent machine learning-based conversational healthcare assistant capable of analyzing user-reported symptoms and predicting potential diseases with high accuracy. The system integrates Natural Language Processing (NLP), machine learning algorithms, and statistical analysis to enable human-like interaction and effective symptom understanding. The proposed chatbot not only provides disease predictions but also offers meaningful insights into pathology reports and delivers location-aware recommendations for nearby healthcare facilities, thereby improving timely medical intervention. A comprehensive review of recent literature from the past five years has been conducted to identify existing gaps and incorporate advanced methodologies, including deep learning techniques and scalable data processing frameworks. Furthermore, the system is designed with a user-friendly web interface, allowing seamless interaction between users and the chatbot. By combining conversational intelligence with predictive analytics and geospatial awareness, the proposed solution aims to bridge the gap between patients and healthcare services. The study highlights the transformative potential of AI-driven chatbots in enabling early diagnosis, improving healthcare accessibility, and supporting self-assessment in a reliable and efficient manner.

IndexTerms - Machine Learning, Conversational Agent, Healthcare Chatbot, Symptom Analysis, Disease Prediction, Machine Learning Artificial intelligence, chatbot, natural language processing, conversational agents, RNN model, geocoder, LSTM , Component, formatting, style, styling, insert.

I. INTRODUCTION

For many healthcare professionals, artificial intelligence (AI) is perceived either as a fantastic notion associated with science fiction or as an overstated technology that is perpetually considered imminent. However, recent technological advancements have impacted this perspective, prompting many to analyze AI's present and potential uses in other domains, including medicine. In particular, advances in natural language processing (NLP) techniques that analyze and produce human-like language have markedly influenced this substantial change in public awareness. Large language models, such as OpenAI's Generative Pretrained Transformer 3 (GPT-3), are indicative of some of the most promising applications in natural language processing (NLP) that are open to the public. Chat-GPT, a recently announced chatbot that makes use of GPT-3, has been trained in conversational tasks using supervised and reinforcement learning. Consequently, it is now considered to be one of the most complete and efficient language processing models currently available. Applied natural language processing (NLP) is related to both human-machine and human-human interactions, and it can be used to evaluate textual data in intelligent healthcare settings. This textual data can be broadly categorized into clinical and non-clinical sources. Clinical text is derived from a wide variety of healthcare institutions and mostly composed of unstructured textual data obtained via electronic health record (EHR) systems.

In recent years, the rapid advancement of digital technologies has significantly transformed the healthcare sector, enabling the development of intelligent systems that enhance accessibility, efficiency, and quality of medical services. With the increasing burden on healthcare infrastructure and the growing population, there is a pressing need for automated solutions that can assist individuals in obtaining preliminary health assessments without immediate dependence on medical professionals. Many people tend to ignore early symptoms due to busy lifestyles, lack of awareness, or limited access to healthcare facilities, which often leads to delayed diagnosis and severe health complications. Therefore, integrating artificial intelligence (AI) and machine learning (ML) into healthcare systems has emerged as a promising approach to address these challenges.

An intelligent conversational healthcare system represents a modern solution that combines natural language processing (NLP), machine learning algorithms, and user-friendly interfaces to provide real-time interaction and support to users. These systems, often referred to as healthcare chatbots, simulate human-like conversations and assist users in identifying potential diseases based on the symptoms they provide. By analyzing symptom patterns and comparing them with trained datasets, such systems can generate

predictive insights and recommend possible medical conditions. This approach not only reduces the workload on healthcare professionals but also empowers users to take proactive steps toward their health.

Symptom-based disease prediction has gained considerable attention in recent years due to its ability to provide quick and cost-effective preliminary diagnoses. Machine learning models such as Decision Trees, Support Vector Machines (SVM), Random Forest, and Gradient Boosting algorithms are widely used for classification tasks in medical diagnosis. These models are trained on large healthcare datasets containing symptoms and corresponding diseases, allowing them to learn complex relationships and patterns. As a result, they can predict diseases with a reasonable degree of accuracy. However, traditional systems often lack user interaction capabilities and do not provide contextual recommendations, limiting their practical usability.

To overcome these limitations, conversational interfaces powered by NLP techniques play a crucial role. Natural language processing enables the system to understand user inputs in plain language, interpret symptoms, and respond appropriately. This makes the system accessible even to users with limited technical knowledge. Moreover, conversational systems can guide users through structured questioning, improving the quality of input data and enhancing prediction accuracy. The integration of NLP with machine learning thus creates a robust framework for intelligent healthcare assistance.

Another significant challenge in healthcare systems is the accessibility of medical facilities. Even after identifying a potential disease, patients often struggle to locate nearby hospitals or healthcare centers that can provide appropriate treatment. This issue is particularly prominent in semi-urban and rural areas where healthcare resources are limited or unevenly distributed. To address this problem, location-aware services can be integrated into the healthcare system. By utilizing geolocation data and mapping technologies, the system can recommend nearby hospitals, clinics, or specialists based on the user's current location and medical needs. This feature enhances the practical utility of the system by bridging the gap between diagnosis and treatment.

The proposed system, titled "An Intelligent Conversational Healthcare System for Symptom-Based Disease Prediction and Location-Aware Hospital Recommendation Using Machine Learning," aims to combine these capabilities into a unified platform. The system leverages machine learning algorithms to predict diseases based on user-provided symptoms, while the conversational interface ensures seamless interaction and improved user experience. Additionally, the integration of location-based services enables the system to recommend nearby healthcare facilities, thereby providing end-to-end support to users.

From a technical perspective, the system involves multiple components, including data preprocessing, feature extraction, model training, and deployment. The dataset used for disease prediction typically contains various symptoms mapped to specific diseases. Data preprocessing techniques such as handling missing values, encoding categorical variables, and normalization are applied to improve model performance. Feature selection methods may also be used to identify the most relevant symptoms for accurate prediction. Once the data is prepared, machine learning models are trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The conversational module is implemented using natural language processing techniques, including tokenization, intent recognition, and entity extraction. These techniques enable the system to interpret user queries and extract relevant symptom information. Advanced NLP models, such as transformer-based architectures, can further enhance the system's ability to understand complex queries and provide context-aware responses. The integration of the chatbot with the prediction model allows for real-time interaction and instant feedback.

The location-aware recommendation module utilizes geographic information systems (GIS) and application programming interfaces (APIs) such as Google Maps to identify nearby healthcare facilities. Based on the predicted disease and user location, the system suggests hospitals or clinics that are best suited to provide treatment. Additional factors such as distance, ratings, and available services can also be considered to improve recommendation quality.

