International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Machine Learning in Healthcare: Developing an AI Assistant Doctor for Symptom-Based Disease Prediction

[1]Dr Dipannita Mondal, [2]Harshad Shinde,[3] Sarang Baghele,[4]Pratham Kadam,[5]Darshan Thengal

[1]HOD of Department of Artificial Intelligence and Data Science Engineering DYPCOEI, Varale, Pune (SPPU), Maharashtra, India. mondal.dipannita26@gmail.com .

[2,3,4,5]Final Year Student of Department of Artificial Intelligence and Data Science Engineering DYPCOEL, Varale, Pune (SPPU), Maharashtra, India. [2]harshadshinde2405@gmail.com, [3]sarangbaghele1@gmail.com, [4]prathamkadam777@gmail.com, [5]darshanthengal336@gmail.com .

---***---

1. ABSTRACT:

The rapid advancements in data science and machine learning have revolutionized various sectors, particularly healthcare. This research introduces an innovative AI Assistant Doctor system that utilizes machine learning techniques, specifically a Support Vector Classifier (SVC), to predict potential diseases based on user-reported symptoms. The system aims to bridge the gap between symptom recognition and understanding their significance, promoting early intervention and improved health outcomes. The AI Assistant Doctor offers a comprehensive healthcare tool that goes beyond disease prediction, providing users with detailed information about illnesses, preventive measures, drug recommendations, nutritional guidance, and personalized exercise suggestions. The SVC model is trained on a diverse dataset of symptoms and diseases, enabling accurate predictions. The system is designed to be accessible through a user-friendly web interface, allowing individuals to input their symptoms and receive immediate insights. In addition to disease prediction, the AI Assistant Doctor equips users with a thorough understanding of the predicted condition, including its symptoms, causes, and typical progression. It also offers preventive strategies, suggested medications, dietary advice, and customized physical activity recommendations to help users manage their health effectively. By leveraging machine learning algorithms and web-based frameworks, this research presents a cutting-edge approach to disease prediction and healthcare management, addressing the growing need for accessible medical information and timely interventions. The AI Assistant Doctor has the potential to alleviate pressure on healthcare systems, improve individual health outcomes, and support medical practitioners in delivering precise diagnoses and personalized treatment plans.

Keywords: Machine learning, Disease Prediction, Support Vector Classifier (SVC), Healthcare, Symptoms, Prevention Personalized recommendations.

2. INTRODUCTIONS:

The swift progress in technological advancements, especially in data science and machine learning, has transformed various sectors, with healthcare being a prime example. As healthcare systems worldwide grapple with issues of accessibility and timely diagnosis, there is a growing need for solutions that can help people interpret their health symptoms independently. Many individuals postpone seeking medical attention due to a lack of understanding about their symptoms, which can result in poorer health outcomes. While early intervention is crucial for effective disease management, there often exists a gap in knowledge between recognizing symptoms and comprehending their potential significance. The research introduces a system that employs machine learning to forecast illnesses based on user-reported symptoms. This comprehensive healthcare tool goes beyond disease prediction, offering a range of health-related information including illness descriptions, preventive measures, drug recommendations, nutritional guidance, and exercise suggestions. By utilizing data-driven approaches, the system aims to enhance health awareness and promote timely intervention in potential disease Management. The AI Assistant Doctor initiative seeks to create a cutting-edge healthcare application that utilizes machine learning techniques, specifically a Support Vector Classifier (SVC), to forecast potential health issues based on symptoms reported by users. This innovative system is intended to meet the increasing demand for readily available medical information and promote early detection in the management of diseases. The AI Assistant Doctor offers several essential capabilities:

1. Disease prediction based on symptoms utilizing a trained SVC model.
2. Detailed explanations of various illnesses.
3. Advice on prevention and recommended actions.
4. Suggestions for potential medication treatments.
5. Dietary guidance specific to the predicted conditions.
6. Personalized exercise suggestions.

