IJEIT
International Journal of Emerging Information Technology
ISSN: XXXX-XXXX / Volume 1 Issue 1 September - October 2025 / Page No: 37-56
Paper Id: IJEIT-V1I1P105

Original Article

Disease Prediction Using Machine Learning: Comparative Analysis of SVM, Naive Bayes, and Decision Tree Models with Gemini API Integration

Kavali Durga Prasad, Nidamanuri Hemanth Gopal, Malli Deepak, Kaligotla Veera Venkata Jitin

Independent Researcher, India.

Received: 28 June 2025 Revised: 03 September 2025 Accepted: 20 September 2025 Published: 10 October 2025

Abstract - *This project explores the use of machine learning algorithms to predict diseases based on user-input symptoms, employing three popular models: Support Vector Machine (SVM), Naive Bayes, and Decision Tree Classifier. The models are trained on this dataset, and their performance is evaluated using metrics such as accuracy, precision, recall, and F1 score. A user interface function allows individuals to input their symptoms, and the trained models predict the most likely disease with a confidence score. This comparative analysis highlights the strengths and weaknesses of each algorithm in the context of dis- ease prediction.*

1. Introduction

The accurate prediction of diseases based on symptoms is a critical challenge in the field of healthcare, as it can enable timely interventions and im- prove patient outcomes. Traditional diagnostic processes often rely heavily on medical expertise and patient history, which can be time-consuming and sometimes prone to human error. With the rapid advancements in machine learning, there is an increasing interest in utilizing data-driven approaches to enhance the accuracy and efficiency of disease diagnosis. Machine learning algorithms have shown great promise in various healthcare applications, including disease prediction, patient risk assessment, and personalized treatment recommendations. These algorithms can analyze large datasets, identify patterns, and make predictions based on the relationships between input features and target labels.

2. Literature Review

The literature review explores existing research on the application and per- formance of Decision Tree, Naive Bayes, and SVM algorithms in different domains. Here are some key references:

Paper Title: Accurate AI-Based Chatbot to Diagnose Heart Diseases Pre-Human Doctor Consultation

- **Authors:** Batool Ali Majeed, Ammar Yasir Hardan, Batool Yasir Hardan, Dunya Faeq Munaf
- **Institution:** Ministry of Education, Iraqi Gifted School, Baghdad, Iraq
- Corresponding Author Email: ba5074963@gmail.com
- **Published In:** Revue d'Intelligence Artificielle
- Volume: 38, Issue: 1, February 2024, Pages: 213-220
- Special Issue: Machine Learning and Artificial Intelligence for Context-Aware Network Optimization
- **DOI**: https://doi.org/10.18280/ria.380121

Methodology	Issue Addressed	Research Gap
This study introduces an AI-based chatbot that aims to provide initial assessments for heart disease based on user-provided symptoms. Utilizing the Support Vector Machine (SVM) algorithm, the chatbot processes the symptoms reported by users, employing Natural Language Processing (NLP) to create a smooth, interactive conversation flow. By gathering essential information, the chatbot then applies SVM to predict the likelihood of heart disease and suggests whether further medical consultation is needed.	With heart disease being a significant global health concern, early detection and intervention can be critical. This chatbot provides a user-friendly and accessible option for individuals who may not have immediate access to in-person healthcare. By delivering an easy-to-access platform, the chatbot offers a preliminary layer of health guidance, which can be particularly beneficial in underserved areas or for individuals seeking quick initial advice.	The paper identifies several areas for potential enhancement, such as increasing diagnostic accuracy by leveraging larger datasets and expanding language capabilities for a broader user base. Future research could focus on refining the chatbot's responses to improve user engagement and expanding its diagnostic scope. Additionally, incorporating continuous learning features and strengthening privacy measures would help to enhance the chatbot's functionality and reliability for users.

Paper Title: A Survey on Healthcare Systems Using Internet of Things

Authors: Chitra Ramesh and Jayalakshmi Sekar

Institutions:

1. Department of Computer Science, Immaculate College for Women, Cuddalore, India
2. Department of Computer Applications, VISTAS, Chennai, India

Corresponding Author Email: chitrashanmougam@gmail.com **Published In**: ITM Web of Conferences

Volume: 37

Year: 2021

DOI: https://doi.org/10.1051/itmconf/20213701015

Methodology	Issue Addressed	Research Gap
This study explores integrating the Internet of Things (IoT) into healthcare systems for continuous patient monitoring using wearable or implantable sensors. It focuses on collecting physiological data like blood pressure, heart rate, and body temperature, which is then transmitted to cloud systems for analysis. Machine learning algorithms are applied for disease detection and prediction, while Big Data processing and fog computing are used to manage and preprocess large data volumes at the cloud edge.	The primary issue addressed by this research is the improvement of healthcare delivery through the seamless integration of IoT devices. Specifically, it aims to tackle the challenges in healthcare monitoring and management, including real-time data collection from various sensors, the secure exchange and storage of health data, and the use of analytics to detect health issues early. The study highlights the need for a unified system where data from different devices can be integrated efficiently to ensure better decision-making and patient care.	Despite the growing use of IoT in healthcare, several challenges remain unaddressed. These include issues related to data security and privacy, device interoperability, and the effective integration of data across different platforms. Furthermore, while there is significant research on the use of machine learning for health monitoring, there is a lack of comprehensive studies that evaluate the practical implementation and performance of these algorithms in real-world healthcare settings. This paper seeks to bridge this gap by exploring the application of IoT and Big Data analytics in healthcare, focusing on system performance, scalability, and security.

‘K-Bot’ Knowledge Enabled Personalized Healthcare Chatbot Authors:
Rahul Pradhan, Jaya Shukla, Mani Bansal

Year Published:
2021

Published In:
IOP Conf. Series: Materials Science and Engineering 1116 (2021) 012185

DOI: 10.1088/1757-899X/1116/1/012185

Methodology	Issue Addressed	Research Gap
The study develops a personalized healthcare chatbot named "K-Bot" that utilizes a Decision Tree algorithm for diagnosing diseases based on symptoms. Users interact with the chatbot, which analyzes symptoms and predicts possible diseases along with a confidence level. It also recommends a specialized doctor based on the predicted condition, helping users save time and money.	The main issue addressed is the high number of individuals seeking medical consultations for minor health issues, which could be managed through simple remedies. The chatbot aims to provide an accessible, accurate, and time-efficient alternative for people to self-diagnose and get recommendations without visiting a doctor.	The research highlights the gap in existing healthcare chatbots, where many do not provide personalized, accurate disease predictions and often lack integration with practical health recommendations. Previous models have limitations like inadequate symptom analysis or the inability to suggest specific healthcare professionals. The study seeks to improve the prediction accuracy and confidence levels while providing a specialized doctor recommendation based on the disease diagnosis.