Despite the numerous advantages, the development of intelligent healthcare systems also presents certain challenges. Ensuring data privacy and security is of utmost importance, as healthcare data is highly sensitive. The system must comply with data protection regulations and implement secure data handling practices. Furthermore, the accuracy of predictions depends heavily on the quality and diversity of the training dataset. Bias or imbalance in the dataset may lead to incorrect predictions, which could have serious consequences. Therefore, continuous model evaluation and updates are necessary to maintain reliability.

Another important consideration is the ethical aspect of using AI in healthcare. While these systems can provide preliminary guidance, they should not replace professional medical advice. Users must be clearly informed that the system is intended for informational purposes only and that they should consult qualified healthcare professionals for accurate diagnosis and treatment.

In conclusion, the integration of machine learning, natural language processing, and location-based services has the potential to revolutionize healthcare delivery by making it more accessible, efficient, and user-centric. The proposed intelligent conversational healthcare system addresses key challenges in symptom-based disease prediction and hospital accessibility, offering a comprehensive solution for modern healthcare needs. By providing real-time interaction, accurate predictions, and location-aware recommendations, the system aims to improve health awareness, enable early diagnosis, and facilitate timely medical intervention.

This data includes medical information, diagnostic reports, digital prescriptions, and other interchangeable documents. Non-clinical or supplementary textual data encompasses all other written content relevant to healthcare situations. This includes surveys related to population screening as well as papers that are provided for evidence-based reference. All intelligent healthcare applications involve communication. This includes interactions between patients and providers during clinical examinations, as well as interactions between humans and robots in rehabilitation therapy. Additionally, communication occurs in programs such as automated translation and graphical interfaces for robots for rehabilitation.

The machine learning system uses the prediction model it developed from the prior data to forecast the outcome when it receives new data. The prediction accuracy grows with the amount of data provided; a larger data set facilitates the development of a model that more accurately predicts the outcome. A branch of artificial intelligence called natural language processing (NLP) uses machine learning (ML) methods to teach computers to comprehend, analyse, and draw conclusions from human language. NLP processes and analyses textual data using computer algorithms and machine learning, which aids in sentiment analysis. NLP is an interdisciplinary area that includes linguistics, computer science, and machine learning. It helps computers comprehend and process natural language data, such as text or speech has several applications in areas i.e. Applications of NLP include information retrieval, Sentiment analysis, Chatbots, and language translation. Through the power of advancing computer capabilities, NLP is bridging the communication gap and enabling human-computer interactions for the betterment of future. Chatbot is a program and is being developed using this proposed with AI and NLP. The growing demand for accessible healthcare services is evident, particularly in regions with doctor shortages or where medical facilities are overburdened. According to the World Health Organization (WHO), nearly half the worldwide people lacks access to indispensable health amenities, while those with access often face long wait times for appointments. Healthcare chatbots address these challenges by offering 24/7 automated support, enabling users to describe their symptoms in natural language and receive instant, personalized feedback. Unlike static symptom-checker tools that rely on rigid decision trees, modern AI chatbots leverage machine learning algorithms to interpret complex queries, understand context, and generate accurate responses.

II. RELETED WORKS

It has several attributes (symptoms) as well as classes (diseases). To train the model, we use this to build the training and testing sets. We obtain the user's symptoms and use the qualified model to predict the disease. On the other side, the medical record is compiled, which results in a review of the clinical report focusing on the most significant symptoms associated with a specific illness. This is used to increase the number of disease symptom pairs in the dataset. This paper describes disease prediction using highly personalised training datasets, as well as related tasks such as scheduling appointments and locating the nearest healthcare centre. Disease Prediction by Machine Learning Over Big Data from Healthcare Communities Using structured and unstructured data from hospitals, this paper proposes a new convolutional neural network-based multimodal disease risk prediction algorithm. To our knowledge, no current work in the field of medical big data analytics has centred on both data forms. In comparison to other popular prediction algorithms, our proposed algorithm achieves a prediction accuracy of 94.8% and a convergence speed that is faster than the CNN-based unimodal disease risk prediction algorithm.

The need of the recommender framework is to assist the customers with welling the organizations differently. The techniques utilized are content put together separating that depends with respect to connection between's highlights of illnesses and the infection inputted by the client. The highlights of each sickness is introduced as a ton of qualities in a report or terms, in a perfect world the causes and indications. Generally speaking a substance put together recommender works with respect to the chief to consequently anticipate contingent on inclinations of the client that is accumulated from their information. Half breed sifting — It is a combination of substance based separating and shared separating more or less. On the off chance that Hybrid sifting is utilized it prompts high exactness. It's so in light of the fact that information is missing in regards to decisions of others in a substance based methodology. Hence, in blend of both, normal information increment is there, this basically helps in effectiveness of suggestions. [3] Entities like vehicle, books, melodies or likewise people, these all can be said as items or administrations and recommender frameworks are utilized to work with the clients by discovering such administrations and items. This all relies on the information gathered from the client. The methodology utilized here is conceivable the equivalent to foresee sicknesses for a client dependent on the illness the client is enduring or is probably going to endure. In a substance based framework the client must be certain and give content, so the client can be suggested.

Gyant[8] is another chatbot that helps users with small, non-emergency medical problems. It does this by asking questions about their symptoms and overall health. A main feature of Gyant is that it uses humor, friendly chats, emojis, and memes to connect with users in a fun way. The software isn't available at the moment because it has been stopped. Symptomate is a chatbot powered by AI that assists patients by allowing them to describe their symptoms. It then provides them with a list of possible health problems they may have. Its accuracy is below 80% when compared to Ada Health. This means we need to enhance and create digital health solutions to ensure they are as effective and dependable as we can make them. Tanmay[9] and others. A study on the "E-Health Bot to Change Medicare" showed that chatbots can greatly enhance and transform the delivery of healthcare services. For example, a medical chatbot with up-to-date information used AI to help address the increasing demand for easy access to healthcare help by quickly providing medical assistance and facts. Matthew[10] and his team developed a system where machine learning techniques were used to enhance the ability of the medical chatbots to predict health problems. Chatbots utilised symptoms and the patient's information in identifying the diseases early and suggesting the appropriate treatments. It resulted in better outcomes for the patients and smoother health care work. The study shows the AI powered medical chatbots have a huge scope for the development and prediction of illness and delivery of health care. The AI powered chatbot utilises AI technology and the machine learning algorithms to help solve the various health care problems and makes it easier to access the health care services, makes them work more efficiently and delivers more personalised diagnoses. Further studies are required to see how AI powered medical chatbots perform in a larger scale, how reliable they are and how effective they are in a real healthcare setting. The growth of AI and other digital health technologies such as chatbots will play an increasingly important role in public health. Research also has a crucial role to play in shaping the growth of digital technologies in public health. The focus of the study was on studying the impact of chatbots and other tools on health programs, learning from user experiences and keeping the user safe and healthy. The research involved studying how the tools such as chatbots affected the health programmes, learning from user experiences and focusing on keeping the user safe and healthy. To develop more effective health programmes, it is required to study the tools such as chatbots and their

impact. We must also try to understand what Users feel or go through and how we can keep them safe and healthy. A study found that NLP and Machine learning can be very helpful in fixing many problems in Healthcare. Many people have used dialogflow and tensor flow to create their Chatbots, to see how well a Chatbot can do they used the best training method. We should try to make a better plan that can show how well a chatbot can do and how good it is at helping. If we can make a simple Question and Answer system we can have better communication.

