IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 1, January 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201018 |

Early Prediction of Lifestyle Diseases

AMRUTHA B V (1), SHILPI PRASAD (2), SHREYA S(3)

Mrs.RAMYAVATHSALA C V

School of Computer Science and Engineering, Presidency University, Bangalore, Karnataka, India

Guide, Assistant Professor G1, School of Computer Science and Engineering, Presidency University, Bangalore, Karnataka, India

ABSTRACT: The prevalence of lifestyle diseases like cardiovascular ailments, skin diseases and respiratory problems has significantly increased as sedentary lifestyles and changing eating patterns occur. To manage and prevent these disorders, early detection and action are essential. This study offers a revolutionary machine learning-based method for early lifestyle illness prediction.

Three well-known classifiers—Support Vector Classifier (SVC), Gaussian Naive Bayes, and Random Forest—are combined in the suggested method. A dataset of symptoms linked to several lifestyle disorders is used to train these algorithms. The preprocessed dataset makes good use of the missing variables. A Label Encoder is used to encode the target variable, converting its categories labels into a numerical representation.
K-fold cross-validation is used to evaluate each model's performance and provide light on how well it can generalize. The project displays the models' accuracy and confusion matrices while evaluating them on training and test datasets. The mode of each classifier's unique predictions is used to arrive at the final prediction, which is the result of consensus.
The trained models are incorporated into a Flask web application along with other required elements like data dictionaries and label encoders. With the use of this application's user-friendly interface, users can input their symptoms and get forecasts for possible lifestyle diseases. In addition to helping with early disease diagnosis, the system gives users knowledge about their health conditions and is an educational tool.
This project's demonstration of machine learning and web development provides a scalable and approachable way to predict diseases early on. The program is a useful tool for boosting health awareness and preventive healthcare management because Flask is used to enable easy deployment.

KEYWORDS: Lifestyle diseases, Machine-learning based method, Prediction, Flask web application, Early detection, Prevention, Symptoms, Classifier algorithms, Health awareness, Preventive healthcare.

I. INTRODUCTION

The modern period is marked by an increase in lifestyle diseases, which include conditions like diabetes and cardiovascular disorders. Dietary changes, increased stress levels, and sedentary lifestyle choices all play a major role in the rising incidence of these illnesses. This phenomenon has its roots in the development of contemporary living, when convenience-driven lives and technology improvements have proliferated.
The World Health Organization (WHO) and other international health organizations have acknowledged lifestyle diseases as a significant public health concern because of this change. The course of these illnesses emphasizes the necessity of preventative actions, with a focus on early identification, prompt treatment, and public education. This study aligns with the larger goal of changing public health practices by utilizing
Machine learning as a tool for early prediction.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201018 |

II. LITERATURE SURVEY

1. SVM-Based Lifestyle illness Prediction (IEEE-4348, 2018): This work presents an SVM model for proactive lifestyle illness prediction. It comes from the St. Francis Institute of Technology in Mumbai. Faculty members and an undergraduate student work together to use a small dataset and the naïve bayes and random forest techniques.

2. (IEEE, 2019) Personalized Health Monitoring: This 2019 research, written by Weneen Wu and colleagues, is about individualized health monitoring using predictive analytics. Published in the IEEE Fifth International Conference on Big Data Computing Service and Applications, it highlights an interdisciplinary approach.

3. Coronary Heart Disease Prediction (ICSECS-ICOCSIM, 2021): This study, carried out by researchers from Bina Nusantara University in Indonesia, uses Random Forest and Naïve Bayes for machine-learning-based coronary heart disease prediction models. given in 2021 at the ICSECS-ICOCSIM conference.

4. Heart Disease Prediction (2021): This work, which examines machine learning methods for heart disease prediction, was written by Sunitha Guruprasad and a team from St. Joseph Engineering College. The publication contains precise methodology in depth.

5. Cardiovascular Disease Prediction (2018): In this 2018 research, the accuracy of various machine learning algorithms for cardiovascular disease prediction is compared, with a focus on supporting vector machines, naive bayes, random forests, gradient boosting, and logistic regression.
6. Early Lung Cancer Prediction (2020): This study employs machine learning methods to predict lung cancer in its early stages. It covers machine learning techniques, workflow methods, and the significance of accurate models and high-quality data.

7. Early Risk Prediction of Cervical Cancer (2022): Using algorithms like Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors, Support Vector Machine, and Multi-Layer Perceptron, this study uses machine learning techniques to achieve high accuracy in early cervical cancer prediction.
8. Liver Disease Prediction (ICCI-2021): This study, which was presented by Dr. AR. Arunachalam and C. Geetha of Dr. MGR University, examines evaluation-based methods for applying machine learning algorithms to predict liver disorders. The full publication most likely contains specific methodology details.7. Early Risk Prediction of Cervical Cancer (2022): Using algorithms like Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors, Support Vector Machine, and Multi-Layer Perceptron, this study uses machine learning techniques to achieve high accuracy in early cervical cancer prediction.