Title Paper:

Mental Health Assist and Diagnosis Conversational Interface using Logistic Regression Model for Emotion and Sentiment Analysis

- **Authors:**
  S Moulya, T R Pragathi
- **Year Published:**
  2022
- **Published In:**
  Journal of Physics: Conference Series 2161 (2022) 012039

Methodology	Issue Addressed	Research Gap
The study develops an AI-powered chatbot for mental health assistance that performs real-time emotion and sentiment analysis using a Logistic Regression model. The chatbot analyzes user emotions during conversations and applies sentiment analysis to categorize emotions into "Neutral," "Positive," and "Negative." The chatbot also uses the Kessler Psychological Distress Scale (K10) to assess the user's mental health and provide relevant feedback. The system continuously interacts with users, providing emotional support and recommending mental and physical exercises as part of a self-healing kit.	This work addresses the challenge of providing accessible, real-time mental health support. It offers an alternative to traditional psychiatric visits, which are limited by time and accessibility. The chatbot can continuously monitor mental health, provide emotional support, and recommend exercises for improvement, all without requiring an appointment or fixed schedule.	The research highlights a gap in mental health care where traditional methods of therapy are often not readily accessible, especially for those hesitant to seek help. The study also addresses the gap in chatbot-based mental health solutions by developing an AI model that can perform detailed sentiment analysis and offer personalized, real-time mental health support. It aims to improve the effectiveness of AI chatbots in mental health care by enhancing their diagnostic capabilities and emotional intelligence.

Title Paper:

Impact of Artificial Intelligence in the field of Health Care

- **Authors:**
  Ragavi V, SanthaSheela A C, GeethaiNarayanani Kannaiyan
- **Year Published:**
  2021
- **Published In:**
  Journal of Physics: Conference Series 1831 (2021) 012006

Methodology	Issue Addressed	Research Gap
The paper discusses the various applications of Artificial Intelligence (AI) in healthcare, focusing on AI algorithms powered by machine learning (ML) and deep learning (DL) technologies. The methodology section outlines how AI algorithms work by processing both structured and unstructured data. Unstructured data, such as medical records, is first converted into structured data through labeling and annotation. Once the data is structured, AI models are trained using large datasets to make accurate predictions or diagnoses. The paper explores applications like virtual medical assistants, automated image diagnosis, and AI-powered chatbots, which help in oncology, cardiology, and radiology.	This work addresses the growing need for efficient and quick decision-making in healthcare by leveraging AI. The increasing volume of healthcare data, especially in the form of Electronic Medical Records (EMR), can overwhelm human clinicians. AI provides a solution by processing this vast amount of data faster and more accurately than traditional methods. AI also assists in various aspects of healthcare, including diagnosis, surgery assistance, and virtual health companions, improving patient outcomes and supporting healthcare professionals.	The research identifies a gap in how healthcare systems manage the growing volume of data and the inefficiencies in traditional healthcare practices. The paper highlights that AI has the potential to optimize clinical practices, making them more efficient and accessible. However, there is also a recognition of challenges in the implementation of AI, including the need for structured data and overcoming resistance to adopting AI technologies in healthcare settings. The paper calls for more exploration into how AI can complement existing healthcare systems and further enhance diagnostic and therapeutic capabilities.

Title Paper:
Smart Medical Chatbot with Integrated Contactless Vital Sign Monitor

- **Authors:**
  Wan Muhamad Asyraf Wan Zaki, Muhammad Faiz Md Shakhih, Muhammad Hanif Ramlee, and Asnida Ab Wahab

- **Year Published:**
  2019

- **Published In:**
  Journal of Physics: Conference Series 1372 (2019) 012025

Methodology	Issue Addressed	Research Gap
This study presents a smart medical chatbot system integrated with a contactless vital sign monitoring device. The chatbot uses artificial intelligence (AI) and natural language processing (NLP) to enable personalized health consultations. The system integrates a vision-based real-time monitoring device that measures vital signs like oxygen level, heart rate, and respiration rate in a non-obtrusive manner. The diagnostic system leverages an automated database recall system for disease identification, using the "maximum likelihood method" to analyze symptoms and assign probabilities to potential diagnoses. Results indicate low error margins for pulse and ECG sensors, demonstrating accuracy in vital sign monitoring without physical contact.	The paper addresses the limited interaction and lack of seamless communication between healthcare providers and patients in traditional healthcare settings. It highlights the need for improved interoperability to provide patients with more accessible and proactive healthcare services. By using a chatbot with contactless monitoring, the system bridges the gap, enabling efficient remote monitoring and personalized care for patients who may be bedridden or at home.	The research highlights a gap in the availability of accessible, real-time, and non-intrusive health monitoring for patients who require continuous care but may not have direct access to healthcare facilities. While contact-based monitoring solutions exist, the need for a non-contact, AI-driven system that provides accurate real-time data is underscored. Additionally, there is limited integration of AI-based chatbots that can assess and predict health conditions using contactless monitoring, making this system innovative in the field. The study suggests further exploration in enhancing AI diagnostic capabilities and expanding the scope of such systems for broader healthcare applications.