At the heart of this system lies a Support Vector Classifier (SVC), a proven machine-learning algorithm renowned for its classification capabilities. The model underwent training using a comprehensive dataset containing various symptoms and illnesses, allowing it to make accurate predictions about potential conditions based on user-provided information. By combining machine learning models with web-based frameworks like Flask, the system becomes widely accessible, offering an easy-to-use interface where users can enter their symptoms and quickly receive insights about possible diseases. The framework is engineered to both diagnose and provide practical health guidance. The research presents an innovative approach to disease prediction using machine learning techniques. This system leverages user-reported symptoms as input data to generate forecasts of potential illnesses. By utilizing machine learning algorithms, the system can analyze patterns and correlations within the symptom data to make predictions about likely health conditions. After generating a prediction, the system equips users with:

1. A comprehensive overview of the anticipated illness, covering its indicators, origins, and typical development.
2. Preventive strategies users can implement to mitigate their condition's progression or avert the disease entirely.
3. Suggested pharmaceuticals typically used to treat the condition, giving users insight into potential therapies.
4. Nutritional advice to aid in managing the illness.

The development of this system has responded to the growing need for accessible healthcare information. With healthcare resources often stretched thin, especially in underserved areas, such tools have the potential to alleviate pressure on healthcare systems by providing preliminary guidance and improving individual health outcomes through timely actions. The use of machine learning not only enhances the accuracy of disease predictions, but also allows for continuous improvement as more data become available.The medical field is being revolutionized by artificial intelligence, which enhances the precision of diagnoses and the effectiveness of treatments. This initiative incorporates sophisticated machine learning algorithms to support medical practitioners. It emphasizes streamlined processes, enhanced precision, and tailored patient treatment approaches.

3. LITERATURE SURVEY:

Paper Details	Problem Discussion	Algorithm /Technique used	Parameter Consider	Outcome
A new Ai Assisted Medical Molecular Image	The problem is to create an AI system that	GAN Model Principle.	Accuracy, Processing Speed, Data Quality,	The proposed GAN-based system produced

Diagnostic Model.	automates and enhances medical image diagnosis.		Model Complexity, Training Data.	high-quality CT and MRI images, outperforming other methods in accuracy and speed, while requiring less storage.
Development of Artificial Neural Network Model for Medical Specialty Recommendation.	The problem is to create a model that recommends medical specialties based on patient symptoms and comorbidities.	MLP Model (Multilayer Preceptron)	Data set size, Data Quality, Feature Engineering	Ai helps pick right doctor for your Symptoms
An optimal deep feature – based Ai chat conversation system for smart medical application.	Existing chatbot models struggle with accurate and efficient answer retrieval for medical queries, often requiring large datasets, complex architectures.	LBDBC Model (Lion-Based Deep Belief Chat box.)	Tokenization Method, Data Cleaning	A medical chatbot was built using a Lancaster stemmer. It can answer health questions with reasonable accuracy.
Data Exchange Standards in Healthcare.	The healthcare industry faces significant challenges in data exchange among stakeholders due to the lack of standardized formats and protocols. This results in interoperability issues, fragmented data systems, inefficient sharing, compliance challenges, and barriers to data-driven decision-making. These challenges hinder effective patient care and resource utilization.	Data Mapping Techniques Data Transformation Techniques Governance and Adoption Strategies: Governance Frameworks Stakeholder Engagement Strategies Technological Enablers: Cloud Computing Solutions APIs (Application Programming Interfaces) Blockchain Technology.	Data Formats and Structures Interoperability Requirements Data Governance and Security Stakeholder Engagement and Adoption Technical Infrastructure and Capabilities Scalability and Flexibility Maintenance and Support Regulatory and Policy Considerations.	Implementing data exchange standards in healthcare leads to improved interoperability, efficient data sharing, enhanced compliance with regulations, better stakeholder engagement, streamlined operations, scalability, support for data-driven decision-making, and effective change management,

