EARLY PREDICTION OF LIFESTYLE DISEASES

A PROJECT REPORT

Submitted by,

Shreya S   **-20201CSD0139**
Shilpi Prasad **-20201CSD0147**
Amrutha B V **-20201CSD0148**

Under the guidance of,

Mrs. Ramyavathsala C V
Assistant Professor G1

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

At

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

JANUARY 2024

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the Project report **"EARLY PREDICTION OF LIFESTYLE DISEASES"** being submitted by "SHREYA S, SHILPI PRASAD, AMRUTHA B V" bearing roll number(s) "20201CSD0139, 20201CSD0147, 20201CSD0148" in partial fulfillment of requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering(Data Science) is a bonafide work carried out under my supervision.

Mrs. Ramyavathsala C V
ASSISTANT PROFESSOR G1
School of CSE&IS
Presidency University

Dr. A. Jayachandran
PROFESSOR & HoD
School of CSE&IS
Presidency University

Dr. C. KALAIARASAN
Associate Dean
School of CSE&IS
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE&IS
Presidency University

Dr. SAMEERUDDIN KHAN
Dean
School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **EARLY PREDICTION OF LIFESTYLE DISEASES** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering (Data Science)**, is a record of our own investigations carried under the guidance of **Mrs. Ramyavathsala C V, ASSISTANT PROFESSOR G1, School of Computer Science Engineering, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Shreya S	**20201CSD0139**
Shilpi Prasad	**20201CSD0147**
Amrutha B V	**20201CSD0148**

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.

ACKNOWLEDGEMENT

LIST OF TABLES

vii

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

1. Introduction

1.1 Context

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.

1.2 Importance of the Study

Globally, lifestyle diseases are a significant burden on public health systems. This study's importance stems from its dedication to tackling this load in two main ways:

1.2.1 Effect on Public Health

The study's emphasis on early prediction intends to reduce the burden on global healthcare infrastructures by promoting more efficient preventive interventions. Prompt measures, guided by precise forecasts, provide the capacity to modify public health regulations and enhance health results.

1.2.2 Bridging awareness gap

In addition to making predictions, the study aims to be a useful teaching tool. The project's web application serves as a link between personal awareness and forecasts based on data. Giving people the knowledge to understand their health risks promotes a culture of informed decision-making and healthy living.

CHAPTER-2

LITERATURE SURVEY

Proposed Model for Lifestyle Disease Prediction Using SVM
IEEE-4348 (2018)

The paper introduces an SVM-based model for predicting lifestyle diseases, authored by a team including an undergraduate student and faculty from Mumbai's St. Francis Institute of Technology. The research focuses on utilizing SVM to offer proactive insights into potential health risks associated with an individual's lifestyle choices. Published in IEEE, it reflects a scholarly approach with collaborative efforts between students and faculty. It uses naïve bayes and random forest for prediction and the dataset contains minimum records.

Personalized Health Monitoring using Predictive Analytics
IEEE (2019)

The paper, authored by Weneen Wu and a team of researchers, presents "Personalized Health Monitoring using Predictive Analytics." Published in the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications, the research focuses on leveraging predictive analytics for personalized health monitoring. The collaboration of authors reflects a multidisciplinary approach to healthcare technology. For a comprehensive understanding, reviewing the complete paper is recommended.

Machine-Learning-Based Prediction Models of Coronary Heart Disease Using Naïve Bayes and Random Forest Algorithms
ICSECS-ICOCSIM (2021)

This research, conducted by Charels Bernando, Eka Miranda, and Mediana Aryuni from Bina Nusantara University, Indonesia, focuses on developing machine-learning-based prediction models for coronary heart disease. Utilizing Naïve Bayes and Random Forest algorithms, the authors present their work in the 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM). The paper explores the application of using naïve bayes and random forest and predicts the accuracy, f-score based on these models.

Heart Disease Prediction Using Machine Learning Techniques (2021)

The paper, authored by Sunitha Guruprasad, Valesh Levin Mathias, and Winslet Dccunha from St. Joseph Engineering College, addresses the topic of heart disease prediction using machine learning techniques. The specific techniques employed in the research are detailed in the paper. For an in-depth understanding of the methodologies and findings, it is recommended to refer to the full content of the publication.

Prediction of cardiovascular disease using machine learning algorithm (2018)

The paper explores machine learning algorithms to predict cardiovascular disease using datasets from different heart disease databases. It compares methods like Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine, and Gradient Boosting for disease prediction accuracy. It emphasizes accurate diagnosis, data preprocessing, and visualization to aid early detection in healthcare. References include various research works on machine learning for heart disease diagnosis and prediction, offering insights into existing knowledge on the topic.