Title Paper:
The Potential Application of Artificial Intelligence in Healthcare and Hospitals

- **Authors:**
  Sunanda Rani, Dong Jining, Dhaneshwar Shah, Siyanda Xaba, and Prabhat Ranjan Singh

- **Year Published:**
  2023

- **Published In:**
  ITM Web of Conferences 53, 01005 (2023) ICDSIA-2023

Methodology	Issue Addressed	Research Gap
This research explores both existing and potential applications of Artificial Intelligence (AI) in the healthcare and hospital sectors. Using a combination of primary and secondary data sources, the study examines the current role of AI in healthcare through published research papers, industry reports, and expert opinions from healthcare professionals, including physicians, hospital administrators, and IT experts. The research aims to identify the existing applications, potential uses, and challenges of integrating AI into healthcare settings. Key AI applications discussed include medical records management, diagnosis and treatment support, patient monitoring, drug discovery, image recognition for diagnoses, automated prescription writing, and AI-assisted surgeries.	The study addresses the need to improve healthcare service quality and efficiency by incorporating AI, given the increasing patient demand and resource constraints in healthcare. It highlights specific challenges, such as regulatory issues, privacy concerns, and data accessibility, which hinder the full implementation of AI in hospitals. Additionally, it emphasizes the need for reliable data-sharing mechanisms and ethical considerations to ensure AI's safe and effective integration in healthcare.	While AI has transformative potential in healthcare, there is limited guidance on addressing the technical and ethical challenges associated with its adoption in hospitals. The study identifies a gap in understanding how healthcare organizations can overcome regulatory and data privacy barriers to implement AI effectively. Further research is suggested to explore methods for enhancing data security, improving AI transparency, and minimizing biases in AI algorithms to make healthcare solutions more reliable and accessible.

Title of Paper:
Towards a Chatbot for Medical Diagnosis Based on Patient Symptoms

- **Authors:**
  Kisito Kabore, Omar Sawadogo, Yaya Traore, Julie Thiombiano
- **Year Published:**
  2024
- **Published In:**
  Digital Health and Informatics Innovations for Sustainable Health Care Systems

Methodology	Issue Addressed	Research Gap
The study developed a medical diagnosis chatbot to assist healthcare providers by predicting diseases based on patient symptoms. Using data from consultation records, the team trained models with algorithms like Random Forest and Logistic Regression, then integrated a Llama-inspired language model to generate reports and prescriptions. Built with Python, FastAPI, and PostgreSQL, the chatbot offers an efficient diagnostic tool for resource-limited healthcare settings.	The study addresses the lack of adequate healthcare infrastructure and shortage of medical professionals in Burkina Faso, particularly in rural areas. By developing an AI-based chatbot, the research aims to provide preliminary diagnostic support, making healthcare more accessible and assisting physicians with initial assessments based on patient symptoms.	There is a significant gap in the availability of accessible AI-driven medical diagnostic tools tailored for regions with limited healthcare resources. While previous studies have explored AI in medical diagnostics, few have focused on chatbots for primary diagnosis based on symptoms, especially in low-resource settings like Burkina Faso. This research contributes by creating a model that can provide initial diagnostic support, potentially improving healthcare delivery where medical resources

Title: Artificial Intelligence (AI) Chatbots in Medicine: A Supplement, Not a Substitute
Authors: Ibraheem Altamimi, Abdullah Altamimi, Abdullah S. Alhumimidi, Abdulaziz Altamimi, Mohamad Hani Temsah
Year Published: 2023
Published In: Cureus
DOI: 10.7759/cureus.40922

Methodology	Issue Addressed	Research Gap
The article primarily uses an editorial format to discuss various aspects of AI chatbot implementation in healthcare, evaluating the roles, benefits, and limitations of AI tools. It emphasizes a narrative methodology by contrasting human medical professionals' irreplaceable skills (such as empathy, experience, and critical thinking) with AI's limitations, while supporting its claims through statistics and references to existing research.	The article addresses the role of AI chatbots in healthcare, emphasizing that while they are valuable tools for enhancing efficiency in routine medical tasks and patient education, they cannot replace medical professionals. It discusses the implications of integrating AI into healthcare, highlighting both the potential benefits and limitations, particularly in terms of the empathy and expertise required in complex patient care.	The article identifies a gap in the expectation that AI chatbots might fully replace healthcare professionals. It stresses the need for a collaborative approach where AI augments human capabilities but does not replace essential human elements, such as empathy and clinical judgment, that are vital in healthcare. This positions AI chatbots as supplementary tools rather than as replacements, encouraging future research to focus on ethical frameworks, regulatory needs, and the role of human expertise in AI-assisted healthcare.

Title: An Artificial Intelligence-Based Chatbot for Prostate Cancer Education: Design and Patient Evaluation Study
Authors: Magdalena Görtz, Kilian Baumgärtner, Tamara Schmid, Marc Muschko, Philipp Woessner, Axel Gerlach, Michael Byczkowski, Holger Sültmann, Stefan Duensing, Markus Hohenfellner
Year Published: 2023
Published In: *[Journal Name Not Provided in Text, Specify if Known or Use Placeholder]*
DOI: Not provided in the text. You may need to search for this to obtain the DOI if necessary.

Methodology	Issue Addressed	Research Gap
The study employed a user-centered evaluation approach, developing and assessing the PROSCA chatbot, an AI-driven communication assistant designed to provide patient education on prostate cancer. Using the SAP Conversational AI (CAI) platform, the chatbot was tailored with specific natural language processing (NLP) models to interpret prostate cancer-related questions and deliver structured responses. This included functions like intent recognition, entity recognition, and sentiment analysis. Ten men suspected of having prostate cancer participated, and nine actively used the chatbot, providing	The study addresses the gap in accessible, patient-centered information on prostate cancer diagnosis, stages, treatment options, and related procedures. Many patients lack easily understandable resources that can assist in informed decision-making. By offering round-the-clock information and addressing common patient questions, the chatbot serves as a supplementary tool to enhance patient knowledge and provide a foundation for doctor-patient communication.	The research highlights the need for more studies evaluating the efficacy of chatbots in healthcare settings, particularly in oncology. While previous work has explored chatbots for mental health and postoperative care, few studies have rigorously assessed chatbots in cancer education. This study suggests the potential for AI-based tools like PROSCA to improve health outcomes and relieve the strain on healthcare professionals by handling repetitive informational queries, allowing medical staff to focus on individualized care. Further research should explore the chatbot's long-term impact on patient engagement and health outcomes in larger, more diverse populations.