				ultimately enhancing healthcare delivery and patient outcomes.
Recommender systems in the healthcare domain: state-of-the-art and research issues.	The vast amount of clinical data scattered across different sites on the Internet hinders users from finding helpful information for their well-being improvement.	Collaborative Filtering (CF) 9, Content-based Filtering, Knowledge-based Approaches, Hybrid Approaches.	Usage context User profiles Item characteristics (e.g., food, drugs, healthcare services	Food recommendation Drug recommendation 9 Health status prediction Healthcare service recommendation Healthcare professional recommendation .
Medicine Recommend system using machine learning.	uses machine learning algorithms to predict diseases and recommend appropriate medicines based on user-provided symptoms	Decision Tree Classifier Random Forest Classifier Naive Bayes Classifier	The system was developed using the three machine learning algorithms mentioned above. The Naive Bayes Classifier achieved the highest accuracy of 98.12% in predicting diseases.	The proposed system can be used by users to quickly get disease predictions and medicine recommendations, especially in situations where accessing a doctor may be difficult, such as during the COVID-19 pandemic.
"AI-based medical e-diagnosis for fast and automatic ventricular volume measurement in patients with normal pressure hydrocephalus,"	This addresses the critical need for an efficient and accurate method to automatically measure ventricular volume (VV) and intracranial volume (ICV) in patients with normal pressure hydrocephalus (NPH).	3D Convolutional Neural Network (3D CNN)	Parameter Minimization: Strategies aimed at developing parameter-free algorithms, thereby reducing complexity and the need for extensive parameter tuning.	The ventricle segmentation algorithms achieve high accuracy, with Dice similarity coefficients and Hausdorff distances exceeding 0.99

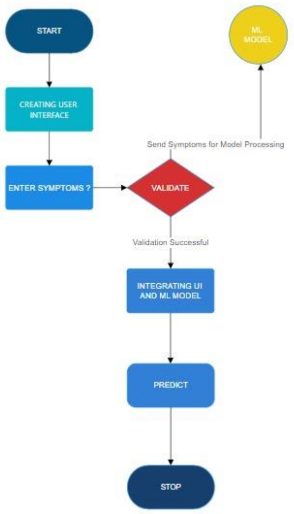
authored by Xi Zhou et al. and published in Neural Computing and Applications in 2022.	Manual segmentation in brain CT and MRI images is time-consuming, labor-intensive, and prone to variability, making it impractical for routine clinical use.		Anatomical Structures: The inclusion of relevant anatomical features, such as papillary muscles, which can significantly affect segmentation outcomes.	compared to manual methods. Processing times are efficient, averaging 3.4 seconds for CT and 1.9 seconds for MRI images. These advancements enhance clinical assessments of ventricular volume and intracranial volume in normal pressure hydrocephalus patients.
Pre-hospital treatment of patients with acute coronary syndrome: Recommendations for medical emergency teams	The document outlines guidelines for managing ACS patients before hospital arrival, focusing on early ECG diagnosis, telemedicine use, and timely treatment.	Early use of P2Y12 inhibitors (e.g., clopidogrel, prasugrel, ticagrelor) in ACS treatment.	Consider contraindications (e.g., bleeding risk, stroke history) and patient factors (age, weight).	Enhanced outcomes by reducing clots and complications, with prasugrel and ticagrelor being more effective than clopidogrel.
Medicine Recommendation System Based On the Patient Reviews 2020.	The growing information overload on the internet, emphasizing the need for personalized healthcare recommendation systems. It identifies the lack of specialized recommender systems in healthcare and the challenges in analyzing patient reviews due to noisy and incomplete data. The need to balance accuracy, efficiency, and scalability is crucial, particularly in medical applications. The study	N- gram Model Evaluation, Light GBM.	N-gram for sentiment analysis LightGBM for building the recommendation model Evaluating and comparing the accuracy of the N-gram and LightGBM models Model Evaluation Metrics 4: Accuracy of the N-gram model (80%)	The proposed medicine recommendation system features five modules: database, data preparation, recommendation model, model evaluation, and data visualization. After preprocessing the data, the LightGBM model achieved

