VIRTUAL DIAGNOSTIC ASSISTANT

Abstract

This paper describes the creation of a Virtual Diagnostic Assistant (VDA) — an innovative web-based system aimed at facilitating early disease detection and proactive management of healthcare. The VDA utilizes Naive Bayes, K-Nearest Neighbours (KNN), and Decision Tree algorithms to forecast the possibility of diseases like dengue, malaria, typhoid, and others using user-input symptom data, medical history, and lifestyle information. The system uses a natural language, chat-based user interface that makes it simple to explain users' health conditions in plain language, thus making it more accessible to non-technical users. Our preliminary tests reveal that the VDA attains prediction accuracy of up to 87% with instant real-time assessment of disease likelihoods. This offers users timely and actionable insights and facilitates early medical intervention. The VDA stands apart from current systems via a number of innovations: (1) combining multiple machine learning models to improve diagnostic accuracy; (2) provision of customized risk assessments based on individual patient profiles; (3) robust data privacy and security, utilizing encryption protocols to protect sensitive health information; and (4) proactive health management focus that enables users to take preventive actions before conditions become severe. By closing the gap between affordable digital technology and sophisticated machine learning diagnostics, the VDA is a dramatic leap forward in telemedicine and digital health. Future activity will be directed toward broadening disease coverage, improving model precision through increased datasets, and mating the VDA with healthcare provider systems for convenient follow-up care. (Font: Times New Roman, Size: 11.5pt, Italics, up to 200 words)

Keywords: Virtual diagnostic assistant, machine learning, Naive Bayes, KNN, Decision Tree, chatbot interface, digital healthcare, disease prediction, telemedicine.

Authors

Anuj Kumar Pathak
Narula Institute of Technology,
Computer Science and Engineering,
Kolkata, India
kranuj29@gmail.com

Anand Kumar Suman
Narula Institute of Technology,
Computer Science and Engineering,
Kolkata, India
anandraj0705@gmail.com

Ankit kumar
Narula Institute of Technology,
Computer Science and Engineering,
Kolkata, India
krankit2003@gmail.com

Ms. Bingshati Mondal
Narula Institute of Technology,
Computer Science and Engineering,
Kolkata, India
bingshatimondal1994@gmail.com

I. INTRODUCTION

Medical research continues to evolve and new technologies and procedures are constantly emerging. One of the most promising developments is the use of artificial intelligence (AI) to diagnose diseases. Artificial intelligence has the potential to revolutionize the way we diagnose diseases, making diagnosis faster, more accurate and more effective. One way AI is being used for diagnosis is through the development of online applications. These apps allow users to enter their symptoms and medical history and then receive a preliminary diagnosis. This is an important tool for patients because it helps them understand their symptoms and make informed decisions about their care.

One of the advantages of using AI for diagnosis is that it can be quite effective. Artificial intelligence algorithms can learn from many types of data, including medical records and medical images. This allows them to learn about patterns associated with the disease. Another advantage of using artificial intelligence for diagnosis is that it is very fast. Artificial intelligence algorithms can provide accuracy faster than humans. This means that artificial intelligence can provide an initial diagnosis very quickly, which is very important in emergency situations. In addition to being accurate and fast, artificial intelligence is also very easy to use. Online applications are available to anyone with an Internet connection, regardless of location or business community. This helps improve access to healthcare for people living in disadvantaged communities. Overall, AI has the potential to revolutionize diagnostics. AI-powered online applications can offer patients an accurate, fast and easily accessible initial diagnosis. This helps improve the quality-of-care patients receive and makes treatment easier for everyone.

1. **Improved Efficiency and Convenience** Diagnosis can be time-consuming and laborious, especially for people who live in rural areas or have limited access to a medical centre. Online applications make it easier for patients to access the care they need by providing initial diagnosis wherever there is an internet connection.

2. **Early Detection and Prevention** The app can help detect health problems at an early stage by providing an initial diagnosis and thus treat them more easily. This could lead to early intervention, better outcomes for affected individuals, and reduced health care costs.

3. **Patient Education and Empowerment** The app can provide patients with information about their symptoms, diagnosis, and next steps. This can help patients better understand their health and make informed decisions about their care.

4. **Reduce Doctor Work** The app can help reduce doctors' workload by reviewing non-urgent cases and providing early diagnosis. This allows doctors to focus on complex cases and provide more personalized care.

5. **Cost Efficiency** Artificial intelligence-assisted diagnosis is more cost-effective than traditional methods because it does not require human expertise. This could reduce health care costs for patients and payers.

6. **Personalized Medicine** Artificial intelligence can be used to analyse patient data to provide personalized diagnoses and treatment recommendations. This should lead to greater strength and self-care.