Paper: "Building a Medical Chatbot Using Support Vector Machine Learning Algorithm"

- **Authors**: Tamizharasi B., Jenila Livingston L.M., S. Rajkumar
- **Year Published**: 2020
- **Published In**: *Journal of Physics: Conference Series* (Proceedings of the National Science, Engineering, and Technology Conference 2020)
- **DOI**: 10.1088/1742-6596/1716/1/012059

Methodology	Issue Addressed	Research Gap
The authors implement a Support Vector Machine (SVM) classifier to develop a chatbot capable of diagnosing diseases based on user-reported symptoms. The chatbot collects basic details from users, uses NLP for conversational flow, and applies the SVM algorithm to make predictions and recommend treatment options.	This work addresses the need for accessible preliminary medical assessments, offering an alternative to traditional in-person consultations by providing a cost-effective and accessible platform for health guidance.	The paper notes opportunities for improvement in expanding diagnostic accuracy through larger datasets, enhancing natural language responses, and integrating multilingual and more complex diagnostic capabilities. Further research could also focus on incorporating feedback loops and advanced privacy protocols.

Paper: "BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding"

- **Authors**: Arun Babu, Sekhar Babu Boddu
- **Year Published**: 2024
- **Published In**: *Exploratory Research in Clinical and Social Pharmacy*
- **DOI**: 10.1016/j.rcsop.2024.100419

Methodology	Issue Addressed	Research Gap
This study introduces a BERT-based medical chatbot that leverages advanced deep learning and natural language understanding to improve healthcare communication. The model is trained on medical datasets such as MIMIC-III, BioASQ, and PubMed, and includes modules for data processing, intent recognition, and contextual management to provide accurate responses to complex medical queries.	Traditional medical chatbots often struggle with understanding medical jargon and maintaining context in conversations. This BERT-based model addresses these challenges by enabling precise interpretation of medical terminology, maintaining context in multi-turn interactions, and providing personalized responses for effective patient engagement	Future enhancements could focus on broadening the chatbot's adaptability to diverse medical contexts, incorporating multilingual support, and addressing computational efficiency for broader deployment. Additional work may also address continuous learning, improved data privacy protocols, and the inclusion of underrepresented medical cases for a more robust chatbot model.

Paper Title: *MAAC-MEDICARE AT A CLICK: A Medical Chatbot Using Machine Learning Algorithm*
Authors: Rani Astya, Puja Kumari, Apoorva Goswami, Abinash Senapati
Published Year: 2021
Published In: *International Research Journal of Modernization in Engineering, Technology, and Science*

Methodology	Issue Addressed	Research Gap
The paper presents a medical chatbot that uses supervised machine learning algorithms—primarily Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes—to predict diseases based on symptom input. The system is trained with a dataset of diseases and symptoms, enabling it to classify user inputs accurately for disease prediction.	This chatbot system addresses the need for remote medical consultation, reducing the necessity for in-person doctor visits, which is especially relevant in pandemic contexts. It provides a user-friendly, conversational interface for health advice, aiming to bridge gaps in healthcare access and convenience	While effective in its current scope, the chatbot could be improved by incorporating real-time data from wearable health monitoring devices to enhance prediction accuracy. Additionally, expanding the dataset to include more diverse medical conditions would increase the application's scope and robustness, providing more comprehensive healthcare support

Paper: "Revolutionizing Diabetes Diagnosis: Machine Learning Techniques Unleashed"

- **Authors**: Zain Shaukat, Wisal Zafar, Waqas Ahmad, Ihtisham Ul Haq, Ghassan Husnain, Mosleh Hmoud Al-Adhaileh, Yazeed Yasin Ghadi, Abdulmohsen Algarni
- **Year Published**: 2023
- **Published In**: *Healthcare*
- **DOI**: 10.3390/healthcare11212864

Methodology	Issue Addressed	Research Gap
The study used six machine learning algorithms (Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, K-Nearest Neighbors, and Naïve Bayes) applied to the PIMA diabetes dataset. The algorithms were evaluated using WEKA and Python, focusing on accuracy, precision, and F-measure.	The challenge of accurately predicting diabetes for improved early diagnosis and patient care.	Future research should aim at increasing model accuracy, using larger datasets, and developing web-based and mobile applications for practical use.

Paper: "Personalized Healthcare: A Comprehensive Approach for Symptom Diagnosis and Hospital Recommendations Using AI and Location Services"

- **Authors**: Seng-Keong Tan, Siew-Chin Chong, Kuok-Kwee Wee, Lee-Ying Chong
- **Year Published**: 2024
- **Published In**: *Journal of Informatics and Web Engineering*
- **DOI**: 10.33093/jiwe.2024.3.1.8

Methodology	Issue Addressed	Research Gap
The authors developed a Flask-based platform that integrates a Random Forest model for symptom checking, an OpenAI API chatbot for diagnosis, and Google Maps API for hospital recommendations. It also includes user profile management and a medicine reminder feature.	The need for a comprehensive, integrated solution for personalized health management and symptom- based diagnostics	Future improvements include adding more languages, supporting more complex medical guidance, and incorporating a larger dataset

Paper: "Chatbot for Healthcare and Oncology Applications"

- **Authors**: (Authors not specified; inferred from content)
- **Year Published**: (Year not specified)
- **Published In**: (Source assumed to be related to healthcare technology or conference proceedings)
- **DOI**: (DOI not provided)

Methodology	Issue Addressed	Research Gap
Literature review of current chatbot implementations in oncology, assessing their capabilities for patient engagement and workflow Efficiency	Evaluating the role of chatbots in aiding cancer diagnosis and treatment, focusing on patient support and operational workflow	The need for further research on chatbot accuracy, user satisfaction, and their adaptation to different healthcare scenarios

Healthcare Chatbot System

- **Authors**: Mark Lawrence, MD Istiyak, Mohd Aman
- **Year Published**: 2024
- **Journal**: International Journal of Novel Research and Development (IJNRD), Volume 9, Issue 3, March 2024

