

## HARNESSING MACHINE LEARNING FOR PREDICTING LIFESTYLE DISEASES

**Aqsa Mohammed Rafique Shaikh**

PG, Student

Dept. of MCA

The Oxford College of Engineering,  
Bommanahalli, Bengaluru- 560068

[aqsamca2025@gmail.com](mailto:aqsamca2025@gmail.com)

**Sowmya J**

Assistant Professor

Dept. of MCA

The Oxford College of Engineering,  
Bommanahalli, Bengaluru- 560068

[sowmyaj@theoxford.edu](mailto:sowmyaj@theoxford.edu)

### ABSTRACT

Lifestyle diseases like brain stroke, lung cancer, liver disease, and chronic kidney disease are increasing day by day. Most health conditions such as stroke, kidney problems, liver and lung diseases creep up on people so slowly that they realize too late. Prediction, as early as possible, is therefore very important. This project involves development of a computer program that will verify medical conditions of an individual with references to their age, blood pressure, and habits like smoke. The algorithm would scrub the information, extract useful entities and train machine learning models. We tested such models as decision tree, random forest, logistic regression and neural networks. Models were best combined giving the best result. The trained model is applied in a simple web application that is built with Flask. One just has to input the health information and can receive a forecast in a few seconds. The system is valid and very convenient to use by all. Tests show that this

way works well, which is good for stopping sickness before it starts. By being able to identify the risks early, the tool can assist individuals and their physicians in making wise decisions, reduce illness and contribute to improved health in all of us.

**KEYWORDS:** *The concept of forecasting such virulent diseases as brain stroke, chronic disease of kidneys, lung cancer, live cancer by using patients data. It touches data pre-processing, feature selection, and classification in order to support early diagnosis and improved healthcare decision-making.*

### INTRODUCTION

Lifestyle diseases have recently become one of the leading causes of illness and even death worldwide. Conditions like strokes, lung cancer, liver problems, and chronic kidney disease usually develop slowly and go unnoticed. When we notice warning signals in the body, it is of great essence to check them early enough so that treatment should commence on time. An example would be that

when the brain does not receive sufficient blood and oxygen, then a person would have difficulties in his/her movement, speech and thinking. This is termed stroke Lung cancer is also very risky as in the majority of them they are detected when there was spreading to other body parts. Liver and kidney diseases such as fatty liver or even cirrhosis disrupt the way our body can use energy and eliminate wastes. So, that is why individuals should be more aware of such diseases and discover them on time. Machine learning may assist physicians in the analysis of large health data and identification of the risk factors that are not readily observed. By so doing, they are able to foresee diseases in advance and warn the people so that the situation does not get serious to an extent that it can be irreversible. This project attempts to develop a system which can predict four of the major lifestyle diseases with the help of trusted datasets and intelligent algorithms. The work will include the data collection, the preparation process, training, and evaluation of the models till the most appropriate result is received. We would also like to develop a simple digital tool, which can be utilized to monitor the health regularly. The ordinary health checkups are very expensive and time-consuming; thus an online platform facilitated by machine learning can provide fast outcomes at a reduced effort. The project at hand aims at using raw health data in order to transform it

into knowledge that can be used by doctors and understood by the patients. It can facilitate a treatment at early stages and encourage people to provide their health with care. Ultimately, this technology will be able to bridge the gap between the simplicity of use and the efficacy of practice in medicine so as to become interested in healthier lives.

## **LITERATURE SURVEY**

In the health sector, health problems have been identified as a lot easier and accurate with the help of machine learning. It also assists in the process of eliminating the redundant part and only choosing the significant features of the huge volume of information in medicine. This enhances the result even more. ML has the ability to sift through patient records and to find minor indicators of chronic conditions such as chronic kidney disease. This assists physicians in identifying the condition at an early stage so that treatment of the patient can commence immediately. Various scholars have developed the application of machine learning to enhance the detection of an illness at an early stage. It also assists the hospitals to reduce errors and time wastage. In the knowledge domain of liver disease, group learning proves favorable given that it combines the effectiveness of many models to produce more thorough solutions. There are also limitations of Machine Learning It is dependent on medical records and test

results, so it does not always make perfect predictions such as with lung cancer. In stroke risk, the combination of lifestyle and medical data performs more effectively with algorithms such as decision trees, support vectors, and boosting. The knowledge of many diseases in one prediction tool enables the physician to screen or diagnose other medical conditions and provide optimal interventions. The boosting methods can also be useful in terms of locating trouble in the liver whereas some other learning techniques may not give accurate results. These measures can contribute to the development of the powerful plan of early disease prediction, with the help of which doctors can make quick and right decisions.

