JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Review on Symptom Driven Disease Prediction Using ML Algorithms

M. Ganesh[1], M. Shiva[2], B. Niveditha[3], K. Meghana[4], Dr. L. Chandra Sekhar Reddy[5]

[1,2,3,4]UG Students, Department of CSE(DS), CMRCET, Hyderabad, Telangana

[5]Associate Professor, Department of CSE(DS), CMRCET, Hyderabad, Telangana

Abstract : The prediction of diseases based on symptomatic information has emerged as a pivotal area in healthcare, leveraging data-driven methodologies to enhance diagnostic accuracy. This abstract provides a concise overview of the techniques employed in disease prediction via symptoms. By harnessing machine learning, artificial intelligence, and data analytics, researchers and healthcare practitioners have achieved notable progress in forecasting diseases using patient-reported symptoms. This abstract underscore the significance of early detection and proactive management of healthcare through predictive models. Various data sources, including electronic health records, wearable devices, and patient self- reports, contribute essential symptom data for predictive modeling. Challenges associated with noisy and incomplete data are addressed, emphasizing data preprocessing and feature engineering to refine predictions. Machine learning algorithms, such as decision trees, support vector machines, and neural networks, are applied to construct predictive models. Incorporating clinical knowledge and domain expertise enhances model performance. Genetic, demographic, and environmental factors are integrated to bolster robustness. Ethical considerations encompassing patient privacy, data security, and potential biases are discussed, highlighting responsible model deployment. Transparent, interpretable AI techniques aid in deciphering model predictions for informed decision-making. In summary, disease prediction using symptoms offers a promising avenue for early intervention and personalized treatment. This abstract encapsulates the methodologies, challenges, and ethical implications surrounding symptom-based predictive models, emphasizing their potential to reshape healthcare through data-driven insights.

***IndexTerms*-Machine learning, Decision Making, ML Technique**

I. Introduction

The healthcare sector is undergoing a transformation driven by new technologies, especially the integration of machine learning (ML) techniques in disease prediction Among these developments, the symptom-driven prediction emerges as a fundamental approach set to revolutionize medical research and prognosis and depends heavily on knowledge production. But the complexity and variability of symptoms presenting in different diseases pose great challenges when it comes to accurate diagnosis The advent of ML algorithms heralded a new era, enabling healthcare providers to capture a vast amount of medical information in a collection and found complex structures and relationships You can predict the names. This paradigm shift not only facilitates early diagnosis and intervention but also allows for personalized health care tailored to the specific needs of individual patients. This paper aims to deepen the concept of symptom-driven disease prediction, clarify its central role in contemporary healthcare, and explore the transformative potential of ML-based approaches in transformation in new paradigms of diagnostic management Current research and context. Through a comprehensive review of studies, we seek to highlight the importance of symptom-driven disease prognosis in early health care delivery, improving patient outcomes, and ultimately emphasizing advancing healthcare towards a more patient-centred and data-driven. Through the full integration of current research, new approaches, and real-world applications, we seek to elucidate the transformative potential of ML-driven approaches in reframing disease diagnosis and management. By highlighting the implications of disease prognosis from symptoms for health outcomes, resource allocation, and patient-centred care delivery, this paper seeks to chart a course toward a future with predictive health care being an important component of precision medical intervention prioritization.

II. LITERATURE REVIEW

A. Heart Disease

The scientists have implemented new systems based on the machine learning techniques to detect and monitor diseases, specifically coronary heart disease. Otoom [1] constructed a model using the Cleveland heart dataset with 304 instances and 77 attributes. Otoom has performed some experiments with algorithms such as Bayes-Net, SVM and FT that showed promising results, where SVM reached an accuracy of 88.3% after Holdout testing. Besides, Cross Validation testing confirmed that Bayes-Net and SVM had achieved precision of 83.8%, while FT yielded an accuracy rate of 81.5%. Feature selection applied in this context raised the accuracies of these classifiers even higher: Bayes-Net (84.5%), SVM (85.1%), and FT (84.6%).

Similarly, Vembandasamy [3] conducted research utilizing the Naïve-Bayes algorithm to identify heart diseases, drawing data from a well-known institute for diabetic patients in Chennai consisting over five hundred patients respectively. The robustness of this algorithm is shown by its precision value which is equal to 86.419%. Tan also introduced Genetic Algorithm (GA) and SVM in combination as a hybrid approach to improve disease classification rates leading to good outcomes for heart diseases like 84.07% and diabetic.

