# ASTHMA RISK PREDICTION USING MACHINE LEARNING

A Capstone Project Phase -2 report submitted

in partial fulfillment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE AND ENGINEERING**

By



| | |
|---|---|
| **K. MANOJ KUMAR** | **(2003A51090)** |
| **V.GAYATHRI** | **(2003A51081)** |
| **B. VINEELA** | **(2003A51229)** |
| **K. SHIVAKIRAN** | **(2003A51168)** |
| **P. PRAMOD** | **(2003A51189)** |



Under the guidance of

**Dr. R. Ravi Kumar**

Assistant Professor, School of CS&AI.

Submitted to

SR University, Ananthsagar ,Warangal , Telagnana-506371

# ASTHMA RISK PREDICTION USING MACHINE LEARNING

A Capstone Project Phase -2 report submitted
in partial fulfillment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**
in
**COMPUTER SCIENCE AND ENGINEERING**
By


| | |
|---|---|
| **K. MANOJ KUMAR** | **(2003A51090)** |
| **V.GAYATHRI** | **(2003A51081)** |
| **B. VINEELA** | **(2003A51229)** |
| **K. SHIVAKIRAN** | **(2003A51168)** |
| **P. PRAMOD** | **(2003A51189)** |

Under the guidance of
**Dr. R. Ravi Kumar**
Assistant Professor, School of CS&AI.


Submitted to

SR University, Ananthsagar,Warangal,Telagnana-506371

# SR University

Ananthasagar, Warangal.

## CERTIFICATE

This is to certify that this project entitled **"ASTHMA RISK PREDICTION USING MACHINE LEARNING"** is the bonafied work carried out by **K.MANOJ KUMAR, V.GAYATHRI, B.VINEELA, K.SHIVAKIRAN, P.PRAMOD** as a Capstone Project phase-2 for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **School of Computer Science and Engineering** during the academic year 2023-2024 under our guidance and Supervision.

**Dr. R. Ravi Kumar**
Associate Professor, Department of CSE
School of CS&AI,
SR University
Anathasagar,Warangal

**Dr. M. Sheshikala,**
Professor & Head (CSE)
School of CS&AI,
SR University
Ananthasagar,Warangal

**Reviewer-1**
Name:
Designation:
Signature:

**Reviewer-2**
Name:
Designation:
Signature:

# ACKNOWLEDGEMENT

We owe an enormous debt of gratitude to our Capstone project phase-2 guide **Dr. R. Ravi Kumar, Asst. Prof.** as well as Head of the School of CS&AI, **Dr. M. Sheshikala, Professor** for guiding us from the beginning through the end of the Capstone Project Phase -2 with their intellectual advices and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We express our thanks to project Coordinators **Mr. Sallauddin Md, Asst. Prof., and** R. **Ashok Asst. Prof.** for their encouragement and support.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

# ABSTRACT

Asthma is a chronic respiratory condition affecting millions worldwide, characterized by inflammation and narrowing of the airways, leading to symptoms like wheezing, shortness of breath, and coughing. Early identification of individuals at risk of developing asthma is crucial for timely interventions and personalized management strategies. In recent years, machine learning (ML) techniques have emerged as powerful tools for predictive analytics in healthcare, offering the potential to enhance asthma risk Assessment.

Control status are integral components of healthcare, demanding sophisticated methodologies to ensure precise prognosis and effective treatment planning. Our project proposes an innovative solution that combines Support Vector Machines (SVM) and Random Forest algorithms, seamlessly integrated with a user-friendly Python Graphical User Interface (GUI) powered by TensorFlow, hosted within the VS code environment. This integrated system leverages a diverse dataset encompassing environmental, demographic, and behavioral variables, allowing for comprehensive analysis to enhance the accuracy of asthma prediction.

Asthma Control Test (ACT) is a validated questionnaire used to assess asthma control in clinical settings. It consists of five simple questions related to asthma symptoms and their impact on daily activities. In asthma risk prediction using machine learning (ML) projects, ACT scores serve as one of the input features, alongside clinical data and patient-reported outcomes, to build predictive models. These models aim to identify individuals at higher risk of exacerbations or poor asthma control, thus facilitating early intervention and personalized treatment strategies.

Beyond prediction, our system goes further by offering personalized medication recommendations tailored to individual needs, considering factors such as demographics, environmental conditions, and behavioral patterns. Additionally, it provides insights into potential side effects associated with recommended medications, empowering both individuals and healthcare providers to make informed decisions regarding asthma management. By facilitating a deeper understanding of asthma triggers and treatment options, our approach aims to optimize patient outcomes and improve overall quality of life.