A Comment box can likewise be accommodated client input. The Ultimate objective is to guarantee client fulfillment just as an able outcome set of admonitions to the client with respect to the infections they are inclined to that prompts mindfulness and thinking often about their self-wellbeing. [14] Content based separating calculations consider clients decisions and inclinations and in this manner a rundown or a bunch of most comparable decisions or things or expectations are advised to the client. Different sorts of recommender frameworks like synergistic sifting have the accompanying according to an online business site or application like Flipkart: User enrollment: User may enter his/her subtleties. Further approval and verification is finished utilizing login qualifications like username and passwords. Clients can purchase new items. Information recovery framework: Server will follow every one of the changes, decisions, inclinations and will save them in the data set. The need is to set up the association with speak with the clients and adjust every client's exercises in the application's data set. Approval will be accomplished for each customer before they access the application. This progression guarantees no unlawful access ought to be there from the client's side to get to the framework. The infection cautioning framework likewise utilizes qualifications and passwords to approve the client. TF-IDF and cosine likeness are significantly utilized in this proposed model

The establishment of large hospitals where hundreds to thousands of patients are treated, it has created a serious problems of biomedical waste management. The seriousness of improper biomedical waste management was brought to the light during summer 1998. In India studies have been carried out at local / regional levels in various hospitals, indicate that roughly about 1-5 kg/bed/day to waste is generated. Among all health care personnel, ward boys, sweepers, operation theatre & laboratory attendants have come into contact with biomedical waste during the process of segregation, collection, transport, storage & final disposal. The knowledge of medical, paramedical staff & ward boys, sweepers about the biomedical waste management is important to improve the biomedical waste management practices. The biomedical waste requiring special attention includes those that are potentially infectious, sharps, example needle, scalpels, objects capable of puncturing the skin, also plastic, pharmaceutical & chemically hazardous substances used in laboratories etc.

Ilyas et al. [12] addressed the issue of recommending optimised treatments for obese patients and while treatments supported by big data resources aPAld us... Machine learning algorithms when deployed with large and reliable dataset can come up with better prediction of patient disease so it gives an completely new interaction for the existing problem. Published yearly count of publishing the different diseases from 2007 till... It's well documented that the deep learning algorithms are designed to classify few selected diseases using the phrases in description of the published papers, This points to a lot of multiple class' for text mining and clustering [13]. Diagnosis of Heart diseases, with not only one or two features, but attributes like age, sex, blood pressure, eGFR, serum urea, serum sodium, serum potassium [14-17]. The interaction of the environmental factors like life style of the patient which are stress based or on the other hand status of smoking also interact with the disease. This data is put in learning algorithms like SVM, RF, ANN and KNN as input to get the prediction, i.e. If th. This paper is focused on a aspect applying to the diagnosis of performance and fruitful disease with data in result from classification algorithms. And this paper does a very good job in explaining where typically

2.1. Research Gap

Despite the significant advancements in intelligent healthcare systems over the past decade, a detailed analysis of the literature from 2018 to 2026 reveals several critical gaps that limit the effectiveness, usability, and real-world applicability of existing solutions. While many studies have focused on symptom-based disease prediction using machine learning algorithms, the majority of these systems operate as standalone models without integrating complementary functionalities that are essential for practical healthcare support. One of the most prominent gaps is the lack of a unified framework that combines disease prediction, conversational interaction, and hospital recommendation into a single platform. Most existing systems either focus solely on prediction accuracy or chatbot interaction, but fail to provide end-to-end assistance to users, thereby reducing their overall utility.

Another important limitation observed in the literature is the insufficient use of real-time conversational capabilities. Although chatbot-based systems have been developed using natural language processing techniques, many of them rely on predefined responses or rule-based approaches, which restrict their ability to handle complex or dynamic user queries. This results in poor user experience and limits the system's adaptability to diverse linguistic inputs. Furthermore, only a limited number of studies have explored advanced NLP techniques such as transformer-based models or contextual understanding, which are essential for building more intelligent and human-like conversational systems.

In addition to this, the majority of existing research lacks the integration of location-aware healthcare services. While disease prediction provides users with possible medical conditions, it does not address the subsequent and equally important step of accessing appropriate healthcare facilities. Users are often left without guidance on where to seek treatment, especially in emergency situations. This gap is particularly critical in developing regions and rural areas, where access to healthcare infrastructure is already limited. The absence of geolocation-based hospital recommendation systems significantly reduces the practical impact of current healthcare applications.

Data-related challenges also represent a major research gap in this domain. Many studies utilize small, imbalanced, or synthetic datasets, which can lead to biased predictions and reduced generalizability of the models. Moreover, there is limited focus on incorporating real-time or user-generated data, such as inputs from wearable devices or continuous health monitoring systems. The integration of such dynamic data sources could significantly enhance prediction accuracy and enable proactive healthcare management, but it remains underexplored in existing research.

Another critical issue is the lack of personalization in healthcare systems. Most models provide generic predictions without considering individual-specific factors such as age, gender, medical history, lifestyle, or environmental conditions. Personalized healthcare recommendations are essential for improving diagnostic accuracy and treatment effectiveness, yet this aspect has not been adequately addressed in the majority of studies. Additionally, multilingual support is often overlooked, which limits the accessibility of these systems in diverse populations, particularly in countries with multiple regional languages.

Privacy and security concerns also constitute a significant gap in current research. Although some studies have explored privacy-preserving techniques such as federated learning, the implementation of secure and compliant systems is still in its early stages. Healthcare data is highly sensitive, and the lack of robust data protection mechanisms can hinder user trust and large-scale adoption of intelligent healthcare systems.

Furthermore, most existing systems lack continuous learning and adaptability. Machine learning models are often trained on static datasets and are not updated regularly to reflect new medical knowledge or emerging diseases. This limits their long-term effectiveness and relevance. The ability to incorporate continuous feedback and update models dynamically is essential for maintaining accuracy and reliability in real-world applications.