Methodology	Issue Addressed	Research Gap
The chatbot uses machine learning and NLP techniques with Python libraries (Pandas, NumPy, Sklearn, and Gensim). The system allows user interaction for symptom input, disease prediction, and recommendations using models like Decision Trees and SVM. Data preprocessing includes entity recognition and contextual understanding to interpret user input accurately..	Accessibility: Designed to provide medical advice in remote or underserved areas. Privacy: Ethical concerns around user data privacy are addressed. NLP Challenges: Interpreting complex medical queries and providing reliable responses.	Lack of Dynamic Adaptability: Existing chatbots rely on static responses; improvements could allow real-time adaptation in responses based on evolving user needs. Enhanced Contextual Understanding: Limited capacity to handle unstructured data; future work could incorporate deeper contextual analysis and response refinement to improve user engagement

A Medical Chatbot

- **Authors**: Mrs. Rashmi Dharwadkar, Dr. Neeta A. Deshpande
- **Year Published**: 2018
- **Journal**: International Journal of Computer Trends and Technology (IJCTT), Volume 60, Issue 1, June 2018

Methodology	Issue Addressed	Research Gap
Utilizes NLP with algorithms like SVM for disease prediction and Porter Stemming for language processing. User interactions occur via voice-text conversions enabled by Google API. The system includes features for disease prediction based on symptoms and provides a database-driven Q&A format	Limited Access: Aims to eliminate the need for in-person hospital visits for minor concerns. Real-Time Response: Designed to give immediate feedback on health-related queries, enhancing user satisfaction and accessibility.	Scalability: Current chatbot models lack personalization and cannot handle complex, context-sensitive health issues, which future work could address by incorporating more sophisticated AI models. Integration with Health Systems: Further research is suggested on integrating with electronic health records for more personalized recommendations

Title: The Development of A Medical Chatbot Using The SVM Algorithm

Authors: Ryan Matthew, David Agustriawan, Mario Donald Bani, Muammar Sadrawi, Nanda Rizqia Pradana Ratnasari, Moch Firmansyah, Arli Aditya Parikesit

Year Published: 2022
Published In: 4th International Conference on Cybernetics and Intelligent Systems (ICORIS)

DOI: 10.1109/ICORIS56080.2022.10031400

Methodology	Issue Addressed	Research Gap
Developed using RStudio with SVM and NLP integration, the chatbot processes patient input through text preprocessing (using libraries like TM and SnowballC) and classification with SVM to map symptoms to appropriate medical specialists	Improves healthcare accessibility by offering instant, round-the- clock support. Enhances patient self-assessment, reducing unnecessary clinic visits and wait times	Data Reliability: Current data relies on general experiences rather than medically validated information; future improvements could involve licensed medical data. Field Testing: The system has not yet been tested in real healthcare environments, limiting insights into its practical effectiveness and adaptability to complex patient needs

Paper: "Self-Assessment Chatbot for COVID-19 Prognosis Using Deep Learning-based Natural Language Processing (NLP)"

Authors: EKI Thwala, AA Adegun, MO Adigun Year
Published: 2023
Published In: 2023 International Conference on Science, Engineering, and Business for Sustainable Development Goals (SEB-SDG)

DOI: 10.1109/SEB-SDG57117.2023.10124619

Methodology	Issue Addressed	Research Gap
The authors designed a COVID- 19 self-assessment chatbot using a deep learning approach, specifically employing Recurrent Neural Networks (RNNs) and Natural Language Processing (NLP). The chatbot is trained on user input to predict COVID-19 symptoms, providing responses and probability values for each query with high accuracy and quick response time	This work addresses limitations in existing chatbots, which often lack flexibility and the ability to emulate natural conversation in medical contexts. The chatbot aims to aid users in self- assessing COVID-19 symptoms, facilitating early intervention and reducing healthcare strain during the pandemic	Future improvements may focus on enhancing conversational accuracy for broader medical use cases, expanding the chatbot's ability to handle more complex and varied user queries, and improving the model's ability to adapt and learn from ongoing interactions for long-term utility

3. Proposed Work

3.1. Architecture Diagram

The architecture of the proposed work involves preprocessing the dataset, training the three models, and evaluating their performance. The diagram below illustrates the workflow:

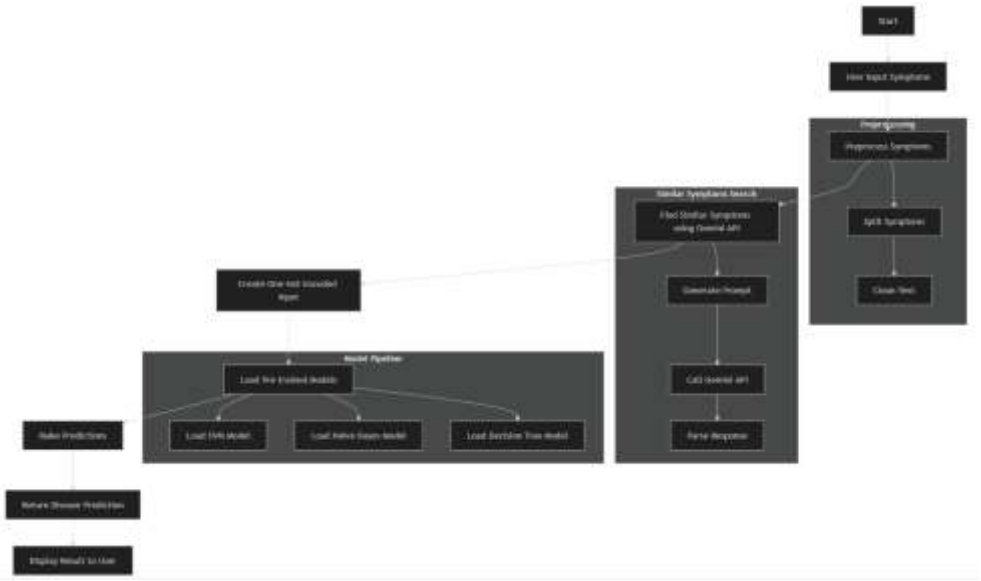

Figure 1. Architecture Diagram of the Proposed Work

3.2. Detailed Explanation

The architecture for disease prediction using machine learning involves the following steps: Users input their symptom list, which is processed by the Gemini API to find similar symptoms in the dataset and generate a modified symptom list. This modified list is then used for disease prediction. If the symptoms match the dataset, the system predicts the disease; otherwise, it returns a message indicating that the symptoms were not found. This approach ensures accurate predictions and appropriate feedback based on user inputs and dataset correlation.