# TABLE OF CONTENTS

# LIST OF FIGURES

# 1. INTRODUCTION

Asthma, a prevalent respiratory condition, presents a significant challenge to individuals and healthcare systems worldwide due to its widespread impact. To address this challenge, accurate predictive model for assessing asthma control status, categorizing cases into well-controlled or poorly controlled. This project proposes an innovative approach that harnesses the power of machine learning algorithms, specifically Support Vector Machines (SVM) and Random Forest, to accurately predict asthma control status.

Healthcare uses machine learning in new ways. Machine learning looks at big data on genes, environment, and lifestyle. It can then predict a person's asthma risk really well. This is very helpful. If we know someone's asthma risk early, we can give them special care. That might stop asthma from getting bad. Machine learning helps healthcare find better solutions for people with asthma. It's an amazing tool that could make asthma much easier to manage.

By leveraging a diverse dataset encompassing environmental, demographic, and Behavioral factors, our system aims to enhance its predictive capabilities. This comprehensive dataset allows for a thorough analysis, enabling the identification of key factors contributing to asthma onset and severity. Additionally, our system offers personalized medication recommendations tailored to individual patient profiles, taking into account various factors such as age, gender, environmental conditions, and lifestyle habits.

The integration of a user-friendly Python Graphical User Interface (GUI) powered by TensorFlow within the VS code environment ensures accessibility and usability for both patients and healthcare professionals. This intuitive interface facilitates seamless interaction with the system, allowing users to input relevant data and receive personalized recommendations and predictions. By bridging the gap between advanced machine learning techniques and practical application in healthcare settings, our system aims to empower individuals and healthcare providers to make informed decisions regarding asthma management, ultimately leading to improved patient outcomes and quality of life.

# 2. RELATED WORKS

Asthma, a continual respiration circumstance characterized through airway irritation and constriction, influences millions global and poses sizeable challenges in its Well-Timed diagnosis and management. This literature survey provides an overview of latest improvements, demanding situations, and future directions in leveraging system gaining knowledge of for asthma danger prediction, highlighting its capability to transform clinical practice and improve the excellent of lifestyles for people suffering from this debilitating Condition.

**Predicting Continuity of Asthma Care Using a Machine Learning Model:[1]**

This project utilized longitudinal electronic health records of asthma patients to predict future exacerbations. Logistic Regression and Decision Trees were employed to model patient data, considering factors such as medication history, demographics, and environmental exposures.
Algorithms Used: Logistic Regression, Decision Trees
Limitations: Limited generalizability due to single-center data, potential biases in electronic health record documentation, and challenges in capturing real-time environmental factors.

**Deep Learning for Asthma Prediction with Environmental Data Integration:[2]**

This study integrated environmental data, such as air quality and pollen levels, with patient health records to predict asthma exacerbations. Convolutional Neural Networks (CNNs) and Long Short- Term Memory (LSTM) networks were used to capture spatial and temporal patterns in environmental data.

Algorithms Used: Convolutional Neural Networks, Long Short-Term Memory networks.

**Predicting asthma attacks in primary care: protocol for developing a machine learning based prediction model:[3]**

This research leveraged data from wearable devices, such as smart inhalers and activity trackers, to predict asthma attacks. An ensemble of machine learning algorithms, including Random Forest, AdaBoost, and Gradient Boosting, was utilized to combine information from multiple sensors.

**Asthma Prediction Using Transfer Learning and Electronic Health Records:[4]**

This study applied transfer learning techniques to asthma prediction by leveraging pre-trained models on large-scale healthcare datasets. Transfer learning was employed to fine-tune deep learning architectures, such as pre-trained CNNs and LSTMs, on electronic health record data specific to asthma patients.

Algorithms Used: Transfer Learning, Convolutional Neural Networks, Long Short-Term Memory networks.

**Development of childhood asthma prediction models using machine learning:[5]**

This Paper deals with the Identification of asthma at very early stages in which children fails to face. Machine learning approach is having better potentiality to experience better performance than any other existing models.

Algorithms used: Regression based models.

**Asthma Prediction Using Federated Learning Across Healthcare Institutions:[6]**

This study investigated federated learning techniques to build asthma prediction models while preserving data privacy across multiple healthcare institutions. Federated learning algorithms, such as Federated Averaging and Secure Aggregation, were employed to train models collaboratively on decentralized data.

Algorithms Used: Federated Learning.

**Asthma Detection System using Deep learning & Machine learning:[7]**

This study implies asthma detection on the basis of sound records which are gathered from various locations of the respiratory parts. Utilizing both of those deep and machine learning models which helps the doctors to diagnose asthma with more efficient and also high accuracy of 99.8%.

