Precision Health Disease Prediction System

Abhilash Gowda. N [1], Prof. K. Sharath [2]

[1] Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India

[2] Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India

ABSTRACT

The Precision Health Disease Prediction System is a web-based application that uses Artificial Intelligence to predict the chances of lung cancer, obesity, and blood pressure-related problems. It applies machine learning models such as Logistic Regression, Random Forest, and Support Vector Machine, trained on medical datasets to provide accurate results. The system is designed to help in early disease detection, provide personalized health advice, and guide patients to the right doctors. It also supports multi-disease prediction in a single step, generates professional PDF reports, tracks patient history, and stores information securely. Developed with Python, Streamlit, and SQLite, it is lightweight, user-friendly, and suitable for both urban and rural healthcare settings.

1. INTRODUCTION

Chronic diseases like lung cancer, obesity, and blood pressure disorders have become global health challenges. According to the World Health Organization, non-communicable diseases are responsible for over 70% of worldwide deaths each year. Most of these conditions are diagnosed late, which reduces treatment success and increases healthcare costs.

Lung cancer is one of the deadliest diseases because it usually shows no signs in its early stages. Once detected at advanced stages, treatment options are limited and survival chances decrease. Major risk factors include smoking, long-term exposure to pollution, and family history.

Obesity has grown into a worldwide epidemic. It is not just about being overweight—it is linked to diabetes, heart disease, joint problems, and even certain cancers. Lifestyle factors like high-calorie diets, lack of exercise, and stress contribute heavily to obesity, making it a major concern for both developed and developing nations.

Blood pressure disorders are equally critical. High blood pressure (hypertension) is often called a “silent killer” because people usually do not notice it until it leads to serious issues like stroke or heart attack. Low blood pressure (hypotension), while less common, can cause dizziness, fainting, or even shock in severe cases. Both conditions are manageable if detected early.

Traditional healthcare tools and apps usually predict only one disease, which forces patients to use multiple systems for different conditions. This process is slow and inconvenient. Our system addresses this gap by predicting multiple diseases from a single data entry and also providing guidance for preventive care.

2. RELATED WORK

Research on AI in healthcare shows strong results for **single-disease prediction** using classic machine-learning models. For lung cancer, studies using Logistic Regression and Random Forest on demographic and exposure variables (e.g., smoking history, pollution) report reliable early-risk detection; obesity work highlights SVM/Random Forest on BMI and lifestyle factors; and hypertension studies often use ensembles or boosted trees on cardiovascular markers—all of which support our choice of LR, RF, and SVM as core models.

To avoid fragmented assessments, newer efforts explore **multi-disease and ensemble frameworks**, combining task-specific models or sharing representations across related conditions. This line of work motivates our approach of running disease-specific models from a single validated input and merging the outputs into one coherent report for the user.

Clinical adoption also depends on **explainability and trust**. Studies emphasize using feature importance, SHAP, or LIME to make model behaviour transparent and to surface potential biases; pairing accuracy with clear explanations is repeatedly recommended to support clinician decision-making and patient understanding. Our reporting mirrors these practices by surfacing key drivers and performance metrics.

Applied studies consistently stress **data quality and validation**—robust preprocessing, handling missingness, and class-balance strategies—as well as automated checks to keep pipelines reliable. We align with this by enforcing input validation, normalization, and reproducible workflows before prediction and storage.

From a systems angle, prior work on **deployment and scalability** highlights containerized, reproducible environments and lightweight local datastores for low-resource or offline contexts, noting privacy–scalability trade-offs between cloud and edge. Our use of a modular design with SQLite and portable packaging follows these engineering lessons.

Finally, patient-centric research shows that **downloadable, clinician-friendly reports** with risk scores, prominent findings, and next-step guidance can improve follow-up and shared decision-making. This directly informs our PDF design and inclusion of specialist referrals to convert predictions into actionable care. Remaining gaps in the literature include limited multi-disease systems, scarce pairing of interpretable outputs with referral pathways, and few solutions that balance offline usability with scale—gaps our system targets explicitly.

3. PROBLEM STATEMENT

The rising number of chronic diseases has put tremendous pressure on healthcare systems worldwide. Early detection is often difficult, especially in regions where diagnostic facilities are limited. Existing AI-based tools are mostly designed for single-disease prediction and do not provide patients with practical next steps. Moreover, many lack secure data handling, making them unsuitable for long-term use.

Patients today need a system that does more than just tell them if they are at risk. They need clear guidance on how to stay healthy, access to trusted medical professionals, and a way to track their progress over time. This is exactly the problem our system aims to solve.