	evaluates machine learning techniques, including N-gram and LightGBM, to optimize these trade-offs. Finally, it highlights the absence of a comprehensive framework to integrate patient review data, sentiment analysis, and recommendation modeli ng.		Accuracy of the LightGBM model (90 %)	90% accuracy, outperforming the N-gram model's 80%. The system effectively integrates sentiment analysis and machine learning techniques to provide reliable drug recommendation s.
Accurate AI-Based Chat Bot to Diagnose Heart Disease Pre-Human Doctor Consultation. [2024]	The paper proposes an AI-based chatbot to improve early diagnosis of heart diseases using machine learning algorithms, focusing on accuracy to prevent life-threatening conditions	XGBoost: Extreme Gradient Boosting Algorithm, SVM: Support Vector Machine Algorithm, LR: Logistic Regression Algorithm	XGBoost: Extreme Gradient Boosting Algorithm, SVM: Support Vector Machine Algorithm, LR: Logistic Regression Algorithm	The paper outlines the development of an AI-based chatbot using machine learning algorithms.
Diet Recommendatio n System Using Machine Learning2023	in the paper is the need for a personalized diet recommendation system that assists individuals in making informed dietary choices based on their unique preferences, health goals, and nutritional requirements. Key challenges include ensuring the system can analyze food content effectively, utilize advanced machine learning techniques for accurate recommendations, provide a user-friendly experience, and maintain scalability to accommodate a growing number of users.	Random Forest, K-Means Clustering, Long Short-Term Memory (LSTM).	USDA Nutrition Dataset, User's Food Intake, User Preferences and Nutritional Information, BMI (Body Mass Index), Relevant Features.	The diet recommendation system offers personalized meal plans based on user details such as age, weight, height, and dietary preferences, recommending three types of diets: weight loss, weight gain, and healthy eating. It calculates the user's Body Mass Index (BMI) to assess their weight category.

4. PROBLEM STATEMENT:

Develop an AI Assistant Doctor to enhance medical diagnosis and treatment recommendation by leveraging artificial intelligence to analyze patient data efficiently, ensure accurate diagnoses, and provide personalized treatment plans, thereby improving healthcare delivery and patient outcomes.

5. METHODOLOGY:



Block diagram of system

5.2 Mathematical Model (SVM):

SVM is used for classification by finding the optimal hyperplane that separates different classes (diseases) in the feature space (symptoms).

- Decision Function:

SVM aims to find the hyperplane $w \cdot X + b = 0$ that maximally separates the classes. For multi-class classification (like this project), we use either a one-vs-one or one-vs-rest approach to generalize SVM for more than two classes.

For each binary classifier:

5.1 System Architecture Diagram (with SVM):

The flow:

1. Data Input (patient symptoms).
2. Data Preprocessing (scaling and cleaning the dataset).
3. Model Training (SVM is trained on labeled data to classify diseases).
4. Prediction (new symptoms are passed into the trained SVM model).
5. Diagnosis Output (the predicted diseases)

$$f(X) = \text{sign}(w \cdot X + b)$$

where:

w is the weight vector defining the hyperplane.

X is the input symptom vector.

b is the bias term.

- **Loss Function**:

The goal is to maximize the margin between classes while minimizing classification errors. The loss function for SVM is the hinge loss**:**

$$L = \max(0, 1 - y_i(w \cdot X_i + b))$$

Where:

yi is the true label, and Xi is the feature vector of symptoms for sample (i)

In multi-class SVM, this can be extended using a one-vs-rest or one-vs-one strategy, where the SVM algorithm is applied to pairs or all classes.

5.3 Algorithm:

Support Vector Machine (SVM) Classifier

Model Selection**:** SVM works well for high-dimensional data and performs robustly in classification tasks, especially when there is a clear margin of separation between classes.

Kernel Choice:

RBF Kernel**:** Used for non-linear data, the RBF kernel maps symptoms into a higher-dimensional space, allowing SVM to find complex decision boundaries.

$$K(x_i, x_j) = \exp\left(-\gamma \cdot \|x_i - x_j\|^2\right)$$

Linear Kernel**:** For linearly separable data, the SVM finds a straight-line boundary

$$K(x_i, x_j) = x_i \cdot x_j$$

When data can be separated linearly and does not require complex decision boundaries, a linear kernel is typically employed. This approach is particularly efficient in terms of computation, making it suitable for datasets that are large or have several dimensions. In addition, it produces a model that is easier to interpret. In scenarios where simplicity and rapid processing are crucial, such as medical diagnostic applications, a linear kernel is an excellent choice for classification tasks.