### **EXISTING WORK**

When people are being able to work or live their normal life, these tools can assist in the detection of diseases such as stroke, liver, kidney or lung cancer as early as the disease has some symptoms. This minimizes responding early. A majority of hospital systems and medical records merely store data, without checking to see patterns to classify health risks safely. This shows that a large portion of clinical data remains unused. It results in missed or lost opportunities in both care that can prevent the onset of the disease at an early age by detecting early indicators. The routine practices are also

time-consuming and expensive and not available in some rural areas where there are limited testing centers. As people now are going mostly for tests and doctor visits every now and then, it is hence impossible to monitor risk on a 24/7 basis. This problem demonstrates the necessity of a data-driven, guess-based system which would support faster health decisions and helps improve stop-care outcomes.

### **PROPOSED SYSTEM**

The proposed system uses a form of smart technologies like machine learning to make prediction early signs of lifestyle diseases like strokes, liver and kidney issues, and lung cancer. This setup does not just use old ways of checking health but also uses big lists of medical information that have data like age, blood pressure, the fat in the blood, if they smoke, and more signs to make good predictions. At the center of the setup is a straightforward web utility constructed with Flask. People can put in their health information, and trained smart models like Random Forests, Gradient Boosting, Machines for Support Vectors, and Brain Networks work on this info. They predictions and scores on what they are certain of, showing the numbers and how sure they are it is correct. The arrangement provides the public with fast responses, thus, making them more aware of health risks and preventing them early, so that major health diseases

occur even when they are prevented, in turn reducing the need of expensive tests at an earlier point. It gives an easy way to use by people not in hospitals. Mostly, this setup mixes easy medical know-how with clear steps to take. This makes it better at finding problems early and pushing health care ahead before problems start.

## METHODOLOGY

It is an web app that could identify an illness such as Brain Stoke, Lung Cancer, Liver Illness and Chronic Kidney Issue. It functions in clever learning manners. It begins by documenting all the important facts relating to each illnesses This is followed by the cleanup of the data and filling of gaps, removal of extras and typecasting of all data. Subsequently, this information is divided into two files. One set will be used in the training of a model to identify patterns and the other will be used to test the accuracy of the model with new cases. Such approaches as Support Vector Machine (SVM), Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN) have their approaches to evaluate the quality of a given method through measures such as correctness, exactness, recall, and F1-score.

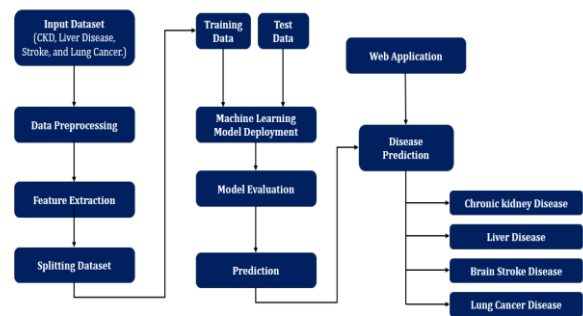


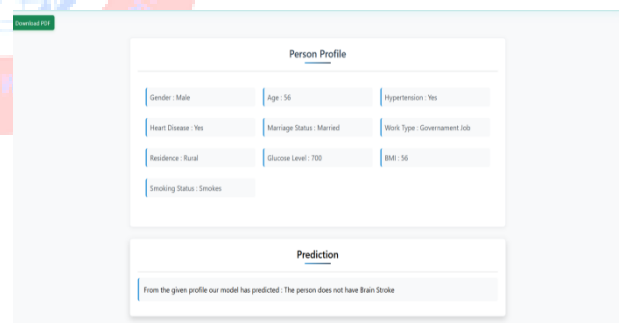
Fig. 1. System Architecture

The best-working model for each sickness is kept for future use. A Flask-based web app will be developed, and in this case, the user will have to enter the health data. Once they submit this, the app works out the best model to predict the disease and displays the results and small descriptions in a clear manner. This model ensures that the disease predictions are accurate, fast and simple to use. Machine learning algorithms are combined to obtain good prediction on diseases. Model employed was Logistic Regression because by the variable nature (yes or no) was very relevant in terms of determining whether an individual would fall sick or not. Support Vector Machine (SVM) was also applied because it performs well with hard health data and classifies properly with distinct boundaries. Random Forest and Extra trees were selected to combine numerous trees in order to improve the forecasts and reduce the errors. K-Nearest Neighbors (KNN) was used to filter the cases in the order of their similarity to previous health files and Decision Tree was used to easily acquire the calculation of the prediction and retrieve this information. Naive

Bayes was brought in for its quick and simple way of sorting based on chance, and AdaBoost helped make weak learners into a strong help. All these methods together made sure that the tool could read different data styles, test how well it works in lots of ways, and pick the best model for use on the web tool.

and the chance score to users to make things clear. To determine its performance, we tested it using numerous open databases containing varieties of amounts of patient data. We evaluated the trained AI models using measures such as rightness, exactness, how often it recalls, F1-score, and ROC-AUC. Group methods, such as Random Forests and Gradient Boosting, performed best overall, and nerve nets performed well on hard data such as lung ills. The models passed more than 90 percent right with little wrongs, and both right type and little wrong checks were true. The web application was also challenging and swift, providing readily obtainable predictions. These facts support the use of the system to get bugs out of the decisions early and are good and simple to check in the young life.