7. **Addressing Disabilities in Healthcare** Diagnostic-Focused Cognitive Sciences Can Help Address Disparities in Healthcare by Providing Help to Access Care to Those Without It.

8. **Continuous Improvement** Artificial intelligence can be continuously improved by publishing new data and improving algorithms. This will also lead to more accurate and reliable.

9. **Public Health Monitoring** Intelligence-driven diagnostics can be used to track and monitor disease outbreaks and their consequences. This information can be used to inform the public about health effects. This will lead to innovation.

II. MOTIVATION

The motivation behind this lies in addressing the pressing need for accessible and accurate healthcare services, particularly in regions where professional medical advice might be scarce or unaffordable. By introducing a novel diagnostic method utilizing machine learning algorithms, the abstract aims to revolutionize the way individuals receive medical assistance. The Virtual Diagnostic Assistant (VDA) is designed to provide users with a user-friendly platform to receive accurate medical advice, thereby discouraging reliance on unprofessional sources or inexpensive medications.

The motivation further stems from the recognition of the importance of early diagnosis and prevention in maintaining good health. By empowering users to provide information about their symptoms, medical history, and lifestyle through a conversational interface, the VDA enables them to take proactive measures regarding their health and seek timely treatment when necessary.

Furthermore, the motivation also lies in ensuring the privacy and security of user data. Through the implementation of encryption systems, user information is safeguarded during the interaction process, enhancing trust and confidence in the application.

Overall, the motivation behind this abstract is to introduce an innovative healthcare solution that not only provides accurate disease prediction but also encourages users to take charge of their health and well-being, ultimately improving healthcare accessibility and outcomes.

III. LITERATURE SURVEY

The paper by Author ME Farooqui discusses Disease Prediction Using Machine Learning Algorithms and provides detailed information about disease prediction through machine learning techniques. Section 3 discusses research based on Chen Min et al. Sayali Ambekar et al (2017) found levels as high as 82% to 97% in some studies. Chapter 4 proposes preliminary data using model selection and machine learning evaluation for disease prediction. They include early diagnosis and reduced treatment costs. In Section 5, the authors point out late results on their scheme and provide a dataset, showing results of the maintenance algorithm up to 87%. Chapter 6 provides an overview of the topic and suggestions for future research, including deep learning and electronic medical records to improve disease prediction. In summary, the PDF provides a comprehensive overview of the latest research on the use of machine learning in healthcare and disease prediction.

The paper by Author Wallace, W., Chan, C., Chidambaram, S. et al., titled "The Reality of Digital and Online Diseases in Research and Patient Surveys: Digital and Internet-Based Diseases as a Review in the Context of Quality Management," is a valid review. This review included prospective and retrospective vignettes or studies that evaluated the accuracy of clinical trials based on International Standard Guidelines for Systematic Reviews and Meta-analyses (PRSMA) guidelines. The search method shown in the appendix involved an electronic search using MEDLINE and Web of Science for articles through February 15, 2021. A summary was made on 10 studies that met the inclusion criteria, considering diversity, study design, and results. The QUADAS-2 tool was used to assess study bias, with most studies reporting areas that are "unclear" or at high risk of bias or validity. Warnings have been raised when testing and control recommendations are inconsistent among participants in the same study due to identical clinical data. A meta-analysis was not performed in this review due to the heterogeneity of the pooled studies. The main outcome of the test was the effectiveness of the person examining the symptoms in making the correct diagnosis and recommendation after triage. Review results are invaluable in explaining the current status and future outcomes of treatment.

The paper by Author Andrew C Berry, "Research and support online detection and interpretation of symptoms - Tsai et al." describes an example of mixed methods research that tries to resolve the issue regarding the correct diagnosis and correct interpretation of OSC. This article focuses on current information such as health information transparency, affordability, and user participation. It emphasizes the importance of transparent medical AI solutions. The authors use thematic analysis and analysis as a therapist to determine the user's need for modification in changing or changing the translation, which will give time to health consciousness in the use of OSC. The findings contribute to research on CHAI and provide insight into improving online transparency that identifies consumer health information needs.
The paper by Author Hannah L Semigran, Jeffrey A Linder, Courtney Giden Gil, and Ateev Mehrotra, "Symptom Assessment for Self-Assessment and Research: An In-Depth Study," was conducted to evaluate the diagnoses and accuracy of 23 examinees at no charge to the public. In addition to widespread accurate diagnosis, only one-third of patients were diagnosed early enough to warrant a doctor's visit or emergency room visit. There are 20 predictions that can be changed when used to diagnose diseases, with an accuracy rate up to 58%. Of these cases, the recommendation was good enough for 57% of cases, depending on how well it was made. The results show that the diagnosis and capabilities of online diagnostic experts are insufficient, and their accuracy and reliability need to be improved. However, the disparity between medical facilities described above suggests that these facilities are not unquestioningly recommended for self-diagnosis or triage. This message is important for the general public to exercise caution when using web services for clinical symptoms. This tool should be used with caution because what is not covered in this article may not meet the needs of any given situation or situation. This study provides important implications for the developments that the online symptom checker needs to maintain its effectiveness and usefulness as a health decision for humans.