4.PROPOSED SYSTEM

The Precision Health Disease Prediction System is designed as a complete AI-driven healthcare platform that combines prediction, reporting, and patient guidance. It has five main components:

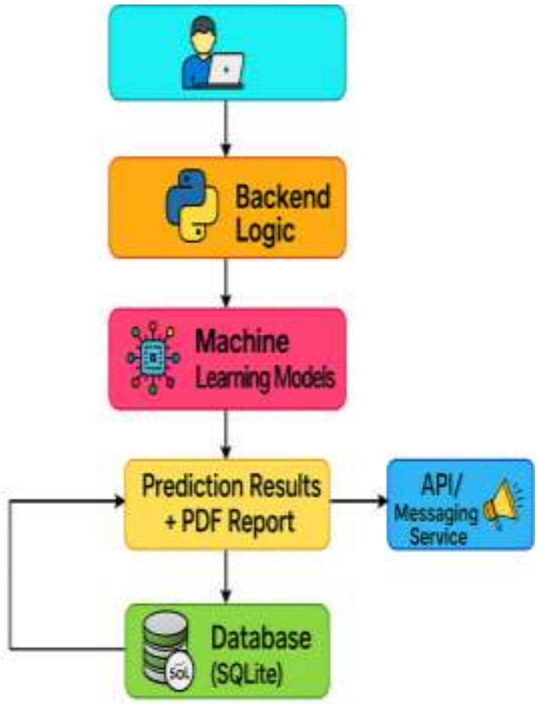
- The **user interface**, developed with Streamlit, allows easy login and data entry.
- The **application logic** manages authentication, validates inputs, and communicates with the machine learning models.
- The **machine learning models**—Logistic Regression for lung cancer, Random Forest for obesity, and SVM for blood pressure—analyze health data and produce predictions with more than 85% accuracy.
- The **database**, powered by SQLite, securely stores patient information and past reports.
- The **reporting module**, built with FPDF, generates professional PDF documents that include results, accuracy, personalized advice, and a list of recommended doctors.

One of the most powerful features of the system is its **multi-disease prediction** capability. Instead of running multiple tests separately, users can get predictions for all supported diseases at once. The system also suggests precautionary measures, such as dietary changes, exercise habits, or regular medical checkups, tailored to the user's condition.

Unlike many other health applications, this system is designed to work even in low-resource settings. Because it uses a lightweight database and can generate reports offline, it is suitable for both urban hospitals and rural clinics.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

5.METHODOLOGY



SYSTEM ARCHITECTURE

The system follows a step-by-step approach. First, users create an account and log in securely. They then enter their personal health details such as age, weight, height, BMI, and lifestyle habits. The system checks the inputs to ensure they are valid and complete. Once validated, the data is sent to the machine learning models, which process it and generate predictions.

The output is presented in two ways: on-screen for instant results, and as a downloadable PDF for future reference. The PDF report includes the prediction, accuracy percentage, confidence level, precautionary measures, and contact information of relevant specialists. It also contains the patient's details and timestamp, so it can be shared with doctors as part of a medical record.

All predictions are stored in the database, allowing users to track their health over time. This history tracking is useful for monitoring progress and planning long-term healthcare strategies.

6.RESULTS AND EVALUATION

Testing with real datasets showed that the system achieved over 85% accuracy in predicting lung cancer, obesity, and abnormal blood pressure. The prediction process was fast, with an average response time of less than three seconds. Users found the platform easy to use, and many appreciated the clarity of the reports and the inclusion of precautionary advice.

Doctors also found the reports useful, as they were professional and detailed enough to support patient consultations. The inclusion of local doctor recommendations helped patients seek immediate medical attention, reducing delays in treatment.


Health Prediction Report

Name:

Age: 1

Disease: Lung Cancer

Prediction Result: Positive

Health Tip: Quit smoking, avoid polluted areas.

Fig: Lung Cancer Prediction Report

The lung cancer prediction report is designed to be clear, professional, and easy for patients to understand. It can be used both for personal health tracking and during doctor consultations. After a user enters their details—such as smoking history, age, family medical background, and exposure to pollution—the system's machine learning model checks the data and predicts whether there is a risk of lung cancer. The report clearly shows the result as either **"Risk Detected"** or **"No Risk Detected"**, along with the accuracy level of the model. It also includes a confidence score, which helps both patients and doctors understand how reliable the prediction is.

Apart from the prediction, the report provides

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

personalized precautionary advice based on the user's profile. This may include suggestions such as quitting smoking, eating a healthier diet, exercising regularly, and scheduling routine medical check-ups. To make the report more useful, it also lists recommended cancer specialists and lung care doctors in Bangalore, with hospital names, addresses, and contact details, so that patients can easily reach out for further medical care if needed.