**Advanced Ensemble Learning Approach for Asthma Prediction:[8]**

This research proposed a hybrid approach combining traditional machine learning models with deep learning architectures for asthma prediction. An ensemble of models, including Random Forest, Gradient Boosting, and Deep Neural Networks, was utilized to capture diverse aspects of patient data.

Algorithms Used: Random Forest, Gradient Boosting, Deep Neural Networks.

**Machine Learning Approaches to Predict Asthma Exacerbations:[9]**

This study of Machine learning (ML) uses mathematical and statistical methods to detect patterns across large datasets including electronic health records (EHR). ML has the potential to augment clinical decision-making and provide appropriate treatment to improve both asthma prognosis and the overall quality of life.
Algorithms Used: Logistic Regression, Random Forest, Gradient Boosting Machine (GBM), Support Vector Machine (SVM).

**A Machine Learning Approach to Predicting Asthma Hospitalizations:[10]**

This Study about Gradient Boosting Machines (GBM): GBM algorithms, such as XG Boost and Light GBM, gained traction for asthma prediction due to their ability to handle large datasets efficiently and produce highly accurate predictions. Ensemble learning techniques using boosted trees were particularly effective in capturing subtle patterns in the data.

Algorithm Used: Gradient Boosting Machine (GBM).

**Predicting Asthma-related Emergency Department Visits Using Big Data:[11]**

This research of Predicting asthma-related ED visits using big data offers the potential to identify high-risk patients, optimize resource allocation, and implement preventive interventions to reduce healthcare costs and improve patient outcomes. However, it also poses challenges related to data privacy, algorithm interpretability, and model generalizability, which require careful consideration during the development and deployment phases.

Algorithm Used: Random Forest, GBM.

# 3. PROBLEM STATEMENT

Asthma, a prevalent respiratory ailment affecting millions worldwide, presents a multifaceted challenge due to its intricate nexus of genetic predispositions, environmental triggers, and individual health characteristics. Despite strides in medical science, accurately foreseeing asthma susceptibility and gauging control status remains elusive, impeding timely interventions and tailored treatment approaches. Current methodologies lack the sophistication needed to deliver precise prognoses and effective management blueprints.

Consequently, there arises an urgent need for an innovative solution harnessing the power of machine learning algorithms to prognosticate asthma risk and control status with unparalleled accuracy. This solution must seamlessly amalgamate heterogeneous datasets spanning environmental conditions, demographic profiles, and behavioral patterns, facilitating exhaustive analysis to pinpoint pivotal factors driving asthma development and severity.

Furthermore, a pressing requirement emerges for an intuitive interface that streamlines interaction with the predictive model, empowering both patients and healthcare professionals to input pertinent data effortlessly and receive personalized medication recommendations, along with insights into potential side effects and natural remedies. Bridging the chasm between cutting-edge machine learning methodologies and their real-world implementation in healthcare settings is imperative, as it equips individuals and healthcare providers with the tools to make well-informed decisions regarding asthma management, thereby fostering enhanced patient outcomes and a heightened quality of life.

# 4. REQUIREMENT ANALYSIS

## 4.1 Functional Requirements:

**i. User Input Interface:**

The website must feature a user-friendly interface where individuals can input relevant data such as age, gender, smoking habits, family history of asthma, and environmental factors like humidity and air pollution levels. This interface should be intuitive and accessible across various devices to ensure a seamless user experience.

**ii. Data Processing and Validation:**

The website should incorporate robust data processing mechanisms to handle user input securely and validate it for accuracy and completeness. This includes techniques for data cleaning, normalization, and handling missing values to ensure the reliability of the input data for the machine learning model.

**iii. Machine Learning Model Integration**:

The website must integrate a machine learning model trained on relevant asthma risk factors to accurately predict the likelihood of asthma development or exacerbation. This model should be capable of analyzing the input data and generating predictions in real-time, providing users with timely insights into their asthma risk.

**iv. Prediction Output Presentation:**

Once the prediction is generated by the machine learning model, the website should present the results to users in a clear and understandable format. This may include visualizations such as risk scores, probability estimates, or categorical predictions, accompanied by explanatory text to help users interpret the findings.

# 5. RISK ANALYSIS

Asthma chance prediction the usage of machine getting to know (ML) holds sizeable potential for revolutionizing early intervention techniques and improving patient consequences. However, like any technological advancement in healthcare, it also gives certain inherent risks that want to be carefully analyzed and mitigated. This threat analysis targets to identify and check capacity dangers related to making use of machine gaining knowledge of algorithms for bronchial asthma hazard prediction, thinking about elements including records best, version accuracy, ethical implications, and deployment challenges.