A study on early prediction of Lung Cancer using machine learning techniques.
(2020)

The use of machine learning (ML) approaches for lung cancer (LC) early prediction and categorization is discussed in this research. It describes a workflow methodology that includes gathering data, getting ready, choosing features, training models, testing them, and deciding which machine learning technique is optimal. It examines several machine learning techniques for LC prediction, assesses the metrics of their performance, and emphasizes the importance of precise ML models for LC detection at an early stage. For efficient ML-based LC prediction, the study highlights the significance of feature selection, dataset splitting, preprocessing, and high-quality data. It also offers directions for future development of LC prediction models.

Early risk prediction of cervical cancer: A machine learning approach (2022)

The study uses machine learning (ML) techniques to predict the risk of cervical cancer in its early stages. Using a dataset from the UCI ML repository, it compares the performance characteristics of eleven machine learning methods. Algorithms such as Decision Tree Classifier (DTC), Random Forest Classifier (RFC), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) achieved the highest accuracy of 93.33% in predicting the risk of cervical cancer by using default settings and hyperparameter tuning. While the study highlights the potential benefits for computer-aided detection methods for cervical cancer, it also highlights the necessity of further data collection to improve healthcare systems and the requirement for comprehensive testing prior to clinical use.

Evaluation-based Approaches for Liver Disease Prediction using Machine Learning Algorithm (ICCI-2021)

The paper, presented by C. Geetha and Dr. AR. Arunachalam from Dr. MGR University, explores evaluation-based approaches for predicting liver diseases using machine learning algorithms. The specific machine learning algorithms and evaluation methodologies used are likely discussed in detail in the paper. For a comprehensive understanding of the research, it is recommended to refer to the complete content of the publication.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

Sl. No	Title	Limitation
1.	Prediction of Cardiovascular Disease using Machine Learning Algorithms (2018)	Biases in healthcare data used to forecast heart disease are absent from the research, along with a comprehensive dataset analysis. Without delving deeply into each method's advantages, it provides a cursory comparison of machine learning techniques and a cursory exploration of tailored feature strategies. For a reliable model assessment, there needs to be a greater focus on robust validation techniques like cross-validation. Furthermore, there is a conspicuous lack of debate on the usefulness of the model for healthcare practitioners in the real world, ethical issues around the privacy of medical data, responsible model use, and practical implementation obstacles.
2.	A study on Early Prediction of Lung Cancer using Machine Learning Techniques(2020)	This research study investigates the use of machine learning algorithms (MLAs) for early lung cancer prediction. For this goal, it explores feature selection, data preparation, collection, and different MLAs. Nonetheless, a number of significant shortcomings are apparent: inadequate focus on diverse, high-quality datasets; disregard for unbalanced data processing; absence of strong feature selection method comparisons; and inadequate validation procedures. Without examining

		more sophisticated approaches like deep learning or ensemble methods, the research concentrates on traditional MLAs. It also fails to address the practical implications and clinical applicability of the predictions. Enhancements in data quality, handling of imbalanced data, robust feature selection, validation approaches, and use of modern machine learning methods have the potential to greatly improve the precision and usefulness of early lung cancer prediction.
3.	Heart Disease Prediction using Hybrid machine learning model. (2021)	The research uses machine learning (ML) techniques to create a hybrid model for heart disease prediction. For prediction, it focuses on Random Forest (RF), Decision Tree (DT), and a combination of the two models. Research gaps now include a lack of investigation into hybrid machine learning techniques and a focus on new and improved models for the prediction of heart disease. Standard datasets like Cleveland, which are frequently used in ML research, are used in this work. In order to support early patient treatment decisions, it tackles a binary classification problem (0 for absence, 1 for existence of cardiac disease). It is noteworthy that it criticizes earlier research for relying too heavily on conventional machine learning techniques like deep learning and DT, RF, KNN, and neural networks without devoting

		enough time to investigating more sophisticated approaches. Furthermore, it emphasizes the more thorough examinations and improved models are required in order to forecast cardiac disease more accurately and practically.
4.	Early risk prediction of cervical cancer: A machine learning approach (2022)	This study report highlights the importance of early detection by examining the early risk prediction of cervical cancer using machine learning (ML) approaches. With data from the UCI ML repository and 11 supervised machine learning algorithms, the study first produced an accuracy of 93.33%, with the Multi-Layer Perceptron (MLP) model gaining the upper hand. With additional refinement using Grid Search Cross-Validation (GSCV), accuracy was increased to 93.33% with DTC, RFC, KNN, SVM, and MLP. The identified shortcomings include issues related to the size of the dataset, clinical validation, management of imbalanced data, interpretability of the model, ethical considerations, generalizability, and transferability. The work shows how ML can be used in healthcare, but it also emphasizes how important it is to conduct thorough validation, ethical review, and real-world testing before implementing ML in a clinical setting.
5.	Evaluation based approach for Liver Disease Prediction using	The study identifies a number of research gaps, including the requirement for larger and more