ABOUT THE PROJECT

In order to create an early lifestyle disease prediction model, a thorough lifestyle data set must be gathered, pertinent features must be extracted, a predictive model must be built using this data, its accuracy must be verified, it must be used for individual risk assessment, the model must be continuously improved, personalized recommendations based on predicted risks must be made, and the ethical use of health information must be ensured. An early lifestyle disease prediction model is used to predict health risks based on lifestyle factors. This allows for proactive interventions and individualized recommendations to stop diseases before they start or worsen, which encourages healthier living and lowers the cost of healthcare.

III. PROPOSED METHODOLOGY

Data Preparation

Source databases with associated disease classifications and symptoms are used for data collection.
Investigative Data Analysis (EDA): Examine and illustrate the properties of the dataset.
Preprocessing: Choose pertinent features, encode labels, and deal with missing data.

Model Development

Model Selection: For classification, use Random Forest, Gaussian Naive Bayes, and Support Vector Classifier (SVC).
Training: Build models using the training set after splitting the data.
Cross-Validation: Make use of k-fold cross-validation to evaluate model performance.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201018 |

Model Evaluation
Evaluation of Individual Models: Determine the accuracy and confusion matrix performance of each model.
Combined Model Evaluation: Combine forecasts from every model by means of a voting system.

Web Application
Combining models with a Flask web application to create an intuitive user experience.
User testing: Get input from users to make changes.
Deployment: For accessibility, host the Flask application on a web server.

Documentation
Code Documentation: Give the codebase comprehensive documentation.
Project Report: Compile a comprehensive report covering dataset insights, model development, and application details.

Model Preservation
Artifact Saving: Use Pickle to store label encoders, dictionaries, and trained models for later use.

IV. OBJECTIVE

Data Preparation
Source databases with associated disease classifications and symptoms are used for data collection.
Investigative Data Analysis (EDA): Examine and illustrate the properties of the dataset.
Preprocessing: Choose pertinent features, encode labels, and deal with missing data.

Model Development
Model Selection: For classification, use Random Forest, Gaussian Naive Bayes, and Support Vector Classifier (SVC).
Training: Build models using the training set after splitting the data.
Cross-Validation: Make use of k-fold cross-validation to evaluate model performance.

Model Evaluation
Evaluation of Individual Models: Determine the accuracy and confusion matrix performance of each model.
Combined Model Evaluation: Combine forecasts from every model by means of a voting system.

Web Application
Combining models with a Flask web application to create an intuitive user experience.
User testing: Get input from users to make changes.
Deployment: For accessibility, host the Flask application on a web server.

Documentation
Code Documentation: Give the codebase comprehensive documentation.
Project Report: Compile a comprehensive report covering dataset insights, model development, and application details.

Model Preservation
Artifact Saving: Use Pickle to store label encoders, dictionaries, and trained models for later use.

V. OUTCOMES

Machine Learning Model Development
The creation of machine learning models with the ability to forecast lifestyle disorders is the main objective. The models—Gaussian Naive Bayes, Random Forest, and Support Vector Classifier, among others—will be trained on an extensive dataset that includes a variety of symptoms connected to various illnesses.

Web Application Creation
The project's goal is to construct an accessible web application to go along with the model development. With the use of this program, users can input symptoms and get early forecasts. The project's possible influence on healthcare management is increased when machine learning is incorporated into a useful instrument.

Intended Illnesses
Although there is a wide range of lifestyle conditions, the study primarily focuses on predicting cardiovascular and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201018 |

diabetic issues. These illnesses were picked because they are quite common and have a wealth of symptom data that may be used to train models.

Worldwide Usability

Given that lifestyle disorders are worldwide in scope, the web application that has been built and the study's results are intended to be broadly relevant. The online application's accessibility guarantees its usage in a variety of cultural contexts and healthcare systems.

VI. CONCLUSION

To sum up, the goal of this project was to create and assess machine learning models for symptom-based lifestyle disease prediction. Support Vector Machine (SVM), Naive Bayes, and Random Forest models were used in the experimentation, and their individual and combined performances were analyzed afterwards.

The results provided significant new information about the significance of symptoms, model accuracy, and the effectiveness of an ensemble method. The suggested web application showed potential in disease prediction, providing a simple user interface for people to enter symptoms and get forecasts.

In addition, the examination of a cost-effective machine learning model as a replacement for DNA testing brought to light possible advances in the prevention of disease. This alternative model offers an effective and economical way to discover potential genetic problems that may arise from unhealthy lifestyles by using lifestyle factors.