## 5.1 Data Quality:

One of the primary risks associated with asthma risk prediction using machine learning is the quality of the data used to train the models. Poor-quality or biased data can lead to inaccurate predictions and potentially harmful outcomes. Factors such as incomplete or outdated medical records, underrepresentation of certain demographic groups, and data privacy concerns can compromise the reliability and fairness of the predictive models. Therefore, rigorous data preprocessing techniques, data augmentation strategies, and adherence to privacy regulations are essential to mitigate this risk.

## 5.2 Model Accuracy and Generalization:

Another significant risk is the accuracy and generalization capabilities of the machine learning models developed for asthma risk prediction. Overfitting to the training data, lack of diversity in the dataset, and insufficient feature selection can undermine the reliability of the models when applied to real-world scenarios. Validation on independent datasets and robust evaluation metrics are crucial for assessing model performance and ensuring that predictions generalize well across diverse patient populations. Additionally, ongoing monitoring and adaptation of the models to evolving healthcare trends and patient characteristics are essential to maintain accuracy over time.

## 5.3 Ethical Implications:

The use of machine learning in healthcare raises ethical concerns regarding patient privacy, autonomy, and equity. Asthma risk prediction models may inadvertently perpetuate biases present in the data, leading to disparities in healthcare access and treatment outcomes.

Furthermore, the interpretation and communication of predictive results to patients must be conducted in a transparent and culturally sensitive manner to uphold patient autonomy and trust. Ethical guidelines and regulatory frameworks, such as informed consent protocols, algorithmic transparency, and fairness-aware machine learning techniques, are necessary to address these ethical considerations and safeguard patient rights.

**5.4 Deployment Challenges:**

Deploying machine learning models for asthma risk prediction in clinical settings poses various logistical and operational challenges. Integration with existing electronic health record systems, interoperability with healthcare infrastructure, and user adoption by healthcare professionals require careful planning and coordination. Moreover, ensuring the reliability, security, and scalability of the deployed systems is essential to minimize disruptions to clinical workflows and ensure the delivery of timely and accurate predictions. Collaborative partnerships between data scientists, clinicians, policymakers, and healthcare administrators are vital for overcoming these deployment challenges and realizing the full potential of machine learning in asthma risk prediction.

**5.5 Model Overfitting:**

Overfitting occurs when a machine learning model captures noise in the training data rather than underlying patterns, resulting in poor generalization to new data. Asthma risk prediction models susceptible to overfitting may yield inaccurate or unreliable predictions when applied to real-world patient populations. To mitigate this risk, techniques such as cross-validation, regularization, and ensemble methods can be employed to ensure model robustness and generalization.

**5.6 Data Privacy:**

Asthma risk prediction models rely on sensitive patient data, raising concerns about privacy and confidentiality. Unauthorized access, data breaches, or misuse of patient information could lead to legal and ethical repercussions, eroding patient trust and compromising healthcare provider credibility. To mitigate privacy risks, data encryption, access control mechanisms, and compliance with regulatory frameworks such essential to safeguard patient privacy and ensure data security.

# 6. FEASIBILITY ANALYSIS

The content outlines a comprehensive approach for asthma risk prediction using machine learning. Let's conduct a feasibility analysis to assess the practicality and viability of the proposed system:

### 6.1. Data Availability and Quality:

Assessing the feasibility of asthma risk prediction using machine learning begins with evaluating the availability and quality of relevant healthcare data. Feasibility hinges on the accessibility of electronic health records, patient demographics, clinical variables, and environmental factors pertinent to asthma risk. A thorough assessment of data completeness, accuracy, and relevance ensures the feasibility of developing reliable predictive models.

### 6.2. Algorithm Selection and Complexity:

Feasibility analysis involves identifying suitable machine learning algorithms and assessing their complexity for asthma risk prediction. Selecting algorithms that balance accuracy with computational efficiency is crucial, especially in healthcare settings with limited resources. Feasibility increases with the availability of algorithms capable of handling diverse data types, such as structured patient data and unstructured clinical notes, while maintaining scalability and interpretability.

### 6.3. Expertise and Resources:

The feasibility of implementing asthma risk prediction using machine learning relies on the availability of skilled personnel and resources within healthcare organizations. Assessing feasibility includes evaluating the presence of data scientists, machine learning engineers, and domain experts capable of developing and validating predictive models. Feasibility is enhanced by access to training programs, collaboration opportunities, and partnerships with industry experts.