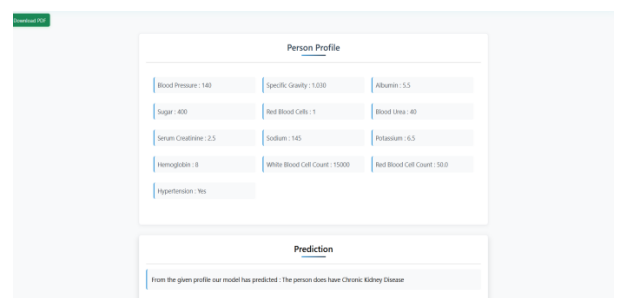
Task	Task Name	Status
1	Collect Disease Dataset	Done
2	Data Preprocessing & Cleaning	Done
3	Train Machine Learning Models	Done
4	Evaluate and Select Module	Done
5	Develop Web-based prediction System	Done



**Fig. 2. Brain Stroke Prediction Result**

## EXPERIMENTAL RESULTS

The predicting lifestyle diseases system tested with a Flask web app. In it, the users enter health information such as age, gender, flow of blood, smoking, and body signs. It also gave quick ones on the risk of brain stroke, liver disease, Chronic Kidney Disease, and lung cancer within a few minutes, some took four and five seconds after it received data. The results displayed the expected risk (ill or not)



**Fig. 3. Chronic Kidney Disease Prediction Result**

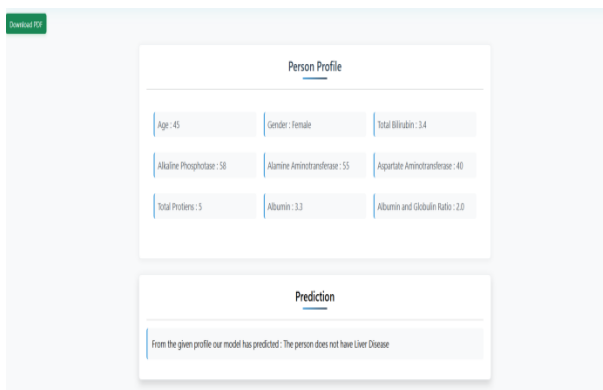


Fig. 4. Liver disease Prediction Result

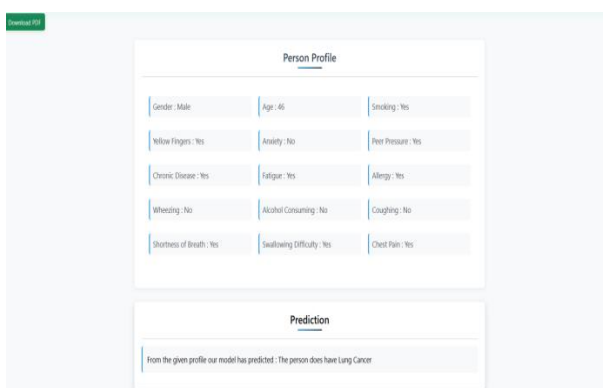


Fig. 5. Lung Cancer Prediction Result

## CONCLUSION

Disease prediction system shows how machine learning can help predicting early lifestyle diseases like stroke, liver disease, kidney disease, and lung cancer. By using medical details like age, blood pressure, and how one lives, the system gives right predictions that can help with early care and knowing soon. By adding a Flask web app, the system has an easy space where users can put in their health info and quickly get prediction results with score chances. This makes the gap between hard machine learning models and real use small, making

care before sickness easy to get. The check of the system shows that it gives sure results fast, showing its worth as a good tool to help make decisions. In the end, the system helps one take care of their health soon and cuts need for costly tests for early checks. With more work, it could become a big answer that makes care before sickness better.

## REFERENCES

- [1] Sharma, A., Soni, R., & Kapoor, D. (2021). Chronic kidney disease prediction using machine learning models. *International Journal of Computer Applications*, 176(7), 28–35. <https://doi.org/10.5120/ijca2021924561>
- [2] Raj, S., & Verma, A. (2022). Machine learning-based techniques for predicting brain stroke risk. *Journal of Healthcare Engineering*, 2022, 1–8. <https://doi.org/10.1155/2022/9708154>
- [3] Kumar, R., & Singh, D. (2021). Prediction of lung cancer through random forest and gradient boosting methods. *Procedia Computer Science*, 192, 1853–1860. <https://doi.org/10.1016/j.procs.2021.07.268>
- [4] Gupta, A., & Bansal, A. (2023). Survey of machine learning approaches for chronic disease prediction. *Journal of Medical Systems*, 47(4), 35–49. <https://doi.org/10.1007/s10916-023-02058-6>