Health Prediction Report

Name:

Age: 1

Disease: Blood Pressure

Prediction Result: Negative

Health Tip: Reduce salt intake, monitor BP weekly

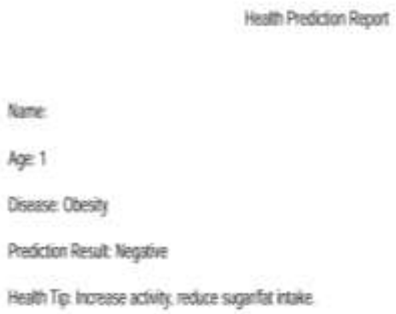
Fig: Blood Pressure Prediction Report

The blood pressure prediction report is designed to give patients a clear picture of their heart health based on the details they provide, such as age, weight, height, lifestyle habits, and any existing medical conditions. Using this information, the system's Support Vector Machine (SVM) model checks whether the user's blood pressure is normal, higher than normal (hypertension), or lower than normal (hypotension). The result is shown on the first page of the PDF as "Normal", "High Blood Pressure Detected", or "Low Blood Pressure Detected", along with the model's accuracy so that users know how reliable the result is.

The report also provides simple, personalized advice to help patients take care of themselves. For high blood pressure, it may suggest reducing salt intake, exercising regularly, keeping a healthy weight, and checking blood pressure at home. For low blood pressure, it may recommend drinking more fluids, slightly increasing salt intake under medical guidance, wearing compression stockings, and avoiding sudden posture changes. These tips are listed in a clear and easy-to-read format.

To make follow-up easier, the report includes a list of recommended cardiologists and internal medicine specialists in Bangalore, with hospital names, addresses, and contact details. The PDF also contains the patient's details, the health inputs used for prediction, and a timestamp, making it useful as part of an official medical record.

By presenting results, practical advice, and doctor contacts in one professional report, the blood pressure PDF helps patients take quick and informed steps to manage their health and reduce risks from uncontrolled blood pressure.



Health Prediction Report

Name:

Age: 1

Disease: Obesity

Prediction Result: Negative

Health Tip: Increase activity, reduce sugar/fat intake

Fig: Obesity Prediction Report

The obesity prediction report is a personalized health summary that explains a person's weight status in a simple and useful way. It analyzes details such as age, gender, height, weight, BMI, eating habits, physical activity, and lifestyle factors. Using this information, the system's Random Forest model classifies the user into one of the categories: "Normal Weight", "Overweight", or different levels of obesity (Class I, II, or III). The result is shown clearly at the top of the PDF, along with the model's accuracy so the user knows how dependable the prediction is.

The report doesn't stop at giving the result—it also provides personalized health tips. For people who are overweight or obese, it may suggest following a calorie-controlled and nutrient-rich diet, exercising daily, cutting down on sugary and fatty foods, and adding regular strength or cardio workouts. For those in the normal weight range, it includes advice on how to maintain a healthy lifestyle through balanced nutrition, regular activity, and routine health check-ups.

To support follow-up care, the report lists nutritionists, dietitians, and bariatric specialists in Bangalore, along with hospital or clinic details and contact numbers. This helps patients quickly reach out to the right professionals if they need further guidance.

Each report is neatly formatted with clear sections, bullet points, and space for doctor's notes if printed. It also includes the patient's details, the inputs used for prediction, and a timestamp, making it suitable as part of a medical record. By combining results, practical advice, and doctor recommendations, the obesity prediction report turns data into a useful health resource that encourages better lifestyle choices and informed medical decisions.

7. CONCLUSION

The Precision Health Disease Prediction System shows how Artificial Intelligence can transform healthcare by enabling early detection and preventive care. Unlike conventional tools, it provides predictions for multiple diseases in one step, stores health history securely, and offers clear next steps for patients.

Its user-friendly design ensures that it can be used by both medical professionals and individuals without technical knowledge. Because it is scalable, more diseases can be added in the future, making it a complete health companion.

By combining technology with healthcare needs, this system demonstrates how AI can help build a healthier, more proactive society. It reduces the burden on hospitals, empowers patients to take charge of their health, and makes healthcare more accessible, especially in regions with limited resources.

8.REFERENCES

[1] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA, USA: MIT Press, 2020.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[5] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[6] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. Multiple Classifier Syst.*, 2000, pp. 1–15.

[7] D. Dua and C. Graff, *UCI Machine Learning Repository*. Irvine, CA, USA: Univ. of California, School of Information and Computer Sciences, 2019.

[8] A. Esteva, B. Kuprel, R. A. Novoa, *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[9] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2001.

[10] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Waltham, MA, USA: Elsevier, 2011.

[11] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[12] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, 2015.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[15] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New Engl. J. Med.*, vol. 380, pp. 1347–1358, 2019.