	Machine Learning Algorithms(2021)	varied datasets, a more thorough examination of model performance measures, and an explanation of the difficulties in applying these ML models in actual clinical settings. Filling up these gaps could improve the study's ML-based liver disease prediction contributions.
6.	Heart disease prediction using machine learning techniques (2021)	In order to provide early identification and assistance for preventive actions, the article focuses on the prediction of heart disease using machine learning. It emphasizes the need for more straightforward, affordable ways to evaluate the risk of heart disease resulting from contemporary lifestyle factors. Nevertheless, there are still research gaps in the areas of managing imbalanced datasets, insufficient evaluation measures beyond accuracy, and little investigation of real-time implementation issues. Furthermore, while not doing rigorous validation on a variety of datasets or populations, the paper lacks comprehensive insights on feature relevance and model interpretability. Additionally disregarded are ethical issues pertaining to the security and privacy of patient data. The study suggests ways to improve long-term predicted accuracy and usability, but it doesn't focus on longitudinal data integration or scalability inside healthcare systems. Filling in these areas could improve the model's validity, practicality.

CHAPTER-4

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.

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.

CHAPTER-5
OBJECTIVES

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 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.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

Sequence diagram: -
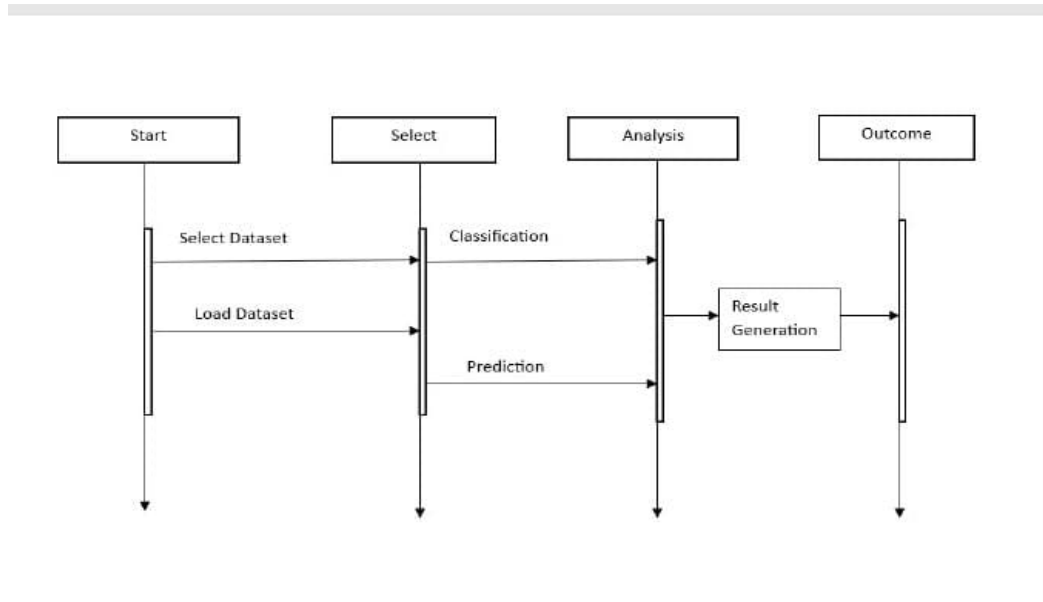
Fig 6.1

Start Sequence:

The system initiates the process when the user engages with the application.

Upon initiation, the user is prompted to select the dataset for analysis.

Select Sequence:

The user interacts with the system to choose a dataset relevant to the disease prediction task.

This selection triggers the subsequent loading of the chosen dataset.

Analysis Sequence:

The loaded dataset undergoes a comprehensive analysis phase, wherein it is preprocessed and prepared for machine learning algorithms.

Key steps in the analysis include data cleaning, feature extraction, and normalization, ensuring the dataset is suitable for training.

Outcome Sequence:

With the analysis completed, the system proceeds to the outcome phase.

The system applies classification algorithms to the prepared dataset, predicting the likelihood of diseases based on the input features.

Prediction Sequence:

During this phase, the system utilizes machine learning models for disease prediction.

The classification models make predictions based on the trained dataset, assigning

probabilities to various disease outcomes.

Result Generation Sequence:

The final stage involves the generation of results.

The system compiles the predictions into a user-friendly format, presenting the outcomes of the disease prediction task.

This sequence diagram encapsulates the intricate process involved in disease prediction, from dataset selection to result generation. The careful orchestration of these sequences ensures a seamless and effective disease prediction system.