4. Dataset Description

4.1. Dataset

The dataset contains disease names along with the symptoms faced by the respective patient. There are a total of 773 unique diseases and 377 symp- toms, with 246,000 rows. The dataset was artificially generated, preserving Symptom Severity and Disease Occurrence Possibility.Several distinct groups of symptoms might all be indicators of the same disease. There may even be one single symptom contributing to a disease in a row or sample. This is an indicator of a very high correlation between the symptom and that particular disease. A larger number of rows for a particular disease corresponds to its higher probability of occurrence in the real world. Similarly, in a row, if the feature vector has the occurrence of a single symptom, it implies that this symptom has more correlation to classify the disease than any one symptom of a feature vector with multiple symptoms in another sample.

4.2. Dataset Parameters

The dataset consists of the following Parameters (symptoms, 377 in total):

1. Anxiety and nervousness.
2. Depression.
3. Sharp chest pain.
4. Dizziness.
5. Insomnia.
6. Shortness of breath.
7. Abnormal involuntary movements.
8. Chest tightness.
9. Palpitations.
10. Irregular heartbeat.
11. Breathing fast.
12. Hoarse voice.
13. Sore throat.
14. Difficulty speaking.

15. Cough.
16. Nasal congestion.
17. Throat swelling.
18. Diminished hearing.
19. Lump in throat.
20. Throat feels tight.
21. Difficulty in swallowing.
22. Skin swelling.
23. Retention of urine.
24. Groin mass.
25. etc...

DETAILED WORKFLOW

Step 1: User Interface

- Display input form
- Show instructions
- Create text area for symptom input
- Add submit

button Step 2: Input Processing

- Clean user input
- Validate input format
- Split into individual

symptoms Step 3: Symptom Processing

- Match each symptom with database
- Use Gemini AI for finding similar symptoms
- Create standardized symptom

list Step 4: Feature Creation

- Generate binary feature vector
- Mark present symptoms
- Prepare for model

input Step 5: Prediction

- Run SVM model prediction
- Process

results Step 6: Output

- Format results
- Display to user
- Show any warnings or additional information

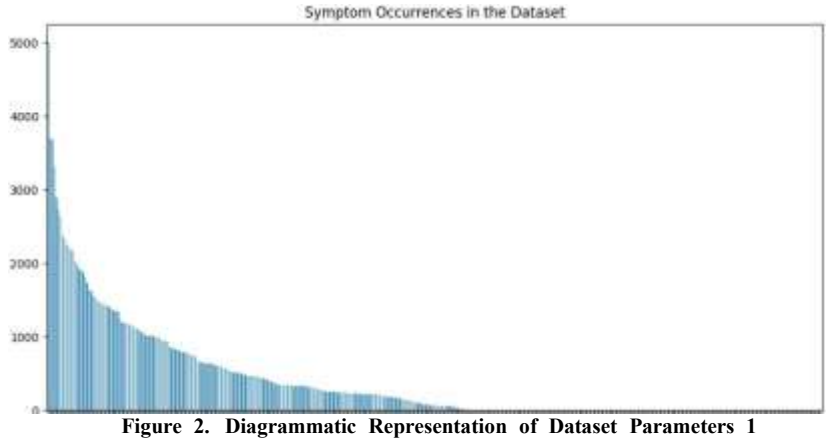

Figure 2. Diagrammatic Representation of Dataset Parameters 1

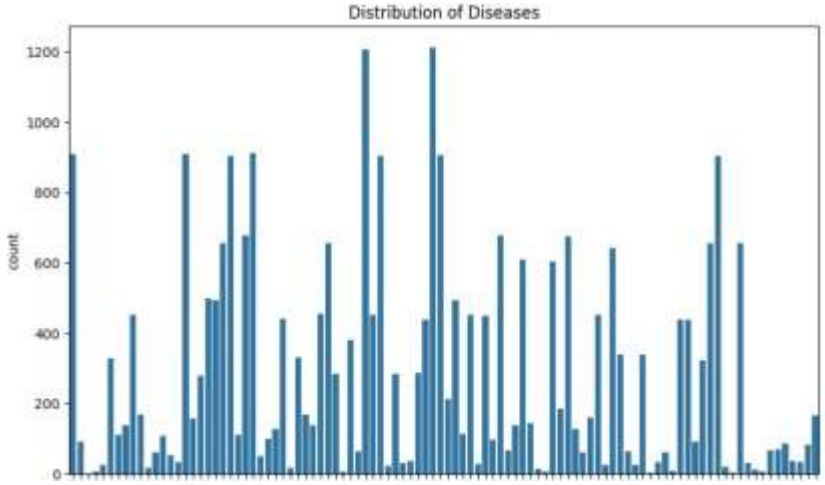

Figure 3. Diagrammatic Representation of Dataset Parameters 2

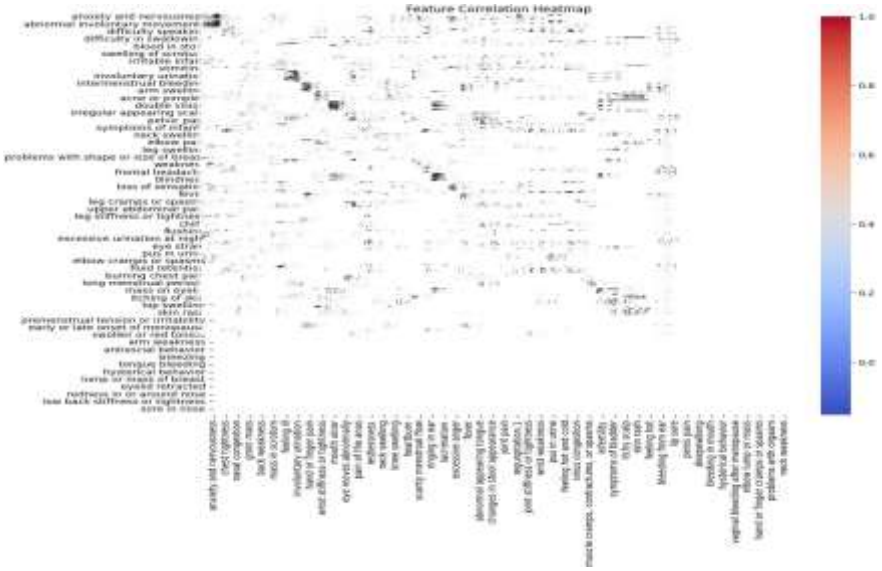

Figure 4. Diagrammatic Representation of Dataset Parameters 3

5. Results and Discussion

5.1. Evaluation Metrics

The following tables and figures represent the evaluation metrics of the three algorithms:

Table 1. Evaluation Metrics for Different Algorithms

Algorithm	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.8136	0.8181	0.8136	0.8136
Naive Bayes	0.8363	0.8442	0.8363	0.8310
SVM	0.8598	0.8731	0.8598	0.8605

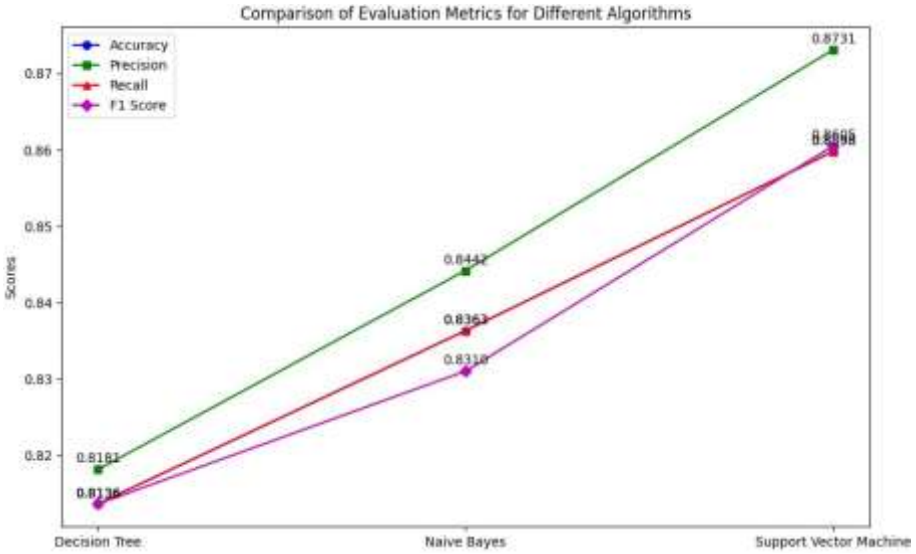

Figure 5. Comparison of Evaluation Metrics

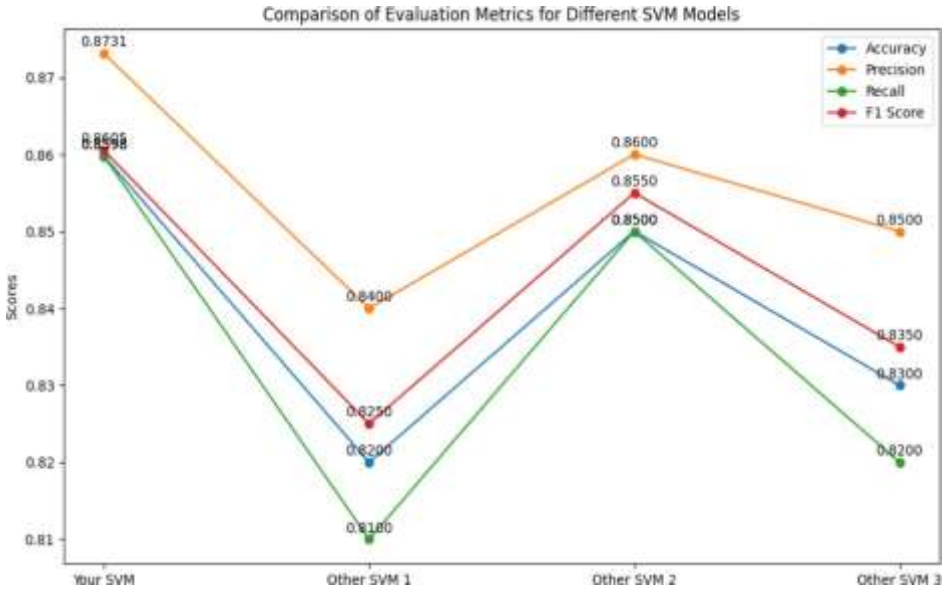

Figure 6. Evaluation Metrics: SVM vs. Other SVM Models

	Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding	BERT (Bidirectional Encoder Representations from Transformers)	0.98	0.97	0.96	0.97	0.98
Building a Medical Chatbot Using Support Vector Machine Learning Algorithm	Support Vector Machine (SVM)	0.92				
Building a Medical Chatbot Using Support Vector Machine Learning Algorithm	K-Nearest Neighbors (KNN)	0.87				
Building a Medical Chatbot Using Support Vector Machine Learning Algorithm	Naive Bayes	0.81				

MAAC-MEDICARE AT A CLICK: A Medical Chatbot Using Machine Learning Algorithm	Support Vector Machine (SVM)	0.94				
MAAC-MEDICARE AT A CLICK: A Medical Chatbot Using Machine Learning Algorithm	Naive Bayes	0.80				
MAAC-MEDICARE AT A CLICK: A Medical Chatbot Using Machine Learning Algorithm	K-Nearest Neighbors (KNN)	0.88				
Our Work	**Decision Tree**	**0.8136**	**0.8181**	**0.8136**	**0.8136**	
Our Work	**Naive Bayes**	**0.8363**	**0.8442**	**0.8363**	**0.8310**	
Our Work	**Support**	**0.8598**	**0.8731**	**0.8598**	**0.8605**	
	Vector Machine					

5.2. Discussion

The results indicate that SVM outperforms both Decision Tree and Naive Bayes in all evaluation metrics. The reasons for the superior performance of SVM include its ability to handle non-linear relationships, robustness to overfitting through regularization, and its focus on maximizing the margin between classes.