**6.4. Regulatory Compliance:**

Feasibility analysis considers regulatory requirements and compliance considerations associated with deploying machine learning models in healthcare. Evaluating feasibility involves understanding regulations such as HIPAA, GDPR, and FDA guidelines governing data privacy, security, and medical device usage. Feasibility increases with adherence to regulatory frameworks and implementation of robust data protection measures to ensure compliance and safeguard patient privacy.

**6.5. Integration with Clinical Workflows:**

Feasibility assessment includes evaluating the integration of machine learning-based asthma risk prediction into existing clinical workflows. Feasibility hinges on interoperability with electronic health record systems, compatibility with healthcare infrastructure, and seamless integration with clinical decision support tools. Feasibility increases with solutions that streamline workflow processes, facilitate data exchange, and provide actionable insights to healthcare providers at the point of care.

**6.6. Cost-Benefit Analysis:**

Feasibility analysis involves conducting a cost-benefit assessment to evaluate the economic viability of implementing machine learning for asthma risk prediction. Assessing feasibility includes estimating upfront costs associated with data acquisition, algorithm development, infrastructure setup, and ongoing maintenance. Feasibility is enhanced by identifying potential cost savings, improved patient outcomes, and operational efficiencies resulting from early risk identification and targeted interventions.

**6.7. Stakeholder Acceptance and User Engagement**:

Feasibility assessment considers stakeholder acceptance and user engagement from healthcare providers, patients, and other stakeholders involved in asthma management. Feasibility is determined by soliciting feedback, addressing concerns, and building trust in predictive models through transparent communication and demonstration of clinical utility. Feasibility increases with active stakeholder engagement, collaboration, and support, ensuring the successful adoption and implementation of machine learning-based asthma risk prediction in clinical practice.

# 7. PROPOSED APPROACH

In our comprehensive project, we harness the power of advanced machine learning methodologies, specifically Support Vector Machines (SVM) and Random Forest algorithms, to delve into the intricate dynamics of asthma prediction and medication recommendation. Our approach goes beyond mere prediction, as we meticulously consider potential side effects when suggesting suitable medications. Through the adept application of these sophisticated algorithms, we meticulously analyze multifaceted datasets containing crucial factors such as genetic predisposition, environmental triggers, and individual responses to medications. This meticulous analysis enables us to provide accurate forecasts of asthma risk, empowering individuals and healthcare professionals alike with invaluable insights.

This emphasis on accessibility and value extends past mere capability, aiming to create an intuitive revel in that caters to the needs of both people and healthcare experts alike. By prioritizing consumer-centric layout ideas, our interface enables efficient navigation and comprehension of complex statistics, fostering knowledgeable selection-making at every stage of asthma control. Through our relentless pursuit of user empowerment, we aim to bridge the distance between facts analytics and practical software, making sure that our platform serves as a treasured resource for optimizing bronchial asthma care strategies and enhancing patient outcomes.

Furthermore, we have meticulously crafted a user-friendly Python graphical user interface (GUI) enriched with the capabilities of TensorFlow. Seamlessly integrated within the versatile Visual Studio (VS) code environment, our interface ensures accessibility and simplicity, catering to a diverse range of users. Whether individuals seek to understand their asthma risk or healthcare professionals aim to optimize treatment strategies, our system stands as a beacon of collaboration and informed decision-making in asthma management. Through our efforts, we aspire to empower users with actionable information, fostering enhanced collaboration and ultimately, improving outcomes in asthma care. Which ensures accessibility and usability for both individuals and healthcare professionals.