Finally, there is a noticeable gap in the evaluation and validation of these systems in real-world scenarios. Many studies focus on experimental results using offline datasets, without conducting user-based testing or deployment in practical environments. This raises concerns about the scalability, usability, and robustness of these systems when implemented in real healthcare settings.

To address these limitations, the proposed research aims to develop an intelligent conversational healthcare system that integrates symptom-based disease prediction, advanced NLP-based interaction, and location-aware hospital recommendation into a unified and scalable platform. By incorporating machine learning models, real-time conversational capabilities, and geolocation services, the system seeks to provide comprehensive, accurate, and user-friendly healthcare support. Additionally, the proposed approach emphasizes personalization, scalability, and practical usability, thereby contributing to the advancement of next-generation intelligent healthcare systems.

III. RESEARCH METHODOLOGY

Our project uses the chatbot to gather information. It collects the patient's information and their symptoms, then predicts the disease based on those symptoms. The disease prediction chatbot is created using ideas from natural language processing and machine learning techniques. The chatbot uses two main ideas from Natural Language Processing: tokenization and WordNet. When the user shares their symptoms, we break down the sentences to pull out the specific symptoms they mentioned. Synset is an easy-to-use tool found in NLTK (Natural Language Toolkit) that helps you find words in WordNet. Synset instances help find words that mean the same as the symptoms. Procedures for the suggested approach Several machine learning methods that are based on a multitude of datasets are used to process input data. To forecast illnesses based on user symptoms, recommend hospitals and medical facilities in the area depending on the user's location for better access to healthcare, and train and test on a large number of datasets for both known and unknown data. user input.

The designed health chatbot system represents groundbreaking integration of AI (Artificial Intelligence) with medical diagnostics to provide convenient and trustworthy preliminary healthcare support. At the heart of it, machine learning algorithm is following a Natural Language Processing (NLP) model which has been coupled with predictive analytics to provide an accurate Symptom Checker and Disease Predictor (its major feature). The chatbot design is divided into core three components; Intelligent Symptom Interpretation, Robust Disease Classification, While the system will ask if the user wants information about local hospitals. If the user agrees, the system determines the nearest healthcare facility by using geo location and distance.

Simple symptom processing module constitutes the most important part of the system. This serves an integral purpose, once the users enter in their respective symptoms in natural language, it converts these textual descriptions to structured numeric data by Term Frequency Inverse Document Frequency and vectorization with machine learning models ability for processing. This results in harmonizing the system to comprehend even if the symptom descriptions are different and put together in various ways.

3.1. Dataset

The dataset used in this study is structured in JSON (JavaScript Object Notation) format and is specifically designed to train an intelligent conversational chatbot for symptom-based disease prediction and healthcare guidance, with a primary focus on mosquito-borne diseases such as dengue and malaria. The dataset plays a crucial role in enabling the chatbot to understand user inputs, identify disease patterns, and provide appropriate medical advice in a structured and efficient manner.

The dataset consists of well-defined categories that include disease labels, associated symptom patterns, and predefined response templates. Each disease entry contains a set of commonly observed symptoms, allowing the system to map user-provided inputs to potential health conditions. For example, in the case of dengue, symptoms such as high fever, headache, joint pain, and fatigue are included, whereas malaria-related entries contain symptoms like chills, sweating, nausea, and recurring fever. These structured symptom patterns enable the machine learning model to perform accurate classification and prediction tasks. In addition to disease identification, the dataset incorporates predefined advisory responses that guide users on preventive measures and early

management strategies. These responses include recommendations such as maintaining hydration, monitoring platelet count (in the case of dengue), seeking medical consultation, and adopting preventive practices like using mosquito nets, repellents, and maintaining hygiene. By integrating such domain-specific knowledge, the chatbot is able to provide not only predictions but also actionable healthcare advice, thereby improving user engagement and practical utility.

A key feature of the dataset is the inclusion of location-based healthcare information, which enhances the system's real-world applicability. The dataset contains details of medical centers, including their names, geographical locations (latitude and longitude), and addresses. This information is utilized by the system to recommend nearby healthcare facilities based on the user's current location. The user's geolocation is determined using IP-based geolocation techniques, which estimate the user's position in terms of coordinates. To identify the nearest medical facilities, the system applies the Haversine formula, a widely used mathematical approach for calculating the distance between two points on the Earth's surface based on their latitude and longitude. By computing the distance between the user's location and each medical center in the dataset, the system generates a list of nearby hospitals or clinics. The results are then sorted in ascending order of distance, ensuring that users are presented with the closest and most accessible healthcare facilities first. This feature significantly improves the usability of the system by bridging the gap between disease prediction and treatment access.

3.2 Flowchart

The design model outlines tools that assist users in identifying health issues and accessing medical information. The user begins by typing in a description of their symptoms. Then, the algorithm looks at this description and pulls out important information. After confirming that the symptoms are real, the information is sent to an AI/ML model for more analysis. The model looks at potential diseases, gives a basic overview, and recommends ways to prevent them. The system will prompt the user to see if they want details about nearby hospitals. If the user agrees, we can find the closest healthcare facility by using their location and measuring the distance. If the user does not ask for more help or if the symptoms are not accurate, the process will end. This design helps make healthcare services work better and faster. The designed health chatbot system represents groundbreaking integration of AI (Artificial Intelligence) with medical diagnostics to provide convenient and trustworthy preliminary healthcare support. At the heart of it, machine learning algorithm is following a Natural Language Processing (NLP) model which has been coupled with predictive analytics to provide an accurate Symptom Checker and Disease Predictor (its major feature). The chatbot design is divided into core three components; Intelligent Symptom Interpretation, Robust Disease Classification, While the system will ask if the user wants information about local hospitals. If the user agrees, the system determines the nearest healthcare facility by using geo location and distance. Simple symptom processing module constitutes the most important part of the system. This serves an integral purpose, once the users enter in their respective symptoms in natural language, it converts these textual descriptions to structured numeric data by Term Frequency Inverse Document Frequency and vectorization with machine learning models ability for processing. This results in harmonizing the system to comprehend even if the symptom descriptions are different and put together in various ways.