6. Gemini Integration Information

The Gemini API is an advanced tool designed to enhance healthcare applica- tions by integrating symptom analysis and disease prediction functionalities. It allows users to input a list of symptoms, which it processes by comparing with a comprehensive medical dataset. The API identifies similar symptoms and modifies the input list to ensure accurate predictions. This integration helps streamline the prediction process, enabling the system to deliver precise and reliable disease diagnoses based on user-reported symptoms and estab- lished medical correlations. The API ensures robust performance in handling symptom-based disease prediction tasks.

6.1. Integration Steps

- **Tool Selection:** Choose appropriate Gemini tools that align with the project requirements.
- **Data Integration:** Integrate the dataset into the Gemini environment for advanced preprocessing and analysis.

- **Model Enhancement:** Utilize Gemini's advanced algorithms and techniques to enhance the performance of existing models.
- **Evaluation:** Compare the performance of Gemini-enhanced models with the original models to assess improvements.

6.2. Benefits of Gemini Integration

- **Improved Accuracy:** Enhanced algorithms provide better accuracy and reliability in predictions.
- **Scalability:** Gemini tools are designed to handle large-scale datasets efficiently.
- **Advanced Analytics:** Access to sophisticated analytical tools for deeper insights into the data.

7. Conclusion

In conclusion, SVM is the most effective algorithm for this dataset, provid- ing a reliable model for classification tasks. Future work could explore the application of these algorithms to other datasets and investigate the impact of different preprocessing techniques and hyperparameter tuning.

References

[1] Batool Ali Majeed, Ammar Yasir Hardan, Batool Yasir Hardan, Dunya Faeq Munaf , "Accurate AI-Based Chatbot to Diagnose Heart Diseases Pre-Human Doctor Consultation" doi.org/10.18280/ria.380121.

[2] Chitra Ramesh and Jayalakshmi Sekar, "A Sur- vey on Healthcare Systems Using Internet of Thing" https://doi.org/10.1051/itmconf/20213701015.

[3] Rahul Pradhan, Jaya Shukla, Mani Bansal, "'K-Bot' Knowl- edge Enabled Personalized Healthcare Chatbot" DOI: 10.1088/1757- 899X/1116/1/012185.

[4] S Moulya, T R Pragathi, "Mental Health Assist and Diagnosis Con- versational Interface using Logistic Regression Model for Emotion and Sentiment Analysis" Conference Series 2161 (2022) 012039.

[5] Ragavi V, SanthaSheela A C, GeethaiNarayanani Kannaiyan, "Impact of Artificial Intelligence in the field of Health Care ," Conference Series 1831 (2021) 012006.

[6] Wan Muhamad Asyraf Wan Zaki, Muhammad Faiz Md Shakhih, Muhammad Hanif Ramlee, and Asnida Ab Wahab, "Smart Medical Chatbot with Integrated Contactless Vital Sign Monitor" Conference Series 1372 (2019) 012025.

[7] Sunanda Rani, Dong Jining, Dhaneshwar Shah, Siyanda Xaba, and Prabhat Ranjan Singh, "The Potential Application of Artificial Intel- ligence in Healthcare and Hospitals" ITM Web of Conferences 53, 01005 (2023) ICDSIA-2023.

[8] Kisito Kabore, Omar Sawadogo, Yaya Traore, Julie Thiombiano, "To- wards a Chatbot for Medical Diagnosis Based on Patient Symptoms"

[9] Ibraheem Altamimi, Abdullah Altamimi, Abdullah S. Alhumimidi, Abdulaziz Altamimi, Mohamad Hani Temsah, "Artificial Intelligence (AI) Chatbots in Medicine: A Supplement, Not a Substitute" DOI: 10.7759/cureus.40922.

[10] Magdalena Görtz, Kilian Baumgärtner, Tamara Schmid, Marc Muschko, Philipp Woessner, Axel Gerlach, Michael Byczkowski, Holger Sültmann, Stefan Duensing, Markus Hohenfellner, " An Artificial Intelligence- Based Chatbot for Prostate Cancer Education: Design and Patient Eval- uation Study"

[11] Tamizharasi B., Jenila Livingston L.M., S. Rajkumar, ""Building a Medical Chatbot Using Support Vector Machine Learning Algorithm"" DOI: 10.1088/1742-6596/1716/1/012059.

[12] Arun Babu, Sekhar Babu Boddu, "BERT-Based Medical Chatbot: En- hancing Healthcare Communication through Natural Language Under- standing" DOI: 10.1016/j.rcsop.2024.100419.

[13] Rani Astya, Puja Kumari, Apoorva Goswami, Abinash Senapati, "MAAC-MEDICARE AT A CLICK: A Medical Chatbot Using Machine Learning Algorithm"

[14] Zain Shaukat, Wisal Zafar, Waqas Ahmad, Ihtisham Ul Haq, Ghas- san Husnain, Mosleh Hmoud Al-Adhaileh, Yazeed Yasin Ghadi, Abdul- mohsen Algarni, "Revolutionizing Diabetes Diagnosis: Machine Learn- ing Techniques Unleashed" DOI: 10.3390/healthcare11212864.

[15] Seng-Keong Tan, Siew-Chin Chong, Kuok-Kwee Wee, Lee-Ying Chong, "Personalized Healthcare: A Comprehensive Approach for Symptom Di- agnosis and Hospital Recommendations Using AI and Location Services" DOI: 10.33093/jiwe.2024.3.1.8.

[16] Author16, "Chatbot for Healthcare and Oncology Applications" doi:10.18280/ria.380136.

[17] Mark Lawrence, MD Istiyak, Mohd Aman, "Healthcare Chatbot Sys- tem"

[18] Mrs. Rashmi Dharwadkar, Dr. Neeta A. Deshpande, "A Medical Chat- bot"

[19] Ryan Matthew, David Agustriawan, Mario Donald Bani, Muammar Sadrawi, Nanda Rizqia Pradana Ratnasari, Moch Firmansyah, Arli Aditya Parikesit , "The Development of A Medical Chatbot Using The SVM Algorithm" DOI: 10.1109/ICORIS56080.2022.10031400.

[20] EKI Thwala, AA Adegun, MO Adigun, "Self-Assessment Chatbot for COVID-19 Prognosis Using Deep Learning-based Natural Language Processing (NLP)" DOI: 10.1109/SEB-SDG57117.2023.10124619.