6. CONCLUSION:

This study introduces an AI-powered medical assistant that employs machine learning techniques, particularly a Support Vector Classifier (SVC), to predict diseases based on symptoms. The system aims to meet the increasing demand for accessible and precise medical advice, providing a user-friendly interface for early disease detection. By examining symptoms reported by patients, the AI not only forecasts potential illnesses but also delivers a comprehensive health package, including preventative strategies, drug recommendations, and tailored diet and exercise plans. The incorporation of machine learning technology enhances the accuracy of predictions, enabling users to make informed health decisions. This innovation has the potential to reduce the workload of healthcare professionals by offering initial diagnostic tools and encouraging individuals to take a proactive approach to their wellbeing. As the model continues to improve and larger datasets become available, this AI-based solution could play a significant role in enhancing healthcare delivery and patient outcomes.

7. REFERENCES:

[1] Majeed, B. A., Hardan, A. Y., Hardan, B. Y., & Munaf, D. F. (2024). Accurate AI-Based Chatbot to Diagnose Heart Diseases Pre-Human Doctor Consultation. Revue d'Intelligence Artificielle, 38(1), 213.

[2] G. Premananthan, B. Nagaraj, J. Jaya, A new AI assisted medical molecular image diagnostic model, Journal of Intelligent & Fuzzy Systems 44 (2023) 9027–9037.

[3] Lal, M., & Neduncheliyan, S. (2023). An optimal deep feature–based AI chat conversation system for smart medical application. Personal and Ubiquitous Computing, 27, 1483–1494.

[4] Golagi, R., Sravani, V., Mohan Reddy, T., & Kavitha, C. (2023). Diet recommendation system using

machine learning. Dogo Rangsang Research Journal, 13(4), 118-125. DOI: 10.36893.DRSR.2023.V13I04.118-125

[5] Khairnar, P., Avula, V., Hargane, A., Baisware, P., Medicine Recommend System Using Machine Learning, International Journal of Scientific Research in Science, Engineering and Technology 9 (2022) 247-250. DOI.

[6] Zhou, X., Ye, Q., Yang, X., Chen, J., Ma, H., Xia, J., Del Ser, J., & Yang, G. (2022). "AI-based medical e-diagnosis for fast and automatic ventricular volume measurement in patients with normal pressure hydrocephalus". Neural Computing and Applications, 35(22), 16011–16020. DOI: 10.1007/s00521-022-07048-0.

[7] Kubica, J., Adamski, P., Ładny, J.R., Kaźmierczak, J., Fabiszak, T., Filipiak, K.J., Gajda, R., Gąsior, M., Gąsior, Z., Gil, R., Gorący, J., Grajek, S., Gromadziński, L., Gruchała, M., Grześk, G., Hoffman, P., Jaguszewski, M.J., Janion, M., Jankowski, P., Kalarus, Z., Kasprzak, J.D., Kleinrok, A., Kochman, W., Kubica, A., Kuliczkowski, W., Legutko, J., Lesiak, M., Nadolny, K., Navarese, E.P., Niezgoda, P., Ostrowska, M., Paciorek, P., Siller-Matula, J., Szarpak, Ł., Timler, D., Witkowski, A., Wojakowski, W., Wysokiński, A., Zielińska, M., Pre-hospital treatment of patients with acute coronary syndrome: Recommendations for medical emergency teams. Expert position update 2022, Cardiology Journal 29 (2022) 540. DOI: 10.5603/CJ. A 2022.0026.

[8] Hasuki, W., Agustriawan, D., Parikesit, A.A., Sadrawi, M., Firmansyah, M., Whisnu, A., Natasya, J., Mathew, R., Napitupulu, F.I., & Ratnasari, N.R.P. Development of Artificial Neural Network Model for Medical Specialty Recommendation. Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, 2021.

[9] Tran, T.N.T., Felfernig, A., Trattner, C., Holzinger, A., Recommender systems in the healthcare domain: state-of-the-art and research issues, Journal of Healthcare Informatics 15 (2020). Published online: 17 December 2020.

[10] Rao, T.V.N., Unnisa, A., Sreni, K., Medicine Recommendation System Based On Patient Reviews, International Journal of Scientific & Technology Research 9 (2020) 3308.

[11] Apostol, A.-S., Muntean, C., Vernic, C., Bogdan, T., Data Exchange Standards in Healthcare, Anale. Seria Informatica. Vol. X fasc. 2 (2012), 45.