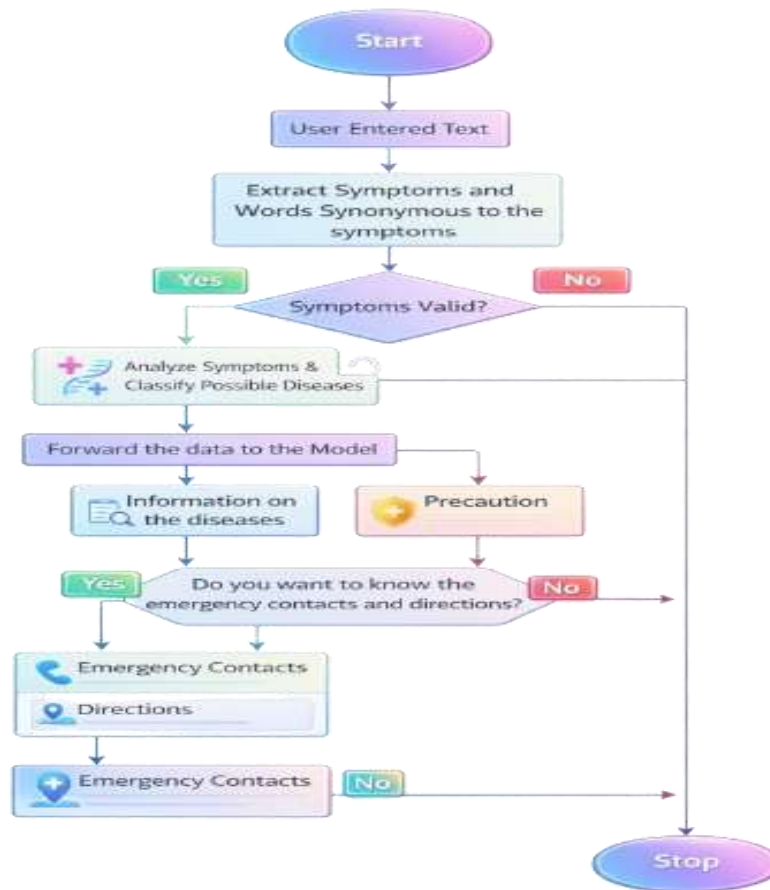


Fig.1 Flowchart of chatbot working

1. Grammar Errors: It is impossible to identify grammatical errors.
2. Predefined or Closed-domain: prior research indicates that the majority of chatbots only respond to enquiries that are either defined in the database or come from a closed domain.
3. Ambiguity: When a sentence's context or meaning is unclear or serves no suitable function.
4. Language Structure: Each language has a different sentence structure. Every language, for instance, has its own set of conventions for punctuation, text organisation, and space usage. However, current chatbots are unable to discern it.
5. Words or phrases in a human natural language format are referred to as semantics. Current chatbots are unable to interpret real language, whether they are analysing queries or just displaying an answer.
6. Data processing: There is no relational database and the current chatbots do not directly handle the structured data. In addition, it is difficult to create datasets, and mapping utterances and entities is essential.
7. User Interface: Current chatbots' user interfaces are poor; they are not intuitive, and their documentation is appalling.

IV. PROPOSED SYSTEM

Collaborative sifting approach will experience the ill effects of 3 issues: cold beginning, trust and security. The disadvantages referenced are intense concerns and significant freedoms can be missed along these lines. The fundamental center is towards critical thinking and building an admonition framework with the assistance of Python since it is simple and slick to carry out calculations on different working frameworks. Python accompanies numerous capacities, modules and libraries that assist us to perform capacities and activities with changes in code. The objective is to utilize such Python libraries and apply Machine Learning Techniques. The dataset that has been made to make the framework work is the most critical piece of the undertaking. The dataset is as a .csv record and stores sicknesses and highlights identified with it. Our dataset contains very nearly 50 columns or preparing instances of model working. These models are as id, depiction design. Where id is from 1 to 50 addressing extraordinary infections and before them each is a hyphen isolated depiction of every specific sickness. It incorporates causes, indications, way of life changes, natural components, explicit variables that lead to such infections and some more. This data.csv document has been connected with our code warning.py and for the present sudden spikes in demand for the neighborhood framework. It's anything but a menu-driven program and is valuable as an easy to use view to info and see the ideal outcomes. The User is approached to enter 1 for use and 0 for exit. On entering 1 client is approached to enter the name of the sickness and requested to enter the quantity of infections he/she needs to get cautioned with. Then, at that point computations are done at the backend. The framework cautions the client with quite a few illnesses inputted

by the client and furthermore scores (TF*IDF) scores are additionally appended aside for reference. In the backend, Python is utilized. Different libraries like Pandas, NumPy, sci-unit learn are utilized. In view of the information given by the client and every one of the models in the dataset, cosine closeness is performed. Cosine comparability is likeness between two vectors. This can be applied to the information accessible in the dataset to track down the closeness between one another by the methods for certain watchwords. The nearness between two vectors (X and Y) can be taken by doing the spot item and partitioning the outcome by the greatness. To get the information in the necessary configuration for cosine closeness utilization of libraries like Pandas and NumPy is there. Pandas is a proficient and quick approach to control information. It gives different functionalities and here it is utilized to peruse the CSV document that is our dataset and some other controls. Tf-Idf vectors are produced using the dataset utilizing different capacities and linear_kernel is additionally utilized for fitting the information. There is an enormous dataset to be managed containing whole data as it is the foundation of the framework. The dataset is as English like a characteristic and straightforward language. The Method is applied to remove includes properly from the dataset that is applicable to expectation. Calculations would then be able to perform productively on giving exact outcomes. Under regulated learning, it's anything but a significant preprocessing venture for organized dataset. Results are put away as key, esteem pair as a word reference that is acquired from cosine similitude network where key is specific illness and worth contains tuples of sickness like key.

The designed health chatbot system represents groundbreaking integration of AI (Artificial Intelligence) with medical diagnostics to provide convenient and trustworthy preliminary healthcare support. At the heart of it, machine learning algorithm is following a Natural Language Processing (NLP) model which has been coupled with predictive analytics to provide an accurate Symptom Checker and Disease Predictor (its major feature). The chatbot design is divided into core three components; Intelligent Symptom Interpretation, Robust Disease Classification, While the system will ask if the user wants information about local hospitals. If the user agrees, the system determines the nearest healthcare facility by using geo location and distance.

Simple symptom processing module constitutes the most important part of the system. This serves an integral purpose, once the users enter in their respective symptoms in natural language, it converts these textual descriptions to structured numeric data by Term Frequency Inverse Document Frequency and vectorization with machine learning models ability for processing. This results in harmonizing the system to comprehend even if the symptom descriptions are different and put together in various ways. On the users' end, a very intuitive web interface has been developed constructed on Flask, HTML, CSS, so that the symptom description can take place easily for the users across the technical spectrum, and response rendered in the manner that is easily understood and solution-oriented Medical advice. Additionally the system has been fortified with input validation for wide-ranging data quality and offers graphical representation for symptom cluster when required to enhance the user understanding.