Architecture diagram: -
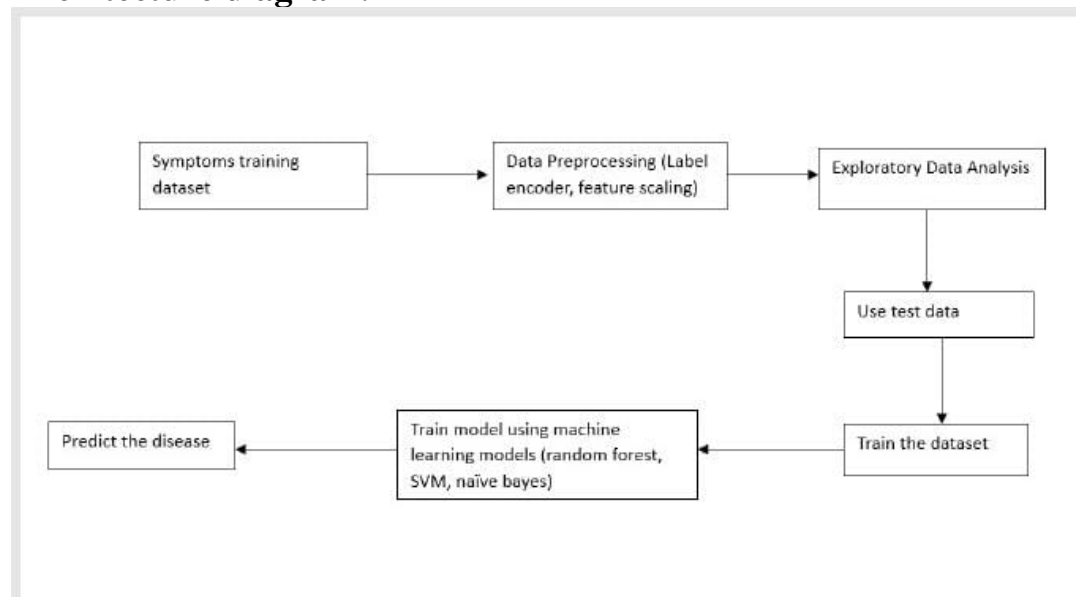
Fig 6.2

Symptoms Training Dataset:

Central to our system, this curated dataset covers diverse symptoms, forming the foundation for subsequent analysis and model training.

Data Preprocessing:

This critical step involves cleaning and structuring the dataset, ensuring uniformity and readiness for robust model training.

Exploratory Data Analysis (EDA):

Before model training, EDA uncovers patterns and insights within the dataset, guiding subsequent decisions.

Use Test Data:

Separate test data evaluates model effectiveness, providing a benchmark for predictive capabilities.

Train the Dataset:

Refined data undergoes training using machine learning models like SVM, Naive Bayes, and Random Forest.

Train Model using ML Models:

Models interpret and learn from the dataset, comprehending intricate symptom-disease relationships.

Predict the Disease:

Trained models accurately predict diseases based on input symptoms, offering actionable health insights.

This architecture seamlessly integrates data-driven methodologies, statistical insights, and machine learning, ensuring a reliable approach to disease prediction.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)

ID	Name	Sep, 2023		Oct, 2023					Nov, 2023					Dec, 2023				
		2...	24 Sep	01 Oct	08 Oct	15 Oct	22 Oct	29 Oct	05 Nov	12 Nov	19 Nov	26 Nov	03 Dec	10 Dec	17 Dec	24 Dec	31 Dec
1	Define goals and project scope		▭														
2	Gather initial requirements			▭													
3	Select the train and test datasets				▭												
4	Setup python and discuss the research papers					▭											
5	Finalize to models to be used						▭										
6	Complete backend ML model							▭									
7	Decide on frontend pages								▭								
8	Complete frontend work									▭							
9	Decide on deployment methods										▭						
10	Streamlit deployment											▭					
11	Flask Deployment												▭				
12	Complete backend and frontend													▭			

Fig 7.1

The study uses machine learning (ML) algorithms to forecast lifestyle diseases early. Phases including research, model building, developing web applications, testing, deploying, and documenting are all part of it. The Gantt chart lists tasks that must be completed within predetermined time constraints, including data analysis, algorithm implementation, web app construction, model testing, gathering user input, and final changes. The diagram acts as a visual management tool for the phases, dependencies, and advancement of the project.

CHAPTER-8
OUTCOMES

1. Model Evaluation

Utilizing cross-validation scores, compare the effectiveness of several machine learning models (SVM, Naive Bayes, Random Forest) for disease prediction.

On the test dataset, compare the SVM, Naive Bayes, and Random Forest models' accuracy.

2. Confusion Matrix Analysis

To comprehend the distribution of true positives, true negatives, false positives, and false negatives, analyze the confusion matrices for each model on the test dataset.

3. Combined Model Performance

Using the mode function, evaluate the accuracy of the composite model that combines predictions from Random Forest, Naive Bayes, and SVM.

4. Symptom Importance

Examine the significance of individual symptoms in the prognosis of an illness. Determine which symptoms have a greater impact on the model's predictions.

5. Web Application Integration

Analyze how well the web application predicts diseases using symptoms entered by the user.