The project's contribution goes beyond the creation of models and includes practical uses in illness prevention and health monitoring. Predictive analytics' incorporation into healthcare systems has enormous potential for early disease identification and individualized health interventions as technology advances.

Future research might examine ways to improve the model's usability and performance, and it would be possible to evaluate the predictive analytics framework's scalability in a variety of healthcare contexts.

Overall, this project signifies a step forward in leveraging machine learning for proactive and personalized healthcare solutions.

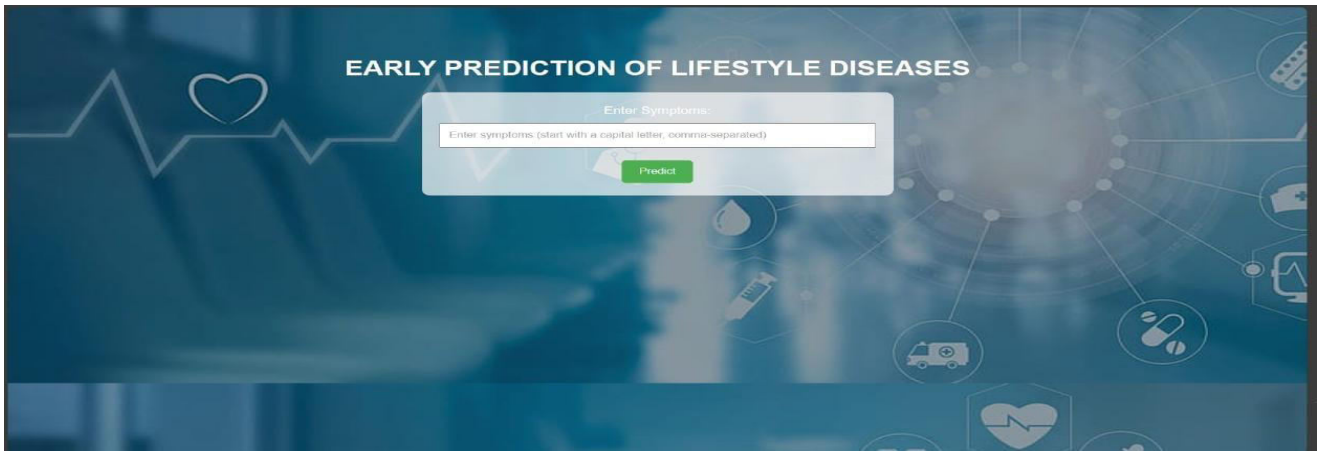
OUTPUT

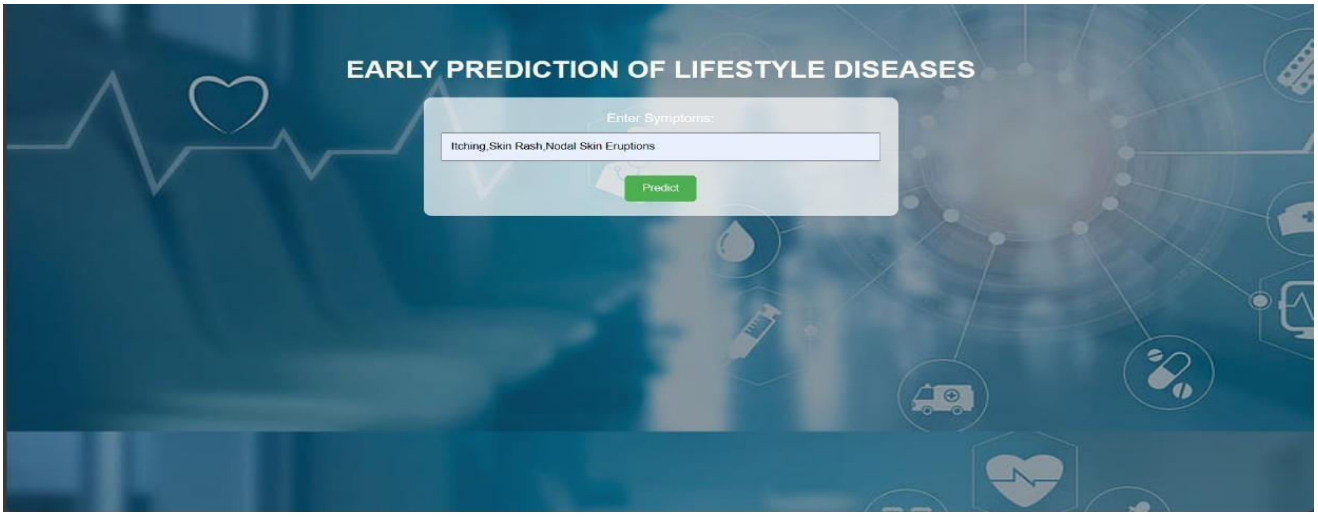
International Journal of Innovative Research in Computer and Communication Engineering