V. LIMITATIONS OF THE PROPOSED SYSTEM

1. As a result of above mentioned reasons, the proposed AI-driven conversational healthcare assistant may come handy for providing primary health guidance. The model is assuming the users are reporting correct symptoms which may not be always the case and lead to incorrect prediction. It also does not try to validate the information given by the users.
2. Another big reason why the system cannot predict all diseases is that the training dataset did not have all diseases and it might also miss some easy but rare conditions. Second, it does not consider clinical test results, medical history or physical examination data to replace professional diagnosis entirely.
3. There will also be challenges with language ambiguity, spelling mistakes or combining different languages with which the chatbot may mixed up the symptoms. Moreover, the success of location-based hospital recommendation depends on how much location info they are getting and hospital data which may not be complete or accurate in locations.
4. The other limitation comes for its not having real-time monitoring or emergency response. In case of an emergency, the user should visit a medical facility directly instead of relying on the chatbot which provides primary healthcare support only.

VI. RESULTS AND DISCUSSION

6.1 Home Page



Fig.2 Home page

The Medical Symptom Checker is a modern online tool that looks at the symptoms people report to give quick and reliable health information. It features a simple chatbot that allows users to share their health problems and get accurate suggestions about possible conditions based on what they say. It is available all day, every day. This tool is easy to use, making healthcare more available by giving trustworthy health advice and helping users find the right care options.

6.2. Symptoms checker page

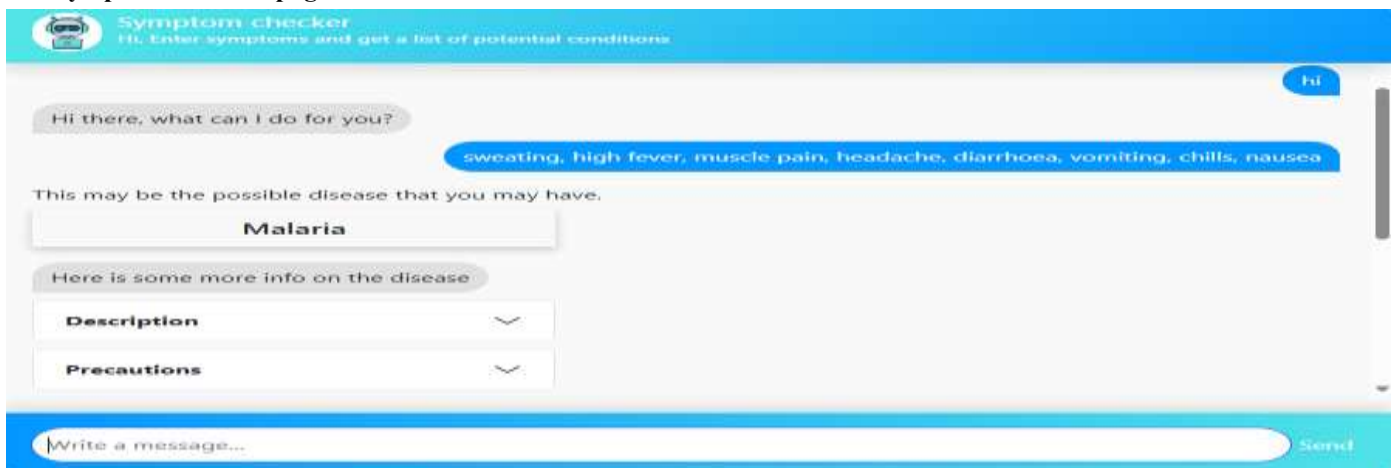


Fig.3 Predication page

This JSON dataset is created to help a chatbot learn how to identify and respond to symptoms of Malaria. It has labels to sort out diseases, common symptoms, and set responses that give advice on Malaria. This includes tips like drinking plenty of water, getting medical help, and using protection like mosquito nets and repellents. The chatbot uses this information to give helpful advice on how to find, handle, and stop diseases effectively. The chatbot based symptom analysis output generated by our system is shown in Figure 3. Symptoms like fever, headache, nausea and vomiting are entered by the user through the conversational interface. By touting those symptoms, machine learning model generates an output predicting the most probable disease. For example, in this case, it may suggest Malaria as a possible disease and give description and precautionary measures of the same. This illustration proves how well our system can understand user symptoms, and provide him with useful healthcare advice in a coherent way.

6.3 Location-Aware Medical Center Recommendation

The location-aware hospital recommendation feature of the system. After disease prediction, the chatbot asks the user whether they want to view nearby medical centers. Upon confirmation, the system retrieves the user's location and displays a list of nearby hospitals along with their distance. This feature assists users in identifying appropriate healthcare facilities quickly, especially in emergency or unfamiliar situations.