Evaluate the application's usability and user experience.

6. Alternative to DNA Testing

Talk about the suggested simulated economic machine learning model as a lifestyle illness prediction tool instead of DNA testing.

Stress the possible advantages of preventing disease through lifestyle variables.

7. Training on Whole Data

Examine the effects of training the final models using both the training set (X_train, y_train) and the full dataset (X, y).

8. Potential Improvements

Make suggestions for possible changes or additions to improve the functionality or performance of the model.

9. Model Persistence

Test the trained models' durability by loading them again after they have been saved to a pickle file.

10. Predictive Analytics in Health Monitoring

Examine the consequences of utilizing predictive analytics in health monitoring and disease prediction.

Consider the developed models' scalability and possible real-world applications.

CHAPTER-9

RESULTS AND DISCUSSIONS

Fig 9.1

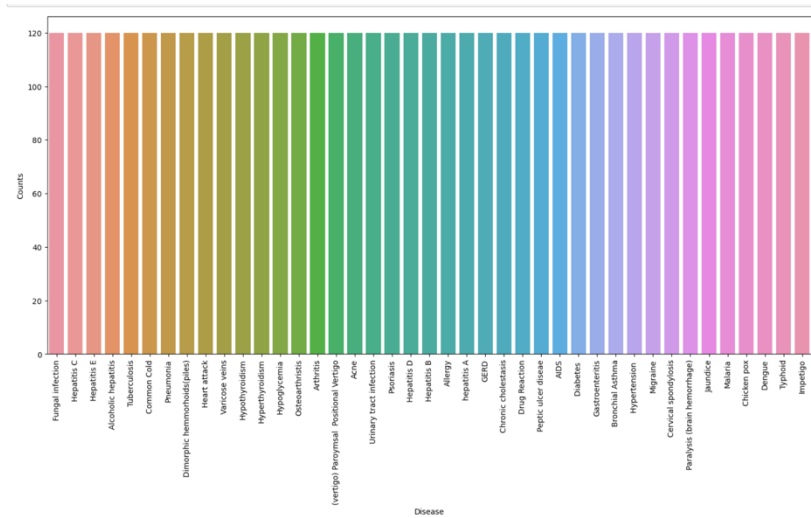It checks whether the target variable (prognosis) is balanced and visualizes the distribution using a bar plot.
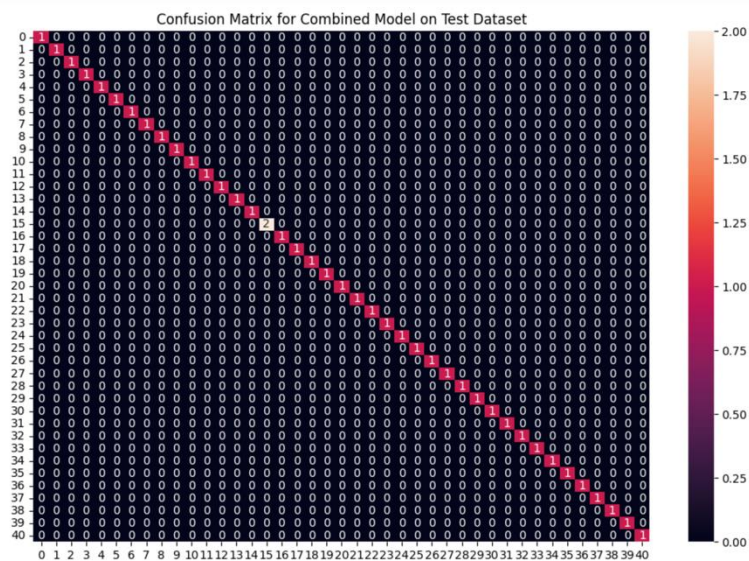

Fig 9.2

Makes predictions on the test data using SVM, Naive Bayes, and Random Forest classifiers. The final prediction is determined by taking the mode of these individual predictions.

Prints the accuracy and generates a confusion matrix for the combined model on the test dataset.

CHAPTER-10
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.