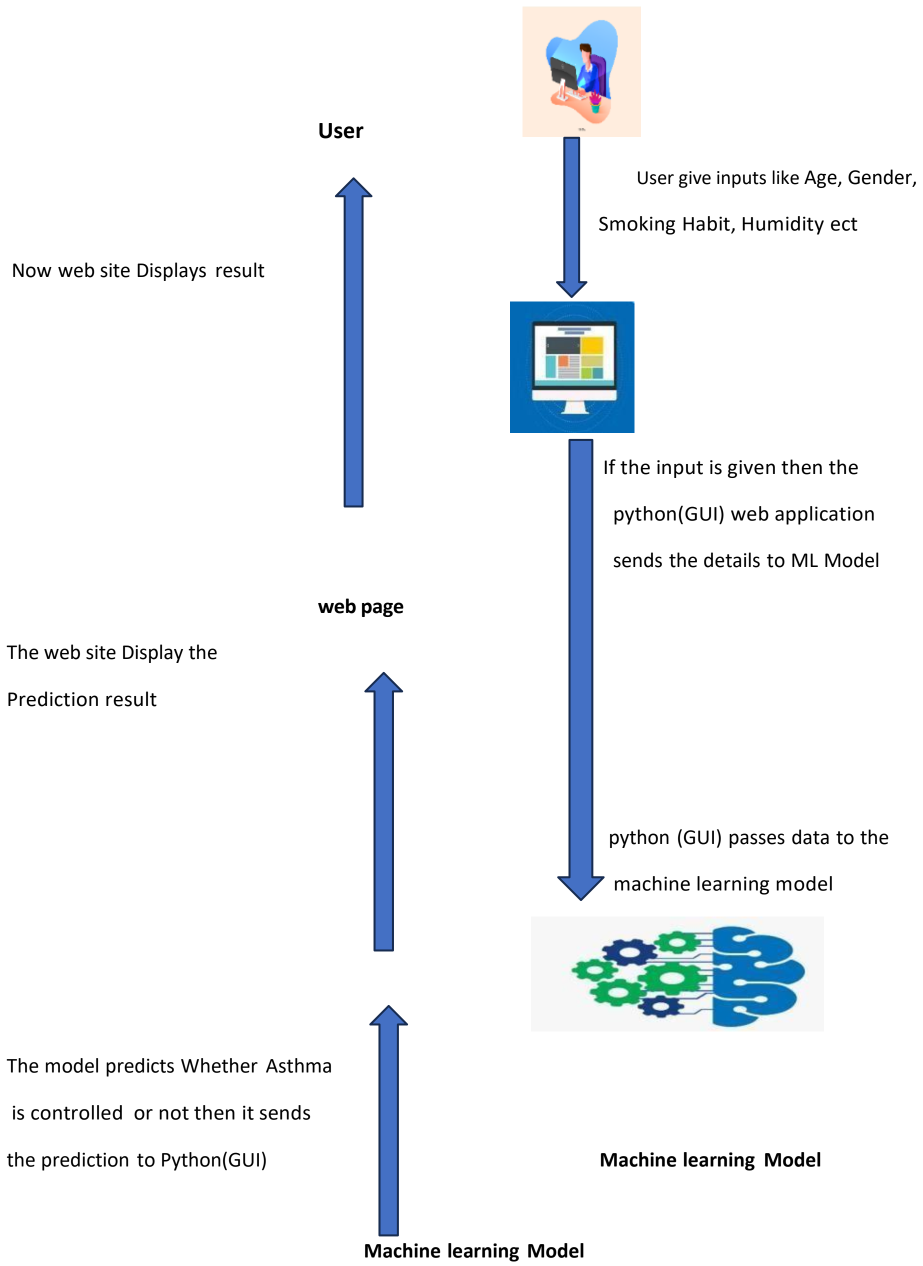


Fig 1: Overall Proposed plan

## Working of Proposed System:

1. Users input details like age, gender, smoking habits, humidity, etc., into the Python GUI application.
2. The Python GUI application processes this input data.
3. The application then forwards the processed data to the machine learning model for prediction.
4. The machine learning model analyses the input data to predict whether asthma is present or not.
5. Once the prediction is made, the result is sent back to the Python GUI application by the model.
6. The Python GUI application receives the prediction and displays it to the user.
7. Alongside the prediction, the application suggests medicines, natural remedies, and rare but serious side effects based on the prediction.
8. Users can review both the prediction and the suggested remedies to make informed decisions about asthma management.
9. This streamlined process enables users to access personalized insights and recommendations for effective asthma management through a user-friendly Python GUI interface.

# 8. ARCHITECTURE

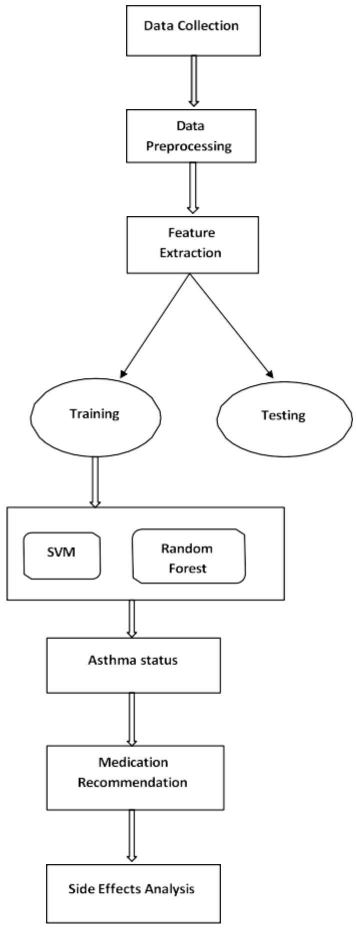


Fig.2, Block-diagram of proposed architecture

It is a simplified mechanism chart for a hypothetical asthma risk prediction using machine learning. Please note that the actual steps and details would depend on the specific design and requirements of the system.