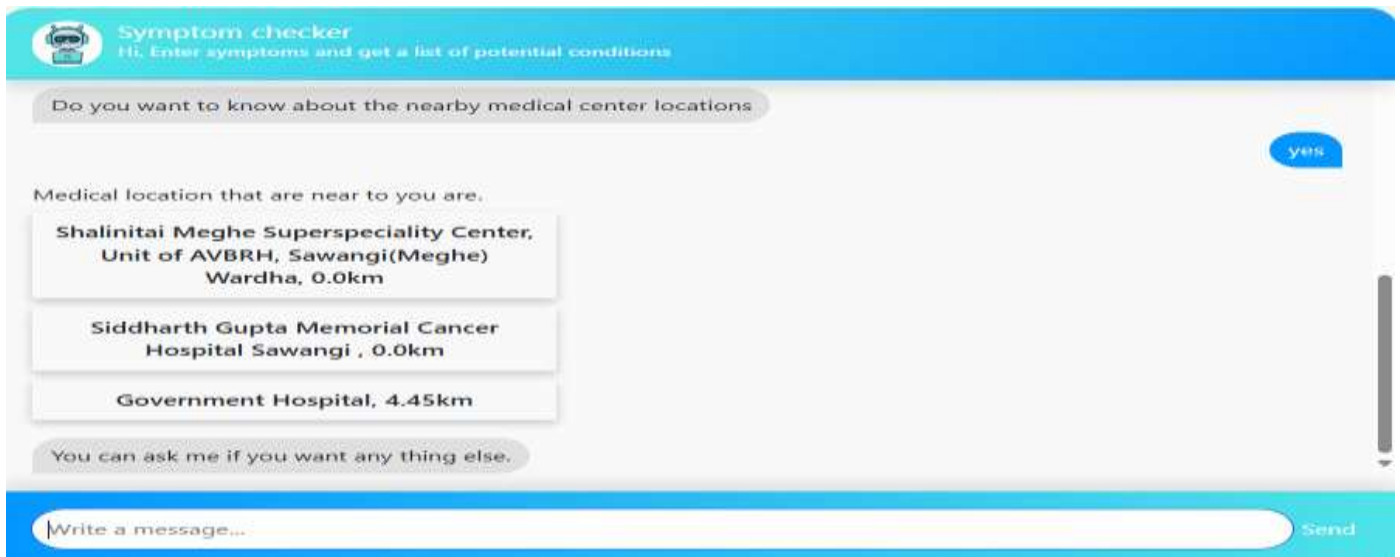


Figure 7: Location-based nearby hospital recommendation

VII. CONCLUSION

In this paper, an intelligent disease prediction chatbot system has been developed to provide real-time interaction and healthcare support to users based on their symptoms. The proposed system effectively combines machine learning, natural language processing, and geolocation services to deliver accurate, relevant, and user-friendly responses. By analyzing both syntactic and semantic similarity in user inputs, the chatbot is capable of understanding symptoms more effectively and providing appropriate guidance. The integration of advanced deep learning techniques, particularly Long Short-Term Memory (LSTM) networks and TensorFlow-based models, enhances the system's ability to process and interpret natural language inputs. Through efficient preprocessing steps such as tokenization, normalization, and word transformation, the system ensures that user input is standardized, leading to improved prediction accuracy. The chatbot not only identifies diseases based on symptom patterns but also understands user intent and responds with meaningful and easy-to-understand advice. A significant contribution of this work is the inclusion of a geolocation-based recommendation system. By utilizing user location data and analyzing information from the *medical_centers.json* dataset using the Pandas library, the system identifies and suggests nearby medical facilities. The application of the Haversine formula ensures accurate distance calculation, and the results are sorted to display the closest healthcare centers. This feature bridges the gap between disease prediction and access to medical services, making the system more practical and impactful in real-world scenarios. Furthermore, the chatbot provides preventive guidance and healthcare recommendations, enabling users to take early action and make informed decisions about their health. The use of AI-driven techniques allows the system to deliver immediate responses, reducing dependency on traditional healthcare processes for initial diagnosis.

VIII. FUTURE SCOPE

The proposed intelligent conversational healthcare system can be further enhanced in multiple ways to improve its accuracy, usability, and real-world impact. One important future direction is the expansion of the dataset by incorporating larger, diverse, and real-time medical data from hospitals, public health records, and global healthcare sources, which can improve prediction performance and enable the system to cover a wider range of diseases. The integration of multilingual and voice-based interaction capabilities using advanced natural language processing techniques can make the system more accessible to users from different linguistic backgrounds, including support for regional languages and speech-based inputs for better user convenience. Additionally, incorporating personalized healthcare features by utilizing user-specific data such as medical history, lifestyle patterns, age, and data from wearable devices like heart rate and activity trackers can significantly improve the accuracy and relevance of health recommendations. The system can also be extended with real-time health monitoring through IoT devices, enabling continuous tracking of user health and generating early warnings or emergency alerts in critical conditions. Another promising enhancement is the development of a smart hospital recommendation system that not only suggests nearby medical centers but also provides additional information such as doctor availability, hospital ratings, waiting time, and online appointment booking options. Furthermore, implementing privacy-preserving techniques such as data encryption and federated learning can ensure secure handling of sensitive health data and increase user trust. The use of advanced AI models, including transformer-based architectures and large language models, can further improve the chatbot's ability to understand complex queries and provide more accurate and human-like responses. Finally, deploying the system on cloud and mobile platforms will ensure scalability, reliability, and easy accessibility, allowing users to access healthcare support anytime and anywhere, thereby transforming the system into a comprehensive, intelligent, and personalized digital healthcare assistant for future applications.