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

APPENDIX-A

PSEUDOCODE

```
define_html_file('front page html')
   create_html_tags:
      add_html_doctype_tag_with_lang_en
      create_html_head_section:
         add_meta_charset_utf_8_tag
         add_meta_viewport_tag_with_content
         add_title_tag_with_content('early prediction of lifestyle diseases')
         add_style_tag_with_body_styles:


background_image_url('https://static.vecteezy.com/system/resources/previews/000/549/219/
non_2x/abstract-background-dark-and-black-overlaps-002-vector.jpg')
            background_size_cover
            background_position_center
            font_family_arial_sans_serif
            color_white
            text_align_center
            padding_50px


      create_html_body_section:
         add_h1_tag_with_content('early prediction of lifestyle diseases')
         add_p_tag_with_content('predicting diseases early is crucial for proactive healthcare.
taking care of your health through a balanced lifestyle can significantly impact your well-
being. enter at least 3 symptoms on the next page to receive an accurate prediction.')
         add_button_tag_with_onclick_attribute('check disease')


define_html_file('home page html')
   create_html_tags:
      add_html_doctype_tag_with_lang_en
      create_html_head_section:
         add_meta_charset_utf_8_tag
         add_title_tag_with_content('early prediction of lifestyle diseases')
```

```
add_style_tag_with_body_styles:
    background_image_url('https://e0.pxfuel.com/wallpapers/128/744/desktop-
wallpaper-healthcare-background-pics.jpg')
    background_size_cover
    background_position_center
    font_family_arial_sans_serif
    color_white
    text_align_center
    padding_50px


create_html_body_section:
    add_h1_tag_with_content('early prediction of lifestyle diseases')
    create_form_tag_with_action_attr('/predict', 'post'):
        create_label_tag_with_content('enter symptoms:')
        create_input_tag_with_type_name_required_placeholder('text', 'symptoms', 'enter
symptoms (start with a capital letter, comma-separated)')
        add_br_tag
        add_button_tag_with_type_submit('predict')


define_html_file('result page html')
    create_html_tags:
        add_html_doctype_tag_with_lang_en
        create_html_head_section:
            add_meta_charset_utf_8_tag
            add_title_tag_with_content('prediction result')
            add_style_tag_with_body_styles:


background_image_url('https://t4.ftcdn.net/jpg/03/61/88/31/240_f_361883158_6gaapuidywa
yuayugl73nz5izbkiqvim.jpg')
    background_size_cover
    background_position_center
    background_repeat_no_repeat
    font_family('segoe ui', tahoma, geneva, verdana, sans-serif)
    margin_zero
```

```
            display_flex
            justify_content_center
            align_items_center
            height_100vh


    create_html_body_section:
        create_table_tag:
            create_table_row_tag:
                create_table_data_tag:
                    add_h1_tag_with_content('predicted disease')
                    add_p_tag_with_content('rf model prediction: {{
predictions.rf_model_prediction }}')
                    add_p_tag_with_content('naive bayes prediction: {{
predictions.naive_bayes_prediction }}')
                    add_p_tag_with_content('svm model prediction: {{
predictions.svm_model_prediction }}')
                    add_p_tag_with_content('final prediction: {{ predictions.final_prediction }}')


define_python_flask_app
    import_needed_libraries:
        numpy, pandas, scipy.stats, matplotlib.pyplot, seaborn, sklearn.preprocessing,
sklearn.model_selection, sklearn.svm, sklearn.naive_bayes, sklearn.ensemble,
sklearn.metrics, flask, pickle


    load_models_from_pickle_file:
        'models.pkl' file:
            final_rf_model, final_nb_model, final_svm_model, data_dict, encoder


    define_flask_app_instance
    set_app_name_as_main


    create_route_for_front_page:
        return_render_template('front_page.html')
```

```
create_route_for_home_page:
    return_render_template('home.html')


create_route_for_prediction:
    if_request_method_is_post:
        get_symptoms_from_form
        process_symptoms_and_make_predictions:
            create_input_data_for_models
            perform_predictions_with_models
            return_render_template('result.html', predictions=predictions)


run_app_with_debug_true
```

APPENDIX-B

SCREENSHOTS

Flask:

Fig 1

Fig 2
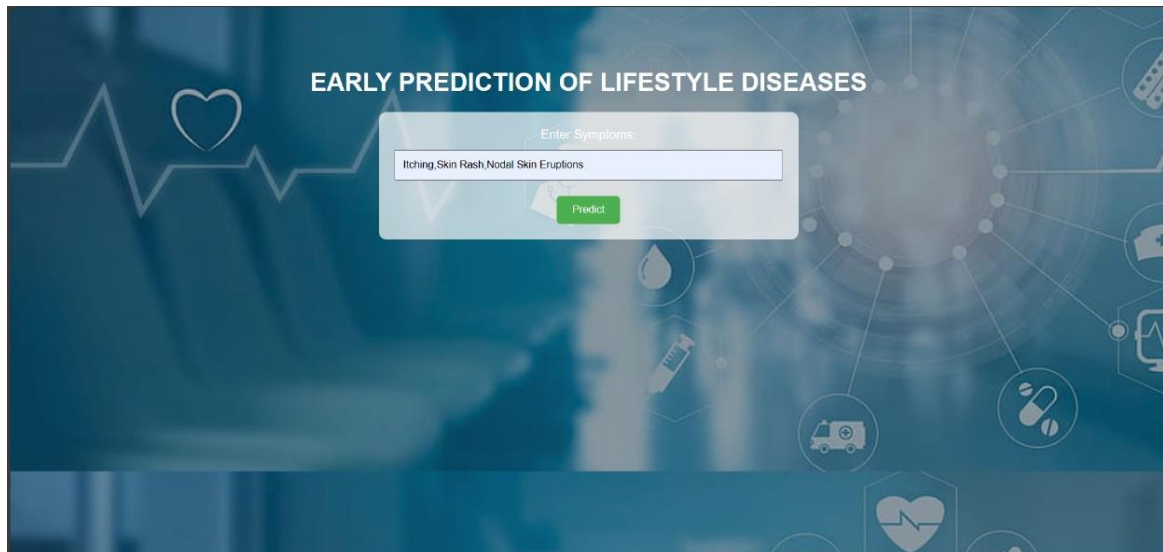
Fig 3
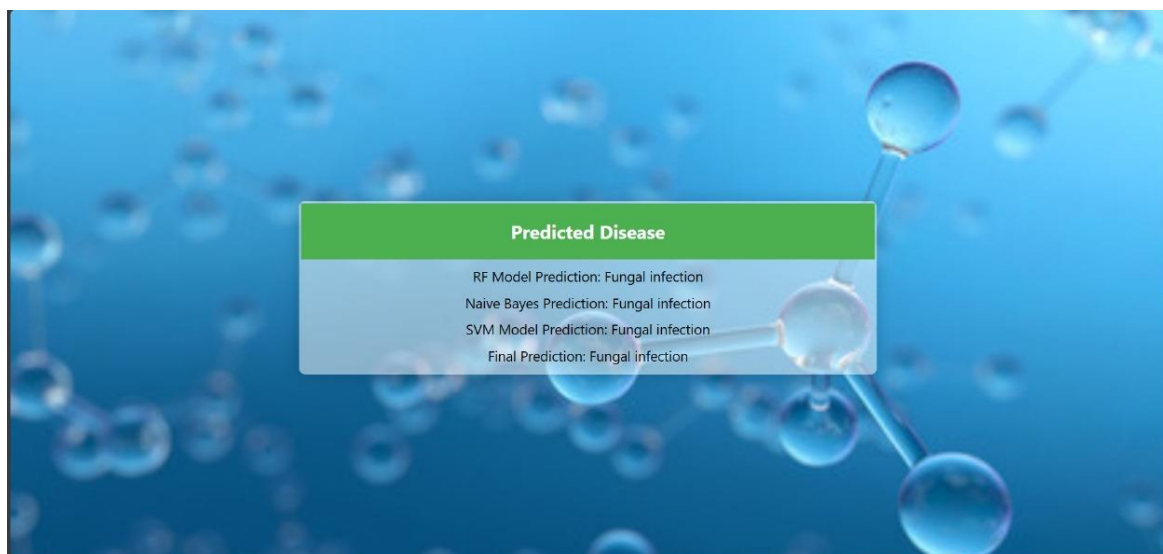
Fig 4

Streamlit:
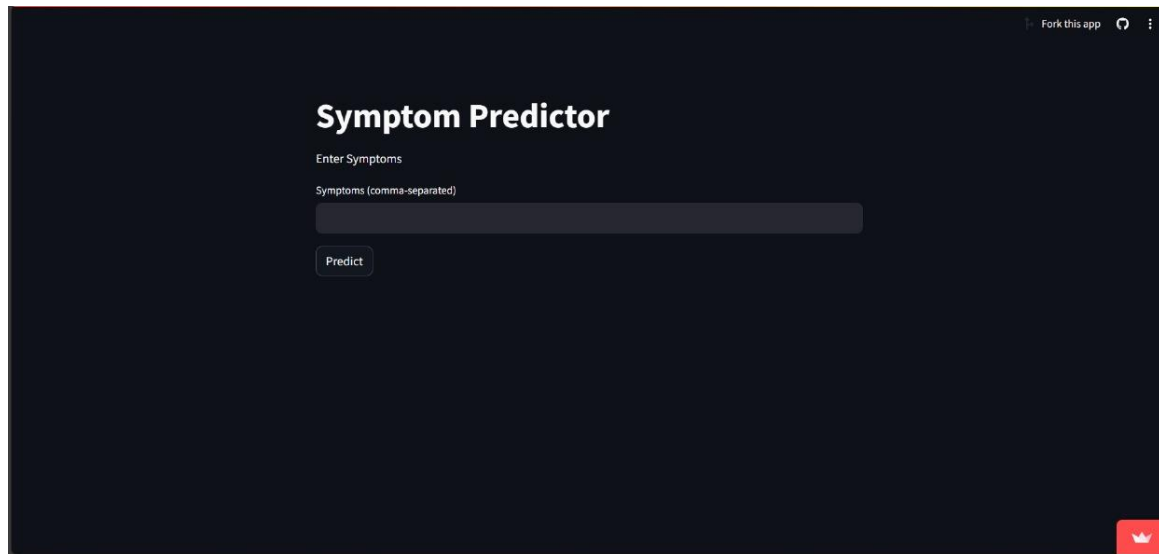
Fig 5

Fig 6

APPENDIX-C

ENCLOSURES

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, Open Access Journal Since 2013)