1. **Collecting the data**: Collecting from all the datasets of asthma patients.
2. **Processing the data:** Processing all the information like age, gender, respiratory issues etc.
3. **Training of the data:** Usage of Support vector machine & Random-forest going to predict risk of asthma.
4. **Status of the user:** Predicting the user level of risk which helps to know the asthma range along with the medication beforehand.
5. **Medication Recommendation:** Getting the medication from expertise doctors to that particular level of risk.

## i. USERCASE DIAGRAM:

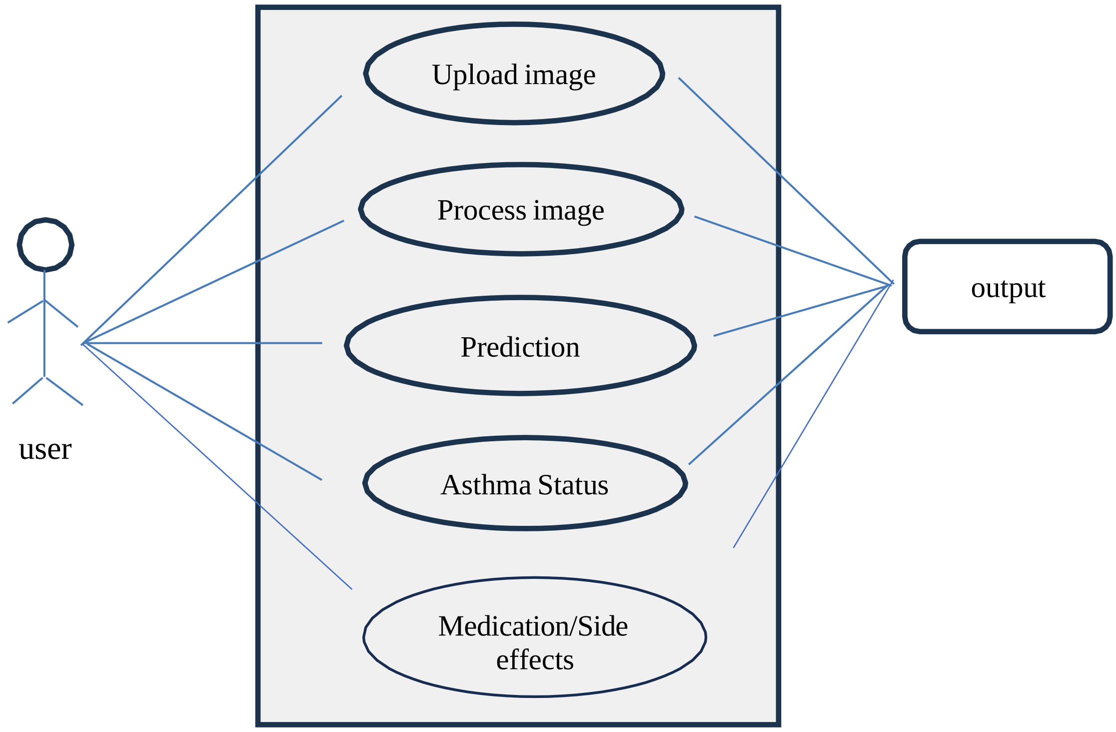


Fig3: Use Case Diagram

**Actors:**

User: Interacts with the system.

System: Represents the Asthma risk prediction System.

**Use Cases:**

Upload user data: User uploads the data for analysis.

View Diagnostic Results: User views the diagnostic data entered by user.

Prediction Results: Range of the asthma is identified along with medication.

Identify Range: System identifies and displays the required medication based on user input data.

Medication /Side effects: Resulted data helps the user to undergo the required traditional/Other medication.

**Relationships:**

User ->Upload User data: Association line connecting user and "Upload User data elements" use case.

User -> View Diagnostic Results: Association line connecting user and "View Diagnostic Results" use case.

User -> Prediction Results: Association line connecting user and "Prediction use case.

System -> Identify Range for asthma: Association line connecting the system and "Status" use case.