REFERENCES

- [1] Afrash, M. R., Erfannia, L., Amraei, M., Mehrabi, N., Jelvay, S., Nopour, R., & Shanbehzadeh, M. (2022). Machine learning-based clinical decision support system for automatic diagnosis of COVID-19 based on clinical data.
- [2] Villavicencio, C. N., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G. (2022). Development of a machine learning based web application for early diagnosis of COVID-19 based on symptoms. *Diagnostics*, 12(4), 821.
- [3] Varshney, D., Zafar, A., Behera, N. K., & Ekbal, A. (2023). Knowledge graph assisted end-to-end medical dialog generation. *Artificial Intelligence in Medicine*, 139, 102535.
- [4] Irawan, P. L. T., Hartanto, C. B. S., & Kelana, O. H. (2022). Medicine Consumption Reminder and Monitoring Application for Patients with Leprosy Disease. *Journal of Community Practice and Social Welfare*, 2(1), 2022.
- [5] Haugeland, I. K. F., Følstad, A., Taylor, C., & Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. *International Journal of Human-Computer Studies*, 161, 102788.
- [6] Athulya N, Jeeshna K, S J Aadithyan, U Sreelakshmi, Hairunizha Alias Nisha Rose, "Healthcare Chatbot", © 2021 IJCRT | Volume 9, Issue 10 October 2021 | ISSN: 2320-2882.
- [7] Dechert, M., Schöning, J., & Barkowsky, T. (2019). *Implementation and Evaluation of a Chatbot to Crowdsourcing Geotagged Images to Detect Mosquito Breeding Sites* (Doctoral dissertation, University of Bremen).
- [8] Tanmay, T., Bhardwaj, A., & Sharma, S. (2020, February). E-Health Bot to change the Face of Medicare. In *2020 Research, Innovation, Knowledge Management and Technology Application for Business Sustainability (INBUSH)* (pp. 49-54). IEEE.
- [9] Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S. (2019, April). Chatbot for disease prediction and treatment recommendation using machine learning. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 851-856). IEEE.
- [10] Giansanti, D. (2023). The chatbots are invading us: A map point on the evolution, applications, opportunities, and emerging problems in the health domain. *Life*, 13(5), 1130.
- [11] Disha, D. N., Rao, S., Bhat, M., Pai, K., & Pandit, A. (2024, July). Revolutionizing Healthcare Triage: A Comparative Analysis of Machine Learning-Driven Symptom Checkers and Triage Bots for Common Diseases and Skin Conditions. In *2024 Second International Conference on Advances in Information Technology (ICAIT)* (Vol. 1, pp. 1-6). IEEE.
- [12] Chakraborty, S., & Dey, L. (2022). The Implementation of AI and AI-Empowered Imaging System to Fight Against COVID-19—A Review. *Smart Healthcare System Design: Security and Privacy Aspects*, 301-311.
- [13] R. Burle, S. Gaurkhede, M. Sharma, U. Pacharaney and P. Verma, "Utilizing Conceptual Deep Learning Networks for Diabetic Retinopathy Based on Feature Extraction and Classification," *2024 (PICET)*, Vadodara, India, 2024, pp. 1-6, doi: 10.1109/PICET60765.2024.10716154.
- [14] Çelik Ertuğrul, D., Toygar, Ö., & Foroutan, N. (2021). A rule-based decision support system for aiding iron deficiency management. *Health Informatics Journal*, 27(4), 14604582211066054.
- [15] Singh, U., & Choubey, M. K. (2021, October). A review: image enhancement on MRI images. In *2021 5th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1-6). IEEE.
- [16] Mohseni, S., Yang, F., Pentyala, S., Du, M., Liu, Y., Lupfer, N., ... & Ragan, E. (2021, May). Machine learning explanations to prevent overtrust in fake news detection. (Vol. 15, pp. 421-431).
- [17] Grünebaum, A., Chervenak, J., Pollet, S. L., Katz, A., & Chervenak, F. A. (2023). The exciting potential for ChatGPT in obstetrics and gynecology. *American Journal of Obstetrics and Gynecology*, 228(6), 696-705.
- [18] Burle, Rushikesh, et al. "Conversational Symptom Checker Chatbot for Disease Prediction and Suggesting Nearby Hospitals using Machine Learning and Location Services." *2025 International Conference on Electronics and Renewable Systems (ICEARS)*. IEEE, 2025.
- [19] Haque, M. R., & Rubya, S. (2023). An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. *JMIR mHealth and uHealth*, 11(1), e44838.
- [20] Kumar, S. S., Ahmed, S. T., Sandeep, S., Madheswaran, M., & Basha, S. M. (2022). Unstructured Oncological Image Cluster Identification Using Improved Unsupervised Clustering Techniques. *Computers, Materials & Continua*, 72(1).
- [21] Ahmed, S. T., Sreedhar Kumar, S., Anusha, B., Bhumika, P., Gunashree, M., & Ishwarya, B. (2020). A generalized study on data mining and clustering algorithms. In *New Trends in Computational Vision and Bio-inspired Computing: Selected works presented at the ICCVBIC 2018, Coimbatore, India* (pp. 1121-1129). Cham: Springer International Publishing.
- [22] Harada, Y., Sakamoto, T., Sugimoto, S., & Shimizu, T. (2024). Longitudinal Changes in Diagnostic Accuracy of a Differential Diagnosis List Developed by an AI-Based Symptom Checker: Retrospective Observational Study. *JMIR Formative Research*, 8, e53985.
- [23] Meropi, P., Billis, A. S., Hasanagas, N. D., Bratsas, C., Antoniou, I., & Bamidis, P. D. (2017, June). Conditional entropy based retrieval model in patient-carer conversational cases. In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 576-581). IEEE.
- [24] R. Burle, S. Gaurkhede, P. Gourshettiwar, S. Izankar, S. Gundewar and U. Pacharaney, "Hybrid Machine Learning Based Predictive Model To Improve Cloud Application Performance In Cloud Saas," *(ICETEMS)*, Nagpur, India, 2024, pp. 167-172, doi: 10.1109/ICETEMS64039.2024.10965099.
- [25] Mendonça YV, Naranjo PGV, Pinto DC. The Role of Technology in the Learning Process. *Emerging Science Journal*. 2022;6(Special Issue):280-95. 10.28991/ESJ-2022-SIED-020
- [26] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].

- [27] Roshani S, Coccia M, Mosleh M. Sensor technology for opening new pathways in diagnosis and therapeutics of breast, lung, colorectal and prostate cancer. medRxiv. 2022:2022.02.18.22271186. 10.28991/HIJ-2022-03-03-010
- [28] Hamidi H, Daraee A. Analysis of pre-processing and post processing methods and using data mining to diagnose heart diseases. International Journal of Engineering, Transactions A: Basics 2016;29(7):921-30. 10.5829/idosi.ije.2016.29.07a.06
- [29] Ye EZ, Ye EH, Bouthillier M, Ye RZ. DeepImageTranslator V2: analysis of multimodal medical images using semantic segmentation maps generated through deep learning. bioRxiv. 2021:2021.10.12.464160. 10.28991/HIJ-2022-03-03-07
- [30] Manjurul Ahsan M, Siddique Z. Machine learning based disease diagnosis: A comprehensive review. arXiv e-prints. 2021:arXiv: 2112.15538.
- [31] Shaji SP, editor Prediction and diagnosis of heart disease patients using data mining technique. 2019 international conference on communication and signal processing (ICCSP); 2019: IEEE. 10.1109/ICCSP.2019.8697977
- [32] Chaudhary J, Joshi V, Khare A, Gawali R, Manna A. A Comparative Study of Medical Chatbots. International Research Journal of Engineering and Technology (IRJET). 2021;8(02).
- [33] Vasileiou MV, Maglogiannis IG. The Health ChatBots in Telemedicine: Intelligent Dialog System for Remote Support. Journal of Healthcare Engineering. 2022;2022. <https://doi.org/10.1155/2022/4876512>
- [34] Burle, Rushikesh, et al. "Conversational Symptom Checker Chatbot for Disease Prediction and Suggesting Nearby Hospitals using Machine Learning and Location Services." 2025 International Conference on Electronics and Renewable Systems (ICEARS). IEEE, 2025.
- [35] Harada, Y., Sakamoto, T., Sugimoto, S., & Shimizu, T. (2024). Longitudinal Changes in Diagnostic Accuracy of a Differential Diagnosis List Developed by an AI-Based Symptom Checker: Retrospective Observational Study. *JMIR Formative Research*, 8, e53985.
- [36] Irawan, P. L. T., Hartanto, C. B. S., & Kelana, O. H. (2022). Medicine Consumption Reminder and Monitoring Application for Patients with Leprosy Disease. *Journal of Community Practice and Social Welfare*, 2(1), 2022.
- [37] "Recommender Systems with Python — Part I: ContentBased Filtering" Accessed February 25, 2021 <https://heartbeat.fritz.ai/recommender-systems-with-python-part-i-content-based-filtering-g-5df4940bd831>
- [38] Joeran Beel, Stefanlanger, Marcel Genzmehr, Bela Gipp, Corinna Breiter, Andreas Nurnberger, "Research Paper Recommendation System: A Quantitative Literature Survey", International Journal On Digital Libraries (2015)
- [39] "TF(Term Frequency)-IDF(Inverse Document Frequency) from scratch in python ." Accessed February 26, 2021 <https://towardsdatascience.com/tf-term-frequency-idf-inverse-document-frequency-from-scratch-in-python-6c2b61b78558>. "Recommender Engine - Under The Hood" Accessed February 27, 2021 <https://www.kdnuggets.com/2018/02/recommender-engine.html>.