IJIRCCE

Impact Factor 8.379

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

SHREYA S

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

Published a paper entitled

" Early Prediction of Lifestyle Diseases"

in IJIRCCE, Volume 12, Issue 1, January 2024

Google scholar   Crossref   Mendeley   doi   INNO SPACE SJIF Scientific Journal Impact Factor

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

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

🌐 www.ijircce.com   ✉ ijircce@gmail.com

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, Open Access Journal Since 2013)

IJIRCCE

Impact Factor 8.379

CERTIFICATE
OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

SHILPI PRASAD

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

Published a paper entitled

" Early Prediction of Lifestyle Diseases"

in IJIRCCE, Volume 12, Issue 1, January 2024

Google scholar   Crossref   Mendeley   doi   INNO SPACE SJIF Scientific Journal Impact Factor

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

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

🌐 www.ijircce.com   ✉ ijircce@gmail.com
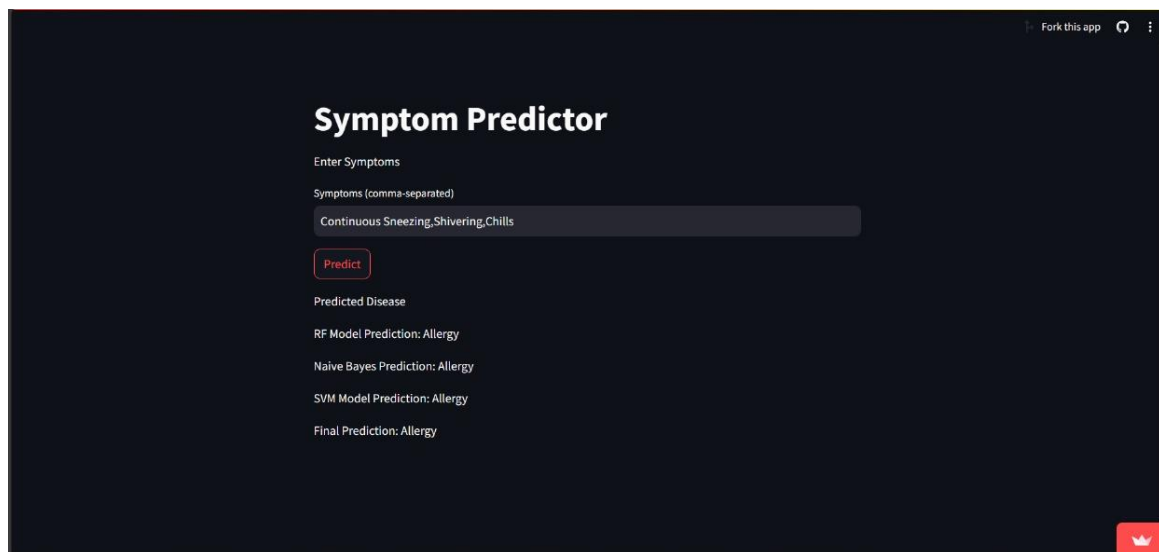
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, Open Access Journal Since 2013)

Impact Factor 8.379

CERTIFICATE
OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

AMRUTHA B V

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

Published a paper entitled

" Early Prediction of Lifestyle Diseases"

in IJIRCCE, Volume 12, Issue 1, January 2024

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

Editor-in-Chief

🌐 www.ijircce.com   ✉ ijircce@gmail.com

2. Similarity Index / Plagiarism Check report clearly showing the Percentage (%). No need of page-wise explanation.

ORIGINALITY REPORT

18%	15%	12%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	2%
2	www.mdpi.com Internet Source	2%
3	"Advances and Applications of Artificial Intelligence & Machine Learning", Springer Science and Business Media LLC, 2023 Publication	1%
4	www.coursehero.com Internet Source	1%
5	Sushree Chinmayee Patra, B. Uma Maheswari, Peeta Basa Pati. "Forecasting Coronary Heart Disease Risk with a 2-Step Hybrid Ensemble Learning Method and Forward Feature Selection Algorithm", IEEE Access, 2023 Publication	1%
6	sumc.lt Internet Source	1%
7	matrusri.edu.in	

SDG

The project work carried here is mapped to SDG-03 Good Health and Well-Being.

The project work carried here acknowledges the significance of universal access to cutting-edge technologies and preventative healthcare practices, guaranteeing that everyone on the planet has the chance for early identification and intervention. The program's reach and the efficiency of early prediction efforts in lowering the total burden of lifestyle diseases are the two main areas of focus for the indicators. The main objective is to proactively manage and mitigate the impact of these preventable diseases in order to promote global health and well-being.