System -> Medication /Side effects: Association line connecting the system and " Medication /Side effects: "

## ii. SEQUENTIAL DIAGRAM:

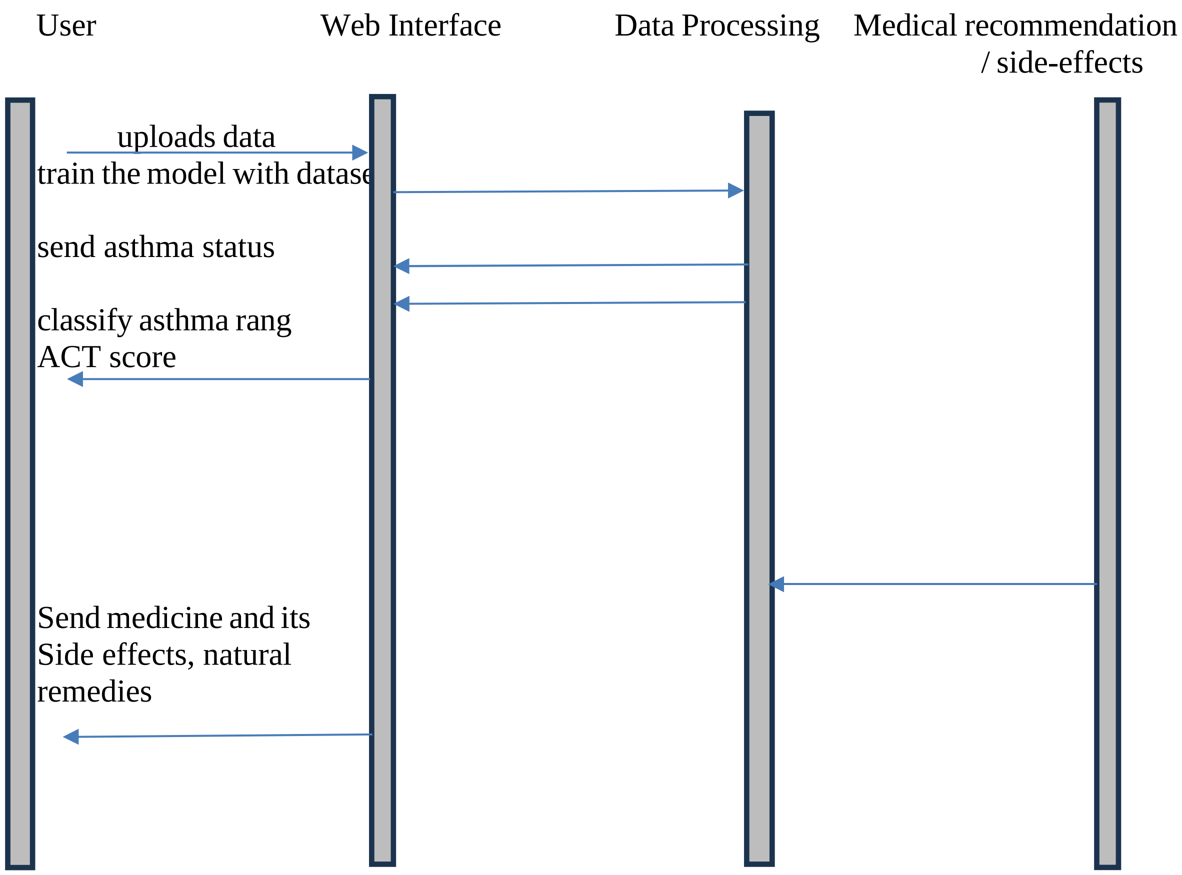


Fig 4: Sequential diagram of proposed method.

# 9. IMPLEMENTATION

## Simulation setup

**A. Hardware Requirements:**

Hardware : intel core i5/i7

Speed : 2.5 GHz

RAM : 8GB

Web camera: HD (720p) resolution

**B. Software Requirements:**

Operating System : Windows/macos

Technology : Machine Learning

Platform : Visual Studio Code, Google Collab

Python Libraries: Pandas, numpy

## IMPLEMENTATION:

The dataset utilized in this project comprises individual-level records encompassing various parameters crucial for asthma prediction and management. These parameters include demographic information, environmental variables, behavioral attributes, and asthma severity scores. Each record encapsulates a unique combination of these variables, providing a comprehensive dataset for analysis and model training.

### 9.1. Data Collection and Preprocessing:

**i. Data Source**: The dataset consists of various parameters such as location, age, gender, outdoor job, outdoor activities, smoking habit, humidity, pressure, temperature, UV index, wind speed, and ACT score.

Data is collected from individuals above the age of 50 from different locations, capturing diverse environmental and demographic factors.

**ii. Data Split**: The dataset is split into training, validation, and testing sets to ensure model generalization and performance evaluation.

Typically, a ratio of 70:15:15 is used for training, validation, and testing, respectively.

**iii. Data