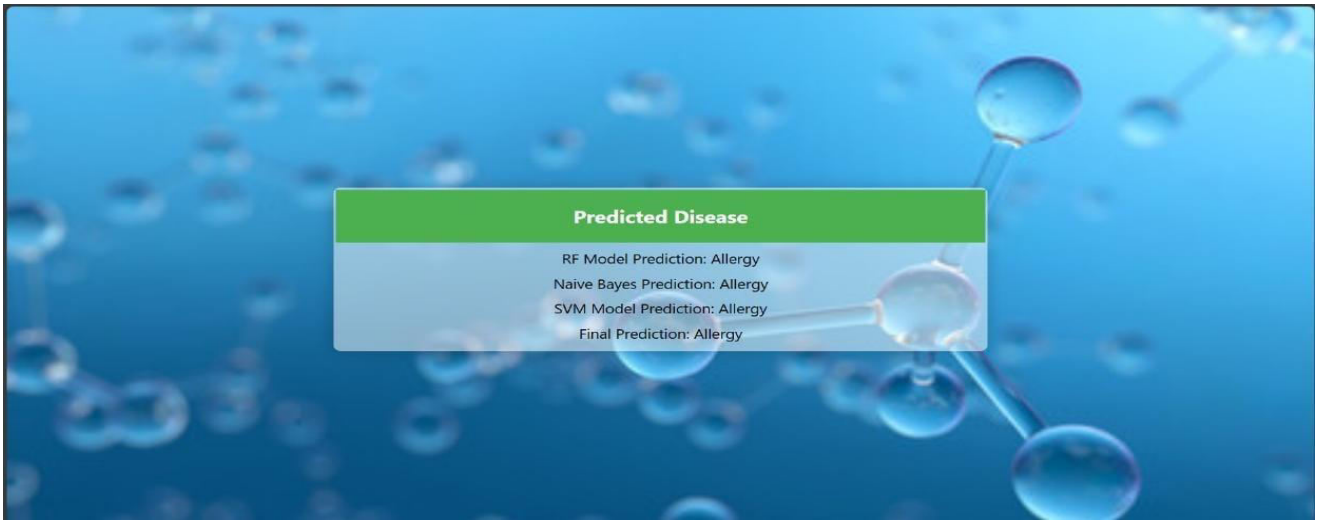
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201018 |







| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201018 |

REFERENCES

[1]Mrunmayi Patil ,Vivian Brian Lobo, Pranav Puranik, Aditi Pawaskar, Adarsh Pai, Rupesh Mishra, U.G. student , Assistant Professor, Department of Computer Engineering, St. Francis Institute of Technology, Mumbai, India.: Proposed Model for Lifestyle Disease Prediction Using Support Vector Machine.IEEE-4348(2018)

[2]Weneen Wu*,Sneha Vadakkemadathil, Poojitha Amin, Nikitha R. Anikireddypally, Suraj Kuhrana.:Personalized Health Monitoring using Predictive Analytics. 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)

[3]Charels Bernando, Eka Miranda , Mediana Aryuni, Bina Nusantara University, Indonesia,: Machine-Learning-Based Prediction Models of Coronary Heart Disease Using Naïve Bayes and Random Forest Algorithms. 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management(ICSECS-ICOCSIM)

[4]Sunitha Guruprasad, Valesh Levin Mathias , Winslet Dccunha, St.Joseph Engineering College.: Heart Disease Prediction Using Machine Learning Techniques.2021

[5]C.Geetha/Assistant Professor , Dr.AR.Arunachalam/HOD-Dean, Dr. MGR University, Bharath Institute of Higher Education and Research,Chennai.: Evaluation based Approches for Liver Disease Prediction using Machine Learning Algorithm.(ICCI-2021)

[6]Ishrak Jahan Ratul, Abdullah Al-Monsur, Bushra Tabassum, Abrar Mohammad Ar-Rafi, Mirza Muntasir Nishat and Fahim Faisal Department of Electrical and Electronic Engineering Islamic University of Technology Dhaka, Bangladesh.: Early risk prediction of cervical cancer: A machine learning approach(2022)

[7]Dr. M. Kavitha1* , G. Gnaneswar1 , R. Dinesh1 , Y. Rohith Sai1 , R. Sai Suraj1 1* Assistant Professor, 1 Student Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.: Heart Disease Prediction using Hybrid machine learning model.

[8]Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020] IEEE Xplore Part Number: CFP20M19-ART; ISBN: 978-1-7281-7089-3 .: A Study on Early Prediction of Lung Cancer Using Machine Learning Techniques

[9]Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India 978-1-5386-3702-9/18. Dinesh Kumar G , Santhosh Kumar D, Arumugaraj K , areeswari V, VIT University.: 2018 IEEE 1 Prediction of Cardiovascular Disease Using Machine Learning Algorithms.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details