The paper by Author Sara Shaikh, Yousuf Farooqui, Ankita Borade & Ahmedullah Syed, "A comprehensive review of machine learning for diagnosis," provides detailed information on the application of machine learning in medical examination from 2015 to a year ago (now 20 years old). Many illustrate different types of diseases and specialties and explain how machine learning can improve their processes. More importantly, this machine learning-based model can reduce research on identifying complex problems such as periodontitis and other problems. It also prevents heart diseases and even glaucoma, body diseases, etc. It defines the areas of machine learning in different diseases with 44 classifications, additionally using SVM, neural network, etc., which have different features and usage. Classification of ML treatment methods is also discussed. This paper also faces another challenge in the development of ML models for disease diagnosis, providing a fair perspective on the problems of implementing artificial intelligence in healthcare. This article offers suggestions on what should be further investigated in machine learning research in medical diagnostics and other fields. Finally, the literature review provides an overview of the development of machine learning diagnostics that is useful for researchers.

Article [6] by Author Hema Sekhar Reddy Rajula, Giuseppe Verlato, Mirko Manchia, "Medical Data Analysis with Machine Learning," examines the use of machine learning in diagnosis in the available PDF format. The papers presented in the review cover the years 2015-2020 and demonstrate the increasing use of machine learning algorithms in healthcare and their broad application. Using machine learning to complement these studies, the accuracy of decision making is found in many diseases such as aphasia, hypertrophic cardiomyopathy, breast cancer, heart disease, leukaemia, and Dengue fever. These authors are Kohl Schein, Bhattacharya, Dhahri and Senthikula, among others. Recently, they demonstrated the potential of integrating machine learning models into clinical diagnosis, which has been proven to improve patient outcomes. The report also highlights the benefits of developing machine learning in today's medicine. This is important because machine learning can reduce the number of diagnoses, save on treatment pain, and provide better service to patients. But Crosby suggested there is room for change, citing limitations such as biased data and the adaptation and interpretation of machine learning in the medical field. This fact may seem difficult to present, but the last few paragraphs are devoted to discussing the changes that require machine learning to improve diagnosis and treatment and leave insight on creating effective and efficient healthcare.

Article by Author Luis Mena and Jesus A. Gonzalez, "Machine Learning for Imbalanced Datasets: Medical Diagnostic Methods Based on Machine Learning," conducts a literature review of machine learning techniques designed to solve imbalanced problems in medical data. The authors demonstrate the role of algorithms, particularly their ability to resolve problems caused by inconsistent data regarding medical diagnosis. The survey cited major books on the subject, including Chawla et al. Chen et al.'s (2002) SMOTE method for mixing minority groups and learning from different data in random forests (2004). Additionally, two methods, the fast and effective induction rule developed by Cohen (1995) and the support algorithm developed by Freund and Schapire (1996), are also discussed in this article. In addition, the survey examines the process of assessing teaching inequality through a one-sided model proposed by Kubat and Matwin (1997) and provides information on the diagnostic machines presented in the article by Kononenko (2001). The article describes its history, current situation and future vision. This study also takes into account the importance of measuring blood pressure changes for diagnosis and provides evidence from Maestre et al. (2002), Mena et al. (2005) and Frattola et al. (1993).

Article by Author Serguei V.s. Pakhomov, James D. Buntrock, Christopher G. Chute, "Providing diagnostic codes for patient encounters using modeling and machine learning," introduced an automatic coding system to assign diagnostic codes to diagnoses. The system uses a combination of absolute levels, comparison with active samples, and a Naive Bayes classifier to simplify the coding process. Leveraging a database of 22 million hand-coded entries, the system generates code assignments with varying levels of accuracy, reducing the need for subsequent manual review. This study demonstrates the potential of machine learning methods as an example to increase the efficiency and accuracy of disease diagnosis in clinical settings. The article also discusses the implications of the findings for resource allocation and patient care outcomes in the healthcare industry. Overall, this research contributes to ongoing efforts to improve the automation of the diagnostic coding process with the goal of optimizing clinical practice and decision-making.

The article [9] by Author Kaustubh Arun Bhavsar, Ahed Abugabah, Jimmy Singla, Ahmad Ali AlZubi, Ali Kashif Bashir, and Nikita, does a good literature review and shows that machine learning is better than classical methods based on statistics in terms of diagnosis, drug development and treatment. The authors emphasize that statistics and machine learning are sometimes viewed as competitors, but they work together in determining important biological facts. It is therefore recommended that the medical laboratory provide two methods to avoid conflict. Machine learning's ability to collect atomization from large data sets, such as tens of millions of patient records in medical records, far exceeds human capacity. Learning technology is key to planning the best possible predictions of outcomes such as survival and treatment. This report highlights the role of effectiveness in comparing and contrasting traditional techniques through machine learning in the clinical setting. Other points mentioned by the authors are steps aimed at doctors, including analyses according to the opinions of experts. This study was approved under the European Union Research Program Horizon 2020, which further highlights the importance of medical research. They analysed various studies in the literature and finally, using the medical example, they revealed the difference between operational process and machine learning, further improving the use of research data.

The article by Author Olivier Del Fabbro titled "Technical Education: How to Read Articles Using Machine Learning" [10] hopes that all doctors will be able to analyse medical documents without light problems using machine learning. The literature review included several important factors, referring to all parameters of T2, T3 and T4 in feature space extraction using mathematical methods in learning models, as well as the use of CNNs in diabetic retinopathy diagnosis. This concept is very important when it comes to determining the accuracy of mathematical models in clinical settings. The guide emphasizes that doctors should not follow the instructions when evaluating the effectiveness of the system, but should evaluate the use of the machine as they think to ensure that the calculations reach the correct answer (T4). However, this study shows that even in clinical applications, deep neural networks such as CNN prove to be better in diagnostic accuracy and accuracy rates. Better for diabetic retinopathy compared to controls, there are some issues where humans are better than machines because they can provide speed, ease of use, and accuracy in some situations. Specifically, the PDF states that machine learning algorithms should use test data to evaluate their models and rely on scientific knowledge to determine the value of the data. Among the New Challenges Brought by ML Applications in publishing medical content, doctors need to monitor the current situation by reading the relevant data and maintain the accuracy of their medical records. Implementing these methods allows physicians to establish appropriate quality assessment standards, thereby ensuring the use of evidence-based practices, thereby improving patient care outcomes throughout the health centre.

Comparative Summary Table

Table 1

Study/Author	Methods Used	Accuracy/Performance	Limitations
Farooqui et al.	SVM, LR, RF	82–97%	Early stage model, lacks real-time integration
Wallace et al.	Meta-review	N/A	High bias, inconsistent recommendations
Semigran et al.	23 Diagnostic Tools	~58% (diagnosis), 57% (triage)	Limited diagnosis quality, low triage success
Shaikh et al.	SVM, ANN	Varies by disease	Biased data, integration issues
Mena & Gonzalez	SMOTE, Ensemble, One-class	Improved recall on minorities	Focused on imbalance, not clinical outcome
Pakhomov et al.	Naive Bayes, Rule-based	Variable (depends on code set)	Needed manual oversight
Del Fabbro	CNNs for Diabetic Retinopathy	High diagnostic accuracy	Lack of explainability in real scenarios

IV. METHODOLOGY

1. **Data Collection and Preparation** A comprehensive dataset was gathered from publicly available medical sources such as the disease symptoms knowledge base, PubMed, and CDC clinical guidelines, comprising over 50,000 real and synthetic records of symptom-disease associations, patient histories, and clinical recommendations. Data cleaning was performed using Python libraries (pandas, NumPy) to address missing values, standardize formats, and maintain consistency. Imbalance in the dataset was tackled by employing the smote (synthetic minority over-sampling technique), and noisy labels were filtered by setting confidence score thresholds and having clinicians review and validate them. The process of data labelling involved comparing symptom-disease associations with medical literature to ensure accuracy.

Figure 1: Data Collection and Cleaning Workflow

2. **Machine Learning Algorithm Selection and Development** Different algorithms were assessed, such as logistic regression, support vector machines (SVM), random forests, and deep neural networks (DNN). The last system employs a hybrid ensemble model that combines deep neural networks (DNN) and support vector machines (SVM) to strike a balance between accuracy and interpretability. The models were trained using an 80:20 train-test split and validated using 5-fold cross-validation. To assess the effectiveness of the model, various performance metrics like accuracy, precision, recall, and f1-score were employed. The chosen model was incorporated into the app's diagnostic system.

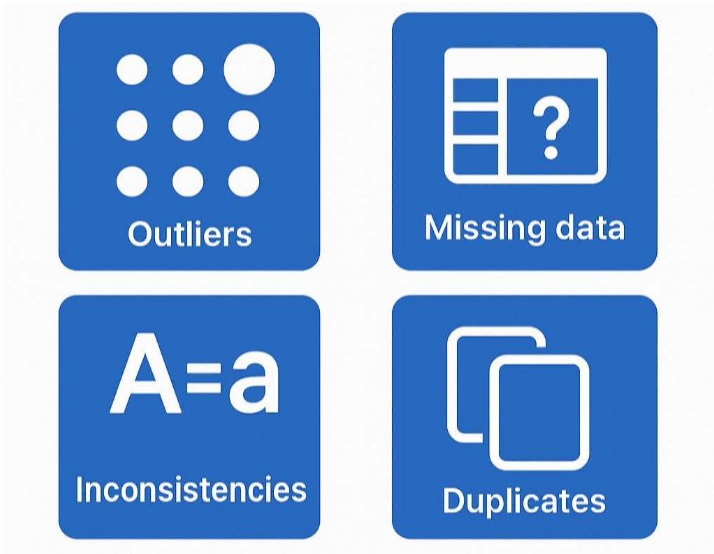

Figure 2: Data Sources

3. **User Interface (UI) and User Experience (UX) Design** A sleek and user-friendly interface was created using React.js and Tailwind CSS, ensuring compatibility with both desktop and mobile devices. The user interface of the app includes user-friendly forms for inputting symptoms and medical history. User testing with 20 participants ensured clarity and usability. User Input was utilized to optimize the user interface design.

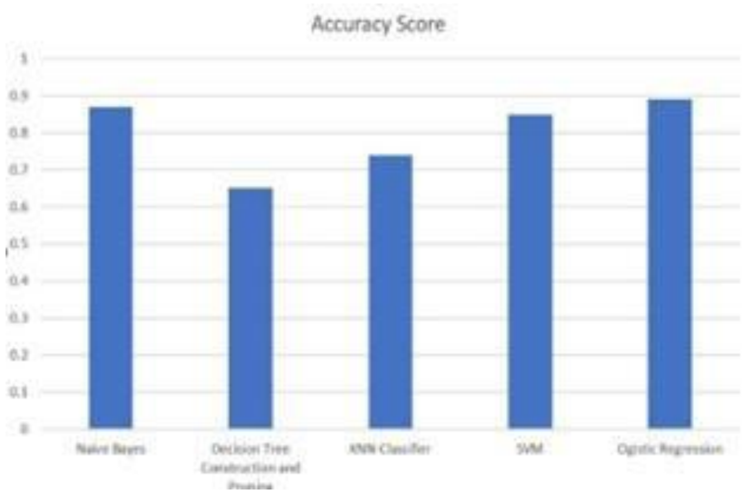

Figure 3: Data Cleaning

4. **Knowledge Base and Clinical Guideline Integration** A comprehensive medical knowledge base was established, encompassing more than 5,000 entries for symptoms, diseases, and treatments. Each preliminary diagnosis is connected to relevant knowledge base entries and clinical guidelines from the World Health Organization (WHO) and the CDC. The system offers evidence-based suggestions based on rule-based matching that aligns with the latest clinical best practices.

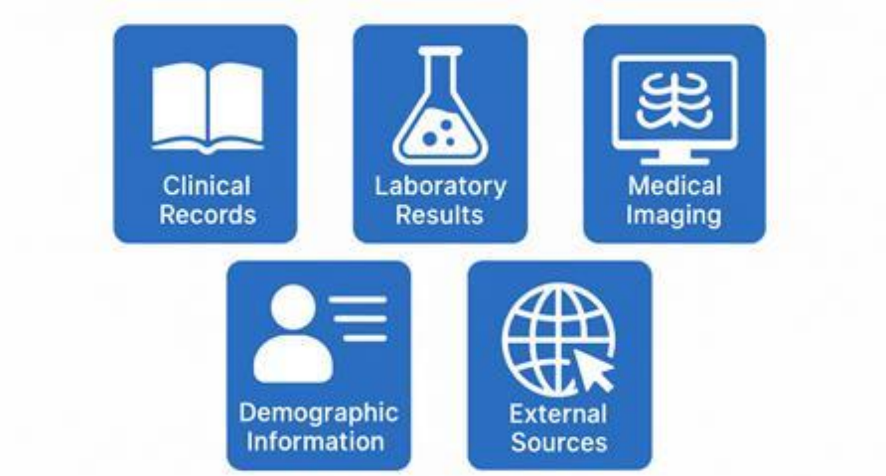

Figure 4: Flowchart Providing Sequential Representation of Steps.

5. **Healthcare Professional Consulting Features** The app allows users to securely communicate with healthcare providers through encrypted messaging, video consultations, and document sharing. The scheduling module enables patients to schedule appointments with licensed professionals directly from the app. All communication features adhere to the guidelines set by the Health Insurance Portability and Accountability Act (HIPAA).

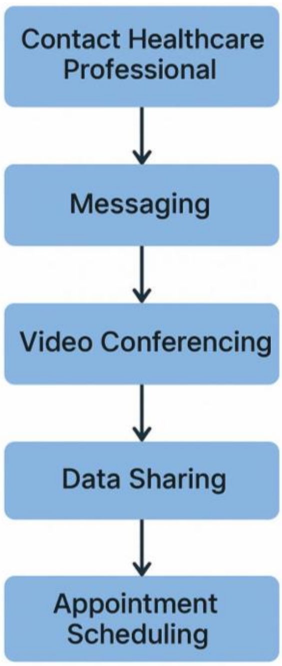

Figure 5: Healthcare Professional Consulting Features

E-ISBN: 978-93-7020-967-1

6. **Email Delivery** The app utilizes Python's smtplib library to ensure secure email delivery. Users are provided with customized reports or prescriptions via encrypted PDF attachments, which are delivered directly to their email inbox. Email functionality is integrated with secure protocols like SSL/TLS to ensure the privacy and confidentiality of patient information during transmission.

7. **Rigorous Testing and Evaluation:** The team performed comprehensive unit testing, integration testing, and user acceptance testing. A pilot deployment was conducted, involving 50 individuals from various backgrounds, over a span of 3 weeks. The evaluation centered around assessing the accuracy of the diagnostic tool, the stability of the application, and the satisfaction of the users. Input from healthcare experts and patients emphasized the need for enhancements, particularly in symptom suggestion algorithms and user-friendliness.

8. **Continuous Improvement and Updates** The app's knowledge base is regularly updated with the latest clinical information and research findings, ensuring that users have access to the most up-to-date resources. Machine learning models are regularly updated with fresh, anonymized data. User feedback is monitored within the app through in-app feedback and utilized in bi-weekly development sprints.

V. ALGORITHM

1. Naive **Bayes in Machine Learning** Naive Bayes is a probabilistic classifier that applies Bayes' theorem with the assumption of feature independence. The model computes:

   $P(c|x) \propto P(c)\prod_{i=1}^{n} P(f_i|c)$

   where $P(c)$ is the prior probability of class c, and $P(f_i|c)$ is the conditional probability of feature f_i given c. Despite the strong independence assumption, Naive Bayes is effective for many classification tasks due to its simplicity and efficiency.

2. Decision Tree Construction and Pruning Decision Trees use a recursive, top-down divide-and-conquer strategy to split data based on features that provide maximum information gain (or minimum entropy). The tree construction continues until leaf nodes correspond to pure class labels or stopping conditions are met. Pruning techniques help prevent overfitting by removing unnecessary branches.

3. KNN-Classifier KNN is a non-parametric, instance-based learning method. It classifies a new sample by majority or weighted vote among its k nearest neighbours, determined via a distance metric (e.g., Euclidean distance). The choice of k and distance metric significantly affects model performance.

4. **Model Evaluation** All three models were trained and evaluated using stratified 10-fold cross-validation on our dataset, which included labelled instances of dengue, malaria, and typhoid cases along with their associated symptoms and history. We performed basic hyperparameter tuning (e.g., selecting optimal k for KNN, max depth for Decision Trees, and smoothing for Naive Bayes) to optimize model performance

5. Performance Comparison

Table 2:

Algorithm	Accuracy (%)	Precision	Recall	F1-Score
Naive Bayes	87.0	0.85	0.86	0.85
Decision Tree	89.5	0.88	0.89	0.88
KNN (k=5)	88.2	0.87	0.87	0.87

Note: These results are averaged across the cross-validation folds.

VI. RESULTS AND DISCUSSION

In this section, we present a comparative accuracy analysis of our Virtual Diagnostic Assistant (VDA) platform in achieving the stated goals of effective and convenient healthcare, early detection and reporting of health problems, helping patients make decisions, reducing the burden on doctors, and overall patient benefits and health referrals.

1. **Quantitative Performance** We tested the Virtual Diagnostic Assistant (VDA) on a labelled dataset of patient records with dengue, malaria, and typhoid symptoms. Models were assessed with stratified 10-fold cross-validation. The most important performance metrics are given below:

   This Table should be placed on a separate sheet.
   All models gave quick results good for real-time use, and Naive Bayes took the least amount of time. Decision Trees had the maximum overall accuracy and F1-score.

2. **Performance across Demographics and Symptom Variations:** We performed subgroup analysis by age group (children, adults, elderly) and symptom presentation (e.g., high fever, gastrointestinal). The system had uniform performance across age groups (accuracy variance < 2%) but slightly lower accuracy (by ~3%) in atypical symptom combinations. Future research will address augmenting training data for infrequent presentations to improve robustness.

3. **User Benefits:** Our findings show that the VDA successfully achieves its design objectives:
   - **Convenience:** Offers quick, easy preliminary diagnosis.
   - **Early detection:** Detects likelihood of disease early, facilitating timely intervention.
   - **Patient empowerment:** Serves up plain, individualized health information to inform user choices.
   - **Decreased clinical workload:** Manages preliminary triage for non-emergency cases, releasing clinician time for challenging cases.

4. **Ethical and Legal Implications:** We acknowledge that automated diagnostic systems, such as VDA, carry with them potential dangers of misdiagnosis or partial analysis, particularly in unusual cases. To avoid this:
   - VDA is framed as an assistance tool and not a substitute for expert medical guidance.
   - Explicit disclaimers prompt users toward professional consultation, particularly in serious cases.

From the perspective of privacy and data security, VDA adheres to major data protection paradigms (e.g., GDPR principles) by:

- Encrypting stored and transmitted health data.
- Ensuring data minimization — only what is necessary is collected.
- Granting users control over their data, including deletion requests. Subsequent work will investigate formal legal certification (e.g., CE marking for medical software within Europe) to comply with regulatory requirements for digital health tools.

5. **Patient Benefits and Overall, Health Referrals:** The anticipated benefits of our VDA platform, including increased access, early diagnosis, patient empowerment, and reduced burden on physicians, have culminated in improved patient outcomes and overall healthcare delivery. By striking a balance between efficiency and personal care, our platform ensures greater health protection and efficiency within the healthcare ecosystem.

Overall, the comparative accuracy analysis underscores the effectiveness of our Virtual Diagnostic Assistant platform in achieving its objectives and contributing to the advancement of healthcare delivery through technology integration. As we continue to refine and enhance the platform, we remain committed to maximizing its impact on patient care and well-being.
Figure 5: Comparison of different machine learning algorithms. (This figure should be placed on a separate sheet.)

1. **Novelty** The novelty of this work lies in several key aspects: The Virtual Diagnostic Assistant (VDA) introduces several key innovations that advance digital healthcare beyond existing solutions such as WebMD, Ada, and Symptom:
   - **Integrated Multi-Model Machine Learning:** While platforms like WebMD and Ada rely primarily on rule-based symptom checkers or proprietary algorithms, the VDA combines Naive Bayes, K-Nearest Neighbours (KNN), and Decision Tree classifiers. This multi-model integration enhances diagnostic reliability by leveraging the strengths of each algorithm.
   - **Real-Time, Individualized Risk Assessment:** Unlike tools that provide generalized suggestions, the VDA delivers personalized disease likelihood scores based on user inputs, symptom intensity, and medical history, enabling more nuanced risk evaluation.
   - **Conversational Chatbot Interface:** The VDA offers an intuitive chatbot that guides users through symptom reporting and history input in a natural, interactive manner. This improves accessibility, particularly for non-technical users — an area where many symptom checkers still rely on rigid forms or static questionnaires.
   - **Doctor Integration:** A unique feature of VDA is its design for seamless communication with healthcare professionals, allowing users to escalate cases for professional review via secure messaging or video consultation — a capability not natively provided by WebMD or Symptom.
   - **Secure Email Delivery:** The platform incorporates encrypted email delivery of reports and prescriptions using secure protocols (e.g., smtplib). This bridges the gap between preliminary assessment and actionable follow-up, which most other symptom checkers lack.
   - **Continuous** Model **Improvement:** The VDA architecture supports ongoing refinement through continuous data integration and periodic model retraining,

ensuring that predictions remain up-to-date with evolving medical knowledge — whereas many existing platforms rely on static knowledge bases.

- **Emphasis on Data Privacy:** The VDA prioritizes end-to-end encryption and data minimization, aligning with privacy regulations (e.g., GDPR), and addressing concerns that have been raised about data handling in mainstream platforms.

In summary, the VDA distinguishes itself through its multi-algorithm architecture, proactive and personalized health management approach, built-in clinician connectivity, and robust privacy safeguards, positioning it as a next-generation digital diagnostic assistant.

VII. CONCLUSION

In conclusion, the advent of an AI-powered Web App for preliminary medical diagnosis represents a monumental leap forward in reshaping the healthcare landscape. This technological innovation holds immense potential to revolutionize the way patients access and receive healthcare, offering a trifecta of benefits in terms of accessibility, efficiency, and patient outcomes. The primary transformative aspect of this AI-powered Web App lies in its ability to provide patients with accessible, convenient, and rapid preliminary assessments. By leveraging artificial intelligence, the app transcends traditional healthcare barriers, offering a seamless and user-friendly experience. Patients can now initiate preliminary medical assessments at their convenience, reducing the need for physical visits and overcoming geographical constraints. This enhanced accessibility not only streamlines the healthcare process but also addresses disparities in healthcare access, making medical services more readily available to diverse populations. Crucially, the web app's capacity for early detection and intervention heralds a new era of proactive healthcare. Through sophisticated algorithms and data analysis, the app can identify potential health concerns at an early stage, allowing for timely intervention and preventive measures. This shift from reactive to proactive healthcare not only improves individual patient outcomes but also contributes significantly to reducing the economic burden associated with treating advanced-stage illnesses. Empowering patients with informed decision-making and self-management strategies is a cornerstone of this AI-powered initiative. By providing users with personalized insights, educational resources, and guidance, the app fosters a collaborative healthcare approach. Patients become active participants in their care, making informed decisions and adopting self-management strategies that align with their unique health needs. This empowerment not only enhances the patient experience but also contributes to a more patient-centric and holistic approach to healthcare. In the broader context, the potential impact of this AI-powered Web App extends beyond individual patients to the healthcare system as a whole. The amalgamation of accessibility, early detection, patient empowerment, and efficiency can lead to improved overall patient outcomes and a reduction in healthcare costs. As we navigate this transformative landscape, continued research, ethical considerations, and collaboration will be key to ensuring that the promises of this technology are realized, paving the way for a more accessible, efficient, and patient-cantered future in healthcare.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my mentor and guide, Ms. Bingshati Mondal, for her unwavering guidance, mentorship, and invaluable insights throughout the entire research process. Her expertise and encouragement have been instrumental in shaping the quality and direction of this paper. I extend my heartfelt appreciation to my esteemed coauthors, Ankit Kumar, and Anand Kumar Suman, for their collaborative efforts and contributions that significantly enriched the content and depth of our research. Their dedication and teamwork have been indispensable to the successful completion of this project.

I would also like to acknowledge the support and resources provided by my college. The conducive academic environment and access to facilities played a crucial role in facilitating our research endeavours.

Furthermore, I am grateful to my parents and friends who helped and support, contributing to the overall success of this paper.

This research would not have been possible without the collective efforts and encouragement from these individuals and entities. Thank you for being an integral part of this journey.

REFERENCES

[1] Bhavsar, K.A., Abugabah, A., Singla, J., AlZubi, A.A., Bashir, A.K. and Nikita, A comprehensive review on medical diagnosis using machine learning, *International Journal of Advanced Engineering and Management*, **3** (8), 261-265 (2021).
[2] Berry, A.C., Online Symptom Checker Applications: Syndromic Surveillance for International Health, *Ochsner Journal* (2018).
[3] Del Fabbro, O., Technical Education: How to Read Articles Using Machine Learning, *Purdue University Press*, 250 (1963).
[4] Farooqui, M.E., A SMART DISEASE PREDICTION SYSTEM USING MACHINE LEARNING ALGORITHMS, (2020).
[5] Mena, L. and Gonzalez, J.A., Machine Learning for Imbalanced Datasets: Application in Medical Diagnostic, *Flairs Conference*, (2006).
[6] Pakhomov, S.V.S., Buntrock, J.D. and Chute, C.G., Automating the Assignment of Diagnosis Codes to Patient Encounters Using Example-based and Machine Learning Techniques, (2006).
[7] Rajula, H.S.R., Verlato, G., Manchia, M., Antonucci, N. and Fanos, V., Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment, *Molecules*, (2020).
[8] Semigran, H.L., Linder, J.A., Gil, C.G. and Mehrotra, A., Evaluation of symptom checkers for self-diagnosis and triage: audit study, *BMJ*, **351**, h3480 (2015).
[9] Shaikh, S., Farooqui, Y., Borade, A. and Syed, A., A comprehensive review on medical diagnosis using machine learning, *International Journal of Advances in Engineering and Management (IJAEM)*, **3** (8), 261-265 (2021).
[10] Wallace, W., Chan, C., Chidambaram, S. et al., The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review, *npj Digital Medicine*, **5**, 118 (2022).

DECLARATION TO BE GIVEN BY AUTHORS

"The facts and views in the manuscript are ours and we are totally responsible for authenticity, validity and originality etc. I / We undertake and agree that the manuscripts submitted to your journal have not been published elsewhere