B. Liver Disease

In the realm of prediction of liver diseases, Vijayarani [9] investigated how effective the Support Vector Machine (SVM) and Naïve Bayes algorithms were in the Indian Liver Patient Dataset (ILPD) based on a dataset from UCI data repository. Vijayarani compared precision and implementation time with 561 instances and 11 attributes. Importantly, Naive Bayes was found to be 61.29% correct within 1770.00 MS whereas SVM had a higher precision of 79.67% for a period of 3410.00 MS. Besides having lower accuracy than Naïve Bayes, SVM takes more time during its implementation stage than it takes for Naïve Bayes to implement it. Moreover, Gulia [4] et al undertook research that involved identifying liver patients using data from UCI repository by employing five intelligent classification methods namely; J48, MLP, Random Forest, SVM and Bayesian Network respectively. With initial merging of datasets, feature selection as well as comparative evaluation undertaken in multiple phases; various degrees of accuracy across algorithmic approaches were demonstrated by Gulia et al. Following feature selection processes MLP, SVM, Random Forest and Bayesian Network had accuracies ranging between 70.668% to 71.8686%.

C. Infectious Disease

The On the other side of infectious disease prognostication, Tirmizi [2] commenced a research study into Malaysia's Dengue epidemic using Data Mining techniques. In hot environments such as Malaysia, Thailand, and Indonesia, dengue is a viral illness posing an enormous threat. Tirmizi took advantage of data from the Department of Health of the State of Selangor by utilizing classification methods like Decision Trees (DT), Artificial Neural Networks (ANN), and Rule-Based Systems (RS). Using WEKA data mining tool through rigorous analysis DT accuracy was observed at 99.96%, ANN at 99.97% and RS gave an overwhelming 101%. Moreover, after employing Percentage Split (PS), both DT and ANN maintained high accuracy rates of 99.93% which shows how robust these methods are in predicting dengue epidemics. Even though Rule-Based Systems appears to have gone beyond 100% accuracy, it had a reasonable precision which was still commendable as it stood at 99.73%. Similarly, Fathima [10] engaged herself with Arbovirus-Dengue diseases prediction using Support Vector Machine (SVM) as the preferred Data Mining method. The study used data derived from surveys done in India's Chennai and Tirunelveli hospitals and labs incorporating 28 characteristics and 5006 samples for analysis. For this purpose, R project version 2.1.

III. EXISTING SYSTEMS

A. System Checker App

Symptom Checker app provides users with a simple and easy way to assess their symptoms and diagnose potential health problems, enabling them to make informed treatment decisions will be explored These apps provide educational resources, increase health literacy and enable individuals to take proactive steps in managing their health. However, while they provide valuable guidance, they are not a substitute for expert or diagnostic medical advice, and users should exercise caution in the interpretation of results In addition the privacy and security concerns surrounding the handling of personal health information underscore the importance of transparency and compliance with privacy laws. Despite the limitations, when judiciously combined with professional medical guidance, symptom assessment apps can play a valuable role in supporting personal health management and well-being.

B. Electronic Health Records (EHRs)

A Electronic health records (EHRs) represent an important innovation in healthcare technology, enabling patient data to be stored, accessed, and managed These digital records provide comprehensive, available information real-time tracking of patients' medical histories, treatments, medications, and diagnostic tests -accessibility of EHRs to authorized healthcare providers in the healthcare system optimizes care planning, reduces medical errors, and improves patient outcomes by better managing data and facilitating information sharing. In addition, EHRs support clinical decision making by providing relevant patient information in a timely manner, enabling healthcare professionals to make informed medical decisions However, challenges such as administrative issues, data security problems, paperwork burden on the healthcare providers They hold great promise to make a difference, for healthcare collaboration, and ultimately, improve the quality and efficiency of patient care.

C. Wearable Device

In healthcare, potential means that the devices will provide half the functionality from Fitness assessment to remote patient monitoring. Collection of device information Though handheld devices are empowered and their health checked in real time, they also impose many restrictions putting their plausible concerns about the reliability of biometric facilities there, especially in terms of data accuracy and safety and security. , by breaking the skeleton Or worries about unauthorized access, worries can also create obstacles to uninterrupted device issues in the existing health care system, as it will in which the benefits of profitability, pressure, pressure and pressure may be affected by factors such as profitability, pressure and pressure. of the names Overall, continued advances in wearable technology hold promise to address these challenges and expand the role of wearable devices in personal healthcare.

IV. LIMITATIONS OF EXISTING SYSTEMS

1.*Accuracy problem:* Both systems can experience accuracy problems, which can lead to incorrect data or diagnoses. This can be caused by factors such as incomplete data, algorithmic limitations, and human error in data entry.

2.Data privacy and Security risks: Vulnerable to data privacy and security risks. Concerns arise about protecting sensitive personal health information, including risks of access, data breaches, or data misuse.

3.Performance challenges: There may be challenges in communicating between two systems, affecting the ability to seamlessly exchange information with health systems or other providers This can hinder care planning and improvement, as well as hinder coordination the effective edge of all types.

4.Complex user interfaces: May suffer from bottlenecks in user interface design, which can lead to usage issues and inefficiencies in user interfaces Complex connections or unclear navigation paths can hinder user experience and performance.

5. Overdependence and abuse: Can be vulnerable to overdependence or abuse. Users or healthcare providers may rely too heavily on systems for diagnosis or decision-making, and may overlook important clinical features or fail to reason effectively.

V. SYSTEM REQUIREMENTS

A, Hardware Requirements:

1. Minimum 4GB RAM
2. Hard Disk: 500GB
3. Processor: Intel Core i3(min)

B. Software Requirements:

1. Operating system: Windows 10(min)
2. Coding Language:, Python
3. Database: MySQL

C. Libraries:

1. Matplotlib
2. NumPy
3. Pandas
4. Tkinter

VI. METHODOLOGY

A GUI-based disease prediction system using machine learning algorithms. The use of a Libraries and modules for user-generated graphics can incorporate attributes, too The system predicts possible disease through machine learning algorithms such as Decision Making tree, random forest, naive Bayes, and nearest neighbours.

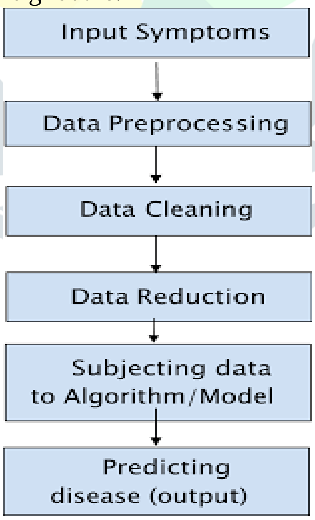


Fig:1. Working

Data Collection: Gather information including disease profiles, symptoms, and labels.
Preprocessing: Clean and preprocess the data set by handling missing values and encodings Variables in groups.
Selection: Identify the appropriate characteristics/factors to predict.
Algorithm selection: Decision Trees, Random Forest, k- . Close to neighbours, and naive Bays.
Model Training: Train the selected algorithms by separating the training and test data sets the scene.
GUI Development: Create a user-friendly GUI with input elements for attribute selection and Algorithm selection.
User Input Processing: Collect user inputs from the GUI, preprocess them, and insert them into the trained ones.
Graphics: Produces a graphic to show the relationship between symptoms and prognosis diseases.
SQLite Integration: Configure a SQLite database to store user input, forecasts, and timestamps.
Deployment: Deploy the application to trained dependencies and models, then deliver in a proper manner.

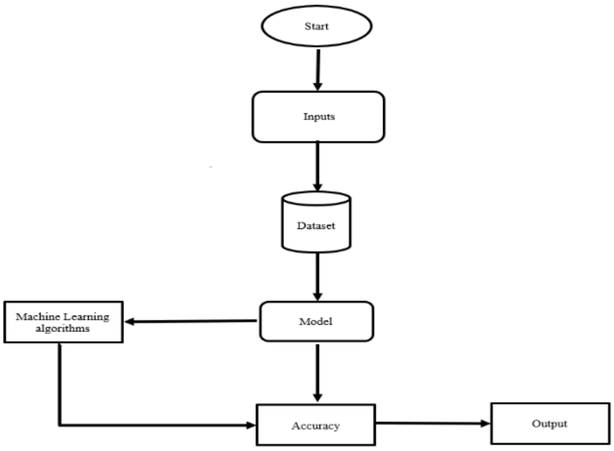


Fig:2. Flowchart

VII. RESULTS

All The application boasts a user-friendly interface, which makes characterization easier for users. It uses classifiers such as polynomial naïve Bayes, decision tree, random forest, K-Nearest Neighbour, etc. to ensure a wide range of disease predictions Users can choose from pre-defined representations, to free them up to accurately capture their experiences. The system evaluates prediction accuracy by comparing predicted diseases with actual scores from a test dataset. Once symptoms are selected and predicted, the system quickly displays the predicted diagnosis or notifies users if no diagnosis is made.

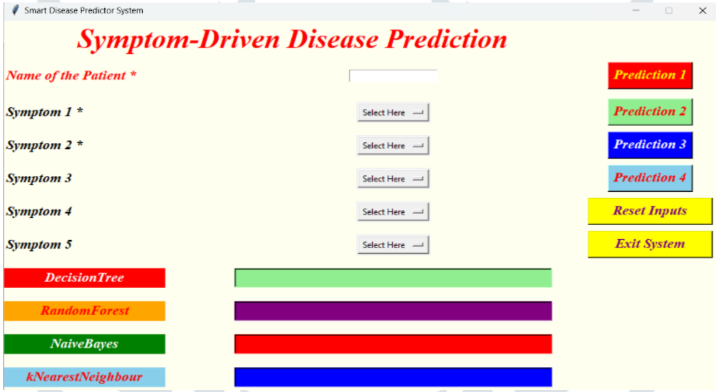


Fig. 3: User Interface

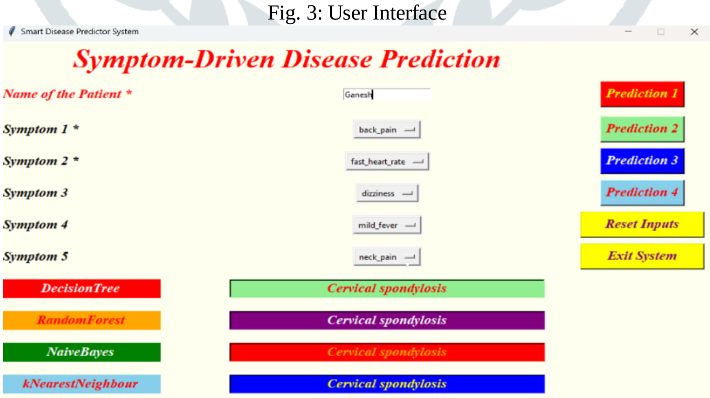
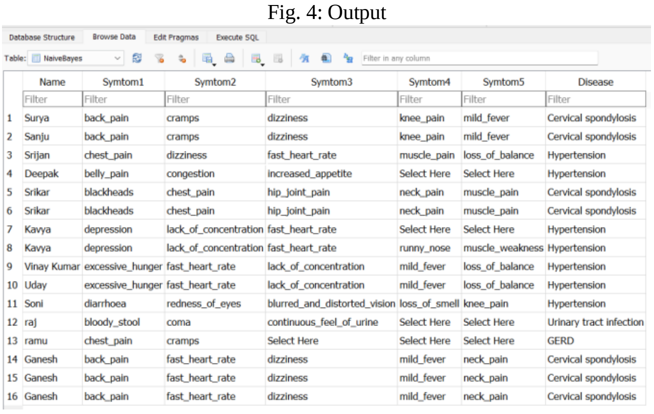


Fig. 4: Output



	Name	Symtom1	Symtom2	Symtom3	Symtom4	Symtom5	Disease
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	Surya	back_pain	cramps	dizziness	knee_pain	mild_fever	Cervical spondylosis
2	Sanju	back_pain	cramps	dizziness	knee_pain	mild_fever	Cervical spondylosis
3	Srijan	chest_pain	dizziness	fast_heart_rate	muscle_pain	loss_of_balance	Hypertension
4	Deepak	belly_pain	congestion	increased_appetite	Select Here	Select Here	Hypertension
5	Srikar	blackheads	chest_pain	hip_joint_pain	neck_pain	muscle_pain	Cervical spondylosis
6	Srikar	blackheads	chest_pain	hip_joint_pain	neck_pain	muscle_pain	Cervical spondylosis
7	Kavya	depression	lack_of_concentration	fast_heart_rate	Select Here	Select Here	Hypertension
8	Kavya	depression	lack_of_concentration	fast_heart_rate	runny_nose	muscle_weakness	Hypertension
9	Vinay Kumar	excessive_hunger	fast_heart_rate	lack_of_concentration	mild_fever	loss_of_balance	Hypertension
10	Uday	excessive_hunger	fast_heart_rate	lack_of_concentration	mild_fever	loss_of_balance	Hypertension
11	Soni	diarrhoea	redness_of_eyes	blurred_and_distorted_vision	loss_of_smell	knee_pain	Hypertension
12	raj	bloody_stool	coma	continuous_feel_of_urine	Select Here	Select Here	Urinary tract infection
13	ramu	chest_pain	cramps	Select Here	Select Here	Select Here	GERD
14	Ganesh	back_pain	fast_heart_rate	dizziness	mild_fever	neck_pain	Cervical spondylosis
15	Ganesh	back_pain	fast_heart_rate	dizziness	mild_fever	neck_pain	Cervical spondylosis
16	Ganesh	back_pain	fast_heart_rate	dizziness	mild_fever	neck_pain	Cervical spondylosis

Fig. 5: Database

VIII.CONCLUSION

In conclusion, Machine learning models of disease prediction represent a groundbreaking advance in the field Health care system. These models can inform how diseases are detected, diagnosed, and managed. By examining complex models across multiple medical contexts, these models can. valuable insights that help healthcare professionals make and deliver informed decisions Individual views. Disease prediction models have the potential to identify early warning signs and risk factors The ability to dramatically improve patient outcomes. Timely and targeted interventions for prevention Interventions can reduce morbidity and mortality. As technology continues to evolve and interdisciplinary collaboration between healthcare Entrepreneurs, data scientists and analysts explore the depth, accuracy and value of attribute-based approaches Predictive models are poised to make great strides. Finally, implementing these models. The ability to significantly reduce the burden of disease on individuals, health systems, and society Another era of holistic, dynamic, patient-centred care.

REFERENCES

Otoom et al., (2015) Effective Diagnosis and Monitoring of Heart Diseases. International Journal of Software Engineering and Its Application. 9, 143-156.

Tarmizi, N.D.A et al., (2018) Malaysia Dengue Outbreaks Detection Using Data Mining Model. Journals of Next Generation Information Technology (JNIT), 4, 96- 107.

Vembandasamy et al., (2015) Heart Disease Detection Using Naive Bayes Algorithms. IJISET-International Journal of Innovative Science, Engineering & Technology, 2, 441-444.

Gulia, A et al., (2014) Liver Patients Classification Using Intelligent Technique. (IJCSIT) International Journal of Computer Science and Information Technology, 5, 5110-5115.

Marshland, S. (2009) Machine Learnings an Algorithmic Perspectives. CRC Press, New Zealand, 6-7.

Tan et al., (2009) A Hybrid Evolutionary Algorithm for Attribute Selections in Data Mining. Journal of Expert Systems with Application,36,86168630.https://doi.org/10.1016/j.eswa.2008.10.013.

Iyer, A., Jeyalatha, S. and Sumbaly, R. (2015) Diagnosis of Diabetes Using Classification Mining Technique. International Journal of Data Mining & Knowledge Management Process (IJDKP), 5, 1-14.

Sarwar, A. and Sharma, V. (2012) Intelligent Naive Bayes Approaches to Diagnose Diabetes Type-2. Special Issues of International Journal of Computer Application (0975-8887) on Issues and Challenges in Networking, Intelligences and Computing Technologies-ICNICT 2012, 3, 14-16

Vijayarani, S. and Dhayanand, S. (2015) Liver Diseases Prediction using SVM and Naive Bayes Algorithm. International Journals of Science, Engineering and Technology Researches (IJSETR), 4, 816-820.

Fathima, A.S. and Manimeglai, D. (2016) Predictive Analysis for the Arbovirus Dengue using SVM Classifications.