

A Web Based Healthcare System Using Machine Learning

M.M. Sangeetha
Dept of Computer Science and
Engineering
Kalasalingam academy of research and
education
Krishnankoil,, Virudhunagar,
Tamil Nadu, India
Sangeetha.m@klu.ac.in

N. Sri Krishna Kowshik
Dept of Computer Science and
Engineering
Kalasalingam academy of research and
education
Krishnankoil, Virudhunagar,
Tamil Nadu, India
9921004491@klu.ac.in

N. Yeshaswini
Dept of Computer Science and
Engineering
Kalasalingam academy of research and
education
Krishnankoil, Virudhunagar,
Tamil Nadu, India
99210041479@klu.ac.in

P.V. Geethika Reddy
Dept of Computer Science and
Engineering
Kalasalingam academy of research and
education
Krishnankoil, Virudhunagar,
Tamil Nadu, India
99210041613@klu.ac.in

P. Harshad
Dept of Computer Science and
Engineering
Kalasalingam academy of research and
education
Krishnankoil, Virudhunagar,
Tamil Nadu, India
99210041101@klu.ac.in

Abstract:

This study provides the development and implementation of a comprehensive web-based healthcare system that leverages machine learning to improve the diagnosis, treatment, and management of a wide range of health conditions. As the demand for efficient, accessible, and personalized healthcare continues to grow, this system is designed to meet these needs by integrating real-time patient data, historical medical records, and advanced predictive algorithms. The platform provides personalized insights, diagnostic support, and treatment recommendations, offering a new approach to healthcare delivery.

The healthcare system collects data from various sources, including electronic health records (EHRs), wearable devices, patient-reported symptoms, and real-time health monitoring tools such as IOT-enabled devices. This data serves as the foundation for machine learning algorithms, which are used to predict disease risk, assess symptoms, and recommend treatments. The algorithms employed include supervised learning for diagnosis, unsupervised learning for identifying patient clusters and personalized care patterns, and natural language processing (NLP) for interpreting unstructured medical data, such as physician notes and patient-reported symptoms.

One of the key features of this system is its prediction models, which are designed to assess disease risk based on the patient's health history and present symptoms. The system utilizes large datasets to identify patterns that may not be easily detected by human practitioners. These models allow for early detection of conditions such as cardiovascular disease, diabetes, and cancer, enabling timely intervention. Additionally, the system's symptom checker provides a real-time analysis of patient-reported symptoms and suggests a ranked list of potential diagnoses. This feature enhances patient engagement and empowers users to take control of their health while reducing the likelihood of misdiagnosis.

The system also offers personalized treatment recommendations, which are continuously updated as new patient data becomes

available. The machine learning models analyse a wide range of factors-including patient health history, lifestyle data, and current symptoms-to recommend personalized treatment plans. These recommendations can include medication dosages, lifestyle modifications, referrals to specialists, and suggestions for further diagnostic tests. By integrating this real-time feedback loop, the system ensures that patients receive the most up-to-date and relevant medical advice.

For healthcare providers, the system acts as a clinical decision support tool, offering data-driven insights to assist with diagnosis and treatment planning, physicians can access a comprehensive view of patient data, which includes historical trends, real-time health metrics, and predicted disease risks. This enables more informed decision-making, reducing the cognitive load on healthcare providers and improving the quality of care.

Another Critical component of the system is remote patient monitoring, which allows for continuous tracking of patient vitals, such as heart rate, heart disease, glucose levels, and other health metrics. By monitoring these indicators, the system can detect anomalies in real time, enabling early interventions in case of chronic conditions or acute health events. The telemedicine feature allows patients to consult with healthcare providers virtually, facilitating care for those in remote or underserved areas. The system tracks and updates symptoms and health data throughout the patient's treatment, improving continuity of care.

1.INTRODUCTION

The rapid advancement of digital technologies has revolutionized healthcare delivery, with web-based platforms and machine learning systems offering new ways to enhance patient care. This introduces an innovative web-based healthcare system that harnesses the power of machine learning to assist in disease diagnosis, brain tumour prediction, and personalized health monitoring. The system is designed to offer comprehensive health

services, from symptom-based disease prediction and brain tumour prediction using the MRI scans, and even allows patients to book medical appointments through the platform. By integrating multiple features into a single accessible platform, this system aims to improve healthcare accessibility, diagnostic accuracy, and patient engagement.

One of the primary functions of the system is its ability to predict diseases based on user-reported symptoms. Patients can input their symptoms into the system, which uses machine learning algorithms to analyze and match them against a database of known conditions. The system then provides a list of possible diseases, diet, workouts, and precautions in managing their health and potentially reduces the burden on healthcare providers by streamlining the initial diagnostic process.

Additionally, the system incorporates a brain tumor prediction module, which leverages machine learning models trained on MRI scans data. By analyzing the images, the platform can detect early signs of brain tumours, offering a fast, accurate and non-invasive diagnostic tool. This capability can be particularly beneficial in improving early detection rates and enabling timely interventions for brain cancer patients, which is critical for better outcomes and survival rates.

Furthermore, the system includes functionality for tracking and managing heart disease. User can log their daily blood pressure readings, and the system monitors trends over time, alerting patients to any significant changes or risks. This continuous tracking provides valuable insights, allowing individuals with hypertension or other cardiovascular conditions to manage their health more effectively.

To enhance patient convenience, the platform also allows users to book medical appointments directly through the website. This integration makes it easy for patients to schedule consultations with healthcare providers, ensuring timely access to medical care.

Whether they need to follow up on their symptom analysis, undergo further diagnostic testing, or consult with a specialist, the system streamlines the appointment process, contributing to a more efficient healthcare experience.

II. RELATED WORKS

This provides discuss the existing research and advancements in machine learning-based healthcare systems, focusing on symptom-based disease prediction, brain tumour prediction from MRI scans, blood pressure monitoring, and online appointment booking. The proposed system incorporates these functionalities into a web-based platform that offers accessible, accurate healthcare services to users. Below, we review the relevant literature in each of these areas, drawing from peer-reviewed research papers and existing systems.

Semigran et al. (2015) evaluated the accuracy of several online symptom checkers, finding that these systems correctly diagnosed the top condition in 34% of cases, with the correct diagnosis being included in the top 3 results 51% of the time.

Chaudry et al. (2019) developed a Bayesian network-based system that uses patient-reported symptoms and medical history to predict possible diagnosis. Their model, trained on a large dataset from the National Health Service (NHS), achieved an accuracy of 71% when predicting the correct diagnosis from a list of possible conditions.

Rathore et al. (2021) developed a Convolutional neural network (CNN) model for brain tumor classification using MRI images. The study utilized the BraTs 2020 dataset and achieved 91.67% accuracy in distinguishing between tumor types (glioma, meningioma, and pituitary tumors).

Zhu et al. (2020) introduced a hybrid machine learning approach combining logistics regression and support vector machines (SVMs) to tract blood pressure readings and predict future health risks. Their system proved effective in forecasting long-term blood pressure trends and tailoring treatment plans based on predicted outcomes.

Dantas et al. (2018) conducted a systematic review of appointment scheduling systems in healthcare, highlighting the importance of using predictive models to optimize patient flow and reduce administrative overhead. The review emphasized the use of real-time data and machine learning algorithms for improving patient and provider satisfaction.

III. SECTION BACKGROUND

a. SYMPTOM CHECKER

PROPOSED WORK:

The proposed system predicts diseases based on user-reported symptoms using the support vector Machine (SVM) model. This model is effective for multi-class classification, making it suitable for identifying various diseases from a set of input symptoms. The dataset includes symptoms, diseases, precautions, medications, diets, and workouts. The system is intended to provide both healthcare professionals and patients with a reliable tool to diagnose diseases and suggest treatments or precautions based on symptoms.

METHODOLOGY:

Data Preprocessing: The symptom-disease dataset was cleaned by handling missing values and normalizing input features to ensure consistency. The symptoms form the input features, and the disease category serves as the target variable. The data was divided into training and testing sets to evaluate the model's performance.

Model Selection: The SVM algorithm was chosen for its strong performance in classification problems, especially for medical diagnosis tasks. The one-vs-rest approach was employed to handle multi-class classification. Model evaluation metrics include accuracy, precision, recall, and F1-score.

Training and Evaluation: The SVM model was trained using the Scikit-learn library in python. The model was optimized to predict diseases based on symptoms. Cross-validation techniques were applied to ensure robust performance and minimize overfitting. A confusion matrix was used to further analyse misclassification.

Flask Web Application: A Flask-based web application was developed where users input their symptoms through an HTML form. The backend process this input to predict possible diseases using the trained SVM model. The output includes the predicted disease along with suggested treatments, precautions, medications, diets, and workouts.

Prediction: The system takes symptoms as input via the web

application and provides real-time predictions for diseases, with treatment suggestions and other related outputs being displayed on the result page.

b. BRAIN TUMOUR DETECTION

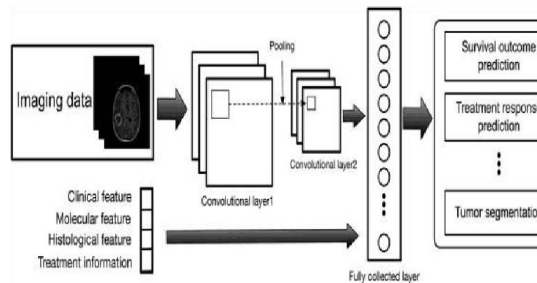
PROPOSED WORK:

The proposed system integrates a CNN-based approach for brain tumour detection. The architecture chosen for this task is VGG-16, a pretrained convolution neural network model that has remarkable accuracy in image classification tasks. The model is fine-tuned using a dataset of brain MRI images to classify the presence of a brain tumour. A Flask web application serves as the front-end interface, allowing users to upload images and view predictions.

The primary goal is to develop an accurate system for healthcare providers and patients, which can assist in early-stage detection of brain-related conditions. The system predicts whether the uploading brain MRI scan is indicative of a tumour or not.

METHODOLOGY:

Data Collection and Preprocessing: The dataset comprises brain MRI images, with image resized to 224x224 pixels to fit the input size required by the VGG-16 model. The data is augmented to prevent overfitting and to improve model robustness.

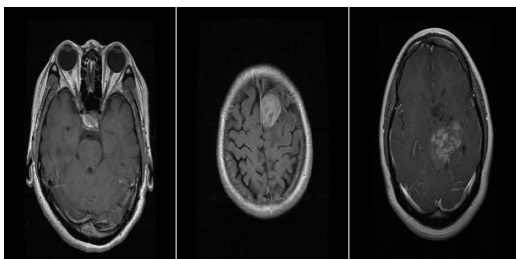


Model Architecture: The VGG_16 model is used as the core of the proposed solution. The model is pre-trained on the ImageNet dataset and fine-tuned on brain MRI image for binary classification. The final layers of the model are customized for the current classification task.

Training Process: The model is trained using a cross-entropy loss function, optimized with Adam optimizer, and evaluated using metrics such as accuracy, precision, and recall. Various hyperparameters such as learning rate, batch, size, and epochs are tuned to achieve optimal performance.

Flask Application: A flask web app is designed to provide a simple and interactive user interface. Users can upload brain MRI images through the trained model, and returns the result on the web page.

Prediction: Upon image upload, the image is preprocessed a (resized, normalized) and passed through the VGG_16 model for prediction. The output is displayed on the web page, alongside the uploaded image.



ALGORITHMS USED:

Convolutional Neural Network: CNNs are widely used for image classification tasks. VGG-16, a specific CNN architecture, is employed here. VGG-16 is known for its deep layers and small filter sizes, which are highly effective for feature extraction in images.

VGG-16: This is a deep VNN-model with 16 laVGG-16: This is a deep VNN-model with 16 layers, including convolutional layers followed by max-pooling layers, and fully connected layers at the end. It is known for its balance between depth and simplicity, making it a strong choice for classification tasks.

c. HEART DISEASE

PROPOSED WORK:

The proposed system predicts the likelihood of developing heart disease based on patient data. The logistic regression algorithm is chosen due to its robustness and interpretability in binary classification tasks. The dataset used for training the model is the Framingham Heart Study dataset, which contains medical records of individuals including features such as smoking static, cholesterol, systolic blood pressure, and glucose levels. The system is designed to provide healthcare professionalas and patients with a simple tool to assess heart disease risk.

METHODOLOGY:

Data Preprocessing: The Framingham dataset was cleaned by removing rows with missing values to ensure the quality of the data. The target variable is the 'TenYearCHD' feature, which indicates the presence or absence of heart disease risk over a 10-year period. Features used for the prediction include 'age', 'currentsmoker', 'BMI', 'sysBP', 'glucose', among others.

Model Selection: Logistic regression was chosen for this due to its effectiveness in predicting binary outcomes. The model evaluated using accuracy as the key metric.

Training and Education: The logistic regression model was trained using the scikit-learn library in python. The dataset was split into training and testing sets, and the model was trained to protect whether an individual is at risk of heart disease. Accuracy, precision, recall, and the confusion matrix were used to evaluate model performance.

Flask Web Application: A Flask-based web application was developed to serve the model predictions. Users input their health parameters (age, cholesterol, blood pressure, etc) via HTML form, and the backend processes the input to make a prediction using the logistic regression model. The prediction result, whether the individual is at risk of heart disease or not, is then displayed on the web page.

Prediction: The input features are collected via the Flask application, and the pre-trained logistic regression model is used to predict the likelihood of heart disease. The prediction is returned in real time, allowing users to get instant results based on their input datazxx

ALGORITHMS USED

Logistic Regression: Logistic regression isa statistical method that models the probability of a binary outcome. It is widely used in medical applications where the objective is to predict the occurrence of a disease based on a set of features.

C. LIVER

PROPOSED WORK:

The proposed system aims to predict liver disease using a machine learning model based on clinical parameters such as age, gender, liver enzyme levels, and more. Random Forest, an ensemble method known for its robustness and ability to handle complex datasets, is employed as the primary model. The study aims to enhance prediction accuracy through hyperparameter tuning and comparative analysis with other classifiers like Logistic Regression and support vector machines.

METHODOLOGY:

Data Preprocessing: The Indian liver patient dataset, consisting of health parameters like age, gender, bilirubin, SGPT, SGOT, and albumin levels, was used. The dataset was cleaned to handle missing values, with missing entries in the Albumin and Globulin Ratio filled using the mean value. Categorical variables such as gender were label encoded (Male = 1, Female = 0).

Model Selection: The Random Forest Classifier was selected for its effectiveness in handling complex data. The dataset was split into training (80%) and testing (20%) sets. The model was trained on the training data and evaluated using accuracy and confusion matrix metrics.

Flask Web Application: A Flask-based web app was developed where users input health parameters through an HTML form. The input is processed by the pre-trained Random Forest model to predict the likelihood of liver disease, and the result is displayed in real-time on the web page.

Prediction: Input features are collected via the web app, and the model outputs whether a patient is likely to have liver disease. The prediction is displayed to the user along with a probability score based on the clinical parameters provided.

ALGORITHMS USED:

Random Forest Classifier: Random Forest is an ensemble learning technique that constructs multiple decision trees and outputs the mode of their predictions. It is highly suitable for liver disease prediction due to its capability to manage complex data and provide accurate predictions.

IV. CONCLUSION

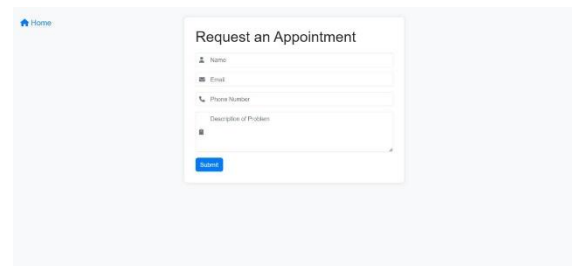
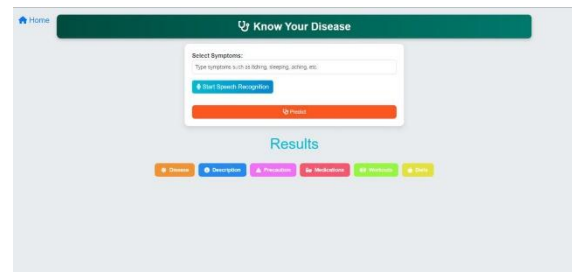
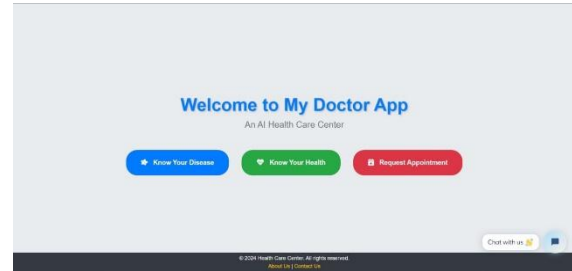
In conclusion, the proposed Flask-based web application provides a comprehensive and user-friendly platform for predicting various health conditions, integrating multiple machine learning models tailored to specific disease prediction. VGG for brain tumour detection, Logistic Regression for heart and lung disease prediction, and Random Forest for liver disease detection, this app offers a robust solution for early diagnosis and health risk assessment. The system empowers users and healthcare professionals alike by offering real-time, personalized health predictions based on input data, improving accessibility to health assessments without requiring extensive medical infrastructure.

The incorporation of the doctor appointment feature enhances the usability of the application, creating a seamless experience for patients to take action based on their health predictions. This project demonstrates the potential of machine learning in healthcare, offering cost-effective,

scalable solutions for preventive care and early detection of diseases.

Future work could involve expanding the database to include more health conditions and improving model accuracy through additional data and fine-tuning of machine learning algorithms. Furthermore, incorporating secure data handling and privacy measures would be crucial to ensure patient confidentiality in real-world applications.

OUTPUTS:



The output of the healthcare prediction system is designed to be user-friendly and informative, presenting the predicted disease or condition based on the input data in a structured format. Alongside the prediction, the system provides relevant details, such as a description of the disease,

suggested precautions, and possible medications. Additionally, the system offers personalized dietary recommendations and physical activity suggestions tailored to the predicted condition. For image-based predictions, such as brain tumour detection, visual representations of the processed medical images may also be displayed. The output ensures users receive clear, actionable insights into their health, helping them make informed decisions.

REFERENCES:

1. Health workforce: the health workforce crisis. World Health Organization. 2009. Jun 24, [2023-05-04]. <https://www.who.int/news-room/questions-and-answers/item/q-a-on-the-heal>.

2. Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Soc Sci Med*. 2015 May; 133:128–35. doi: 10.1016/j.socscimed.2015.04.004. S0277-9536(15)00224-5 - DOI - PubMed

3. Alwashmi MF. The use of digital health in the detection and management of COVID-19. *Int J Environ Res Public Health*. 2020 Apr 23;17(8):906. doi: 10.3390/ijerph17082906. <https://www.mdpi.com/resolver?pii=ijerph17082906> - DOI - PMC - PubMed

4. North F, Ward WJ, Varkey P, Tulledge-Scheitel SM. Should you search the internet for information about your acute symptom? *Telemed J E Health*. 2012 Apr;18(3):213–8. doi: 10.1089/tmj.2011.0127. - DOI - PubMed

5. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? *BMC Musculoskelet Disord*. 2016 Aug 24;17(1):362. doi: 10.1186/s12891-016-1189-2. <https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-016-1189-2> - DOI - DOI - PMC - PubMed

6. Azziz R, Carmina E, Chen Z, Dunaif A, Laven JSE, Legro RS, Lizneva D, Natterson-Horowitz B, Teede HJ, Yildiz BO. Polycystic ovary syndrome. *Nat Rev Dis Primers*. 2016 Aug 11;2:16057. doi: 10.1038/nrdp.2016.57. <https://www.nature.com/articles/nrdp201657> - DOI - PubMed

7. Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod*. 2016;31(12):2841–2855. doi: 10.1093/humrep/dew218. <https://academic.oup.com/humrep/article/31/12/2841/2730240?login=false> - DOI - PubMed

8. Carmina E, Azziz R. Diagnosis, phenotype, and prevalence of polycystic ovary syndrome. *Fertil Steril*. 2006;86(Suppl

1):S7–S8. doi:10.1016/j.fertnstert.2006.03.012. [https://linkinghub.elsevier.com/retrieve/pii/S0015-0282\(06\)00722-9](https://linkinghub.elsevier.com/retrieve/pii/S0015-0282(06)00722-9) S0015-0282(06)00722-9 - DOI - PubMed

9. Deswal R, Narwal V, Dang A, Pundir CS. The prevalence of polycystic ovary syndrome: a brief systematic review. *J Hum Reprod Sci*. 2020;13(4):261–271. doi: 10.4103/jhrs.JHRS_95_18. <https://europepmc.org/abstract/MED/33627974> JHRS-13-261 - DOI - PMC - PubMed

10. Ellis K, Munro D, Clarke J. Endometriosis is undervalued: a call to action. *Front Glob Womens Health*. 2022; 3:902371. doi:10.3389/fgwh.2022.902371. <https://europepmc.org/abstract/MED/35620300> - DOI - PMC - PubMed

11. van der Kleij RMJJ, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJJ, Teichert M, van Luenen S, Vedanthan R, Evers A, Car J, Pinnock H, Chavannes NiH. SERIES: eHealth in primary care. Part 1: concepts, conditions and challenges. *Eur J Gen Pract*. 2019;25(4):179–189. doi: 10.1080/13814788.2019.1658190. <https://europepmc.org/abstract/MED/31597502> - DOI - PMC - PubMed

12. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *J Med Internet Res*. 2010;12(1):e4. doi: 10.2196/jmir.1376. <https://www.jmir.org/2010/1/e4/> v12i1e4 - DOI - PMC - PubMed

13. Timmers T, Janssen L, Kool RB, Kremer JA. Educating patients by providing timely information using smartphone and tablet apps: systematic review. *J Med Internet Res*. 2020;22(4):e17342. doi: 10.2196/17342. <https://www.jmir.org/2020/4/e17342/> v22i4e17342 - DOI - PMC - PubMed

14. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. *J Med Internet Res*. 2015;17(2):e52. doi: 10.2196/jmir.3951. <https://www.jmir.org/2015/2/e52/> v17i2e52 - DOI - PMC - PubMed

15. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, McConnachie A, Raghuvar V, Thomson NC, Mair FS. Digital asthma self-management interventions: a systematic review. *J Med Internet Res*. 2014;16(2):e51. doi: 10.2196/jmir.2814. <https://www.jmir.org/2014/2/e51/> v16i2e51 - DOI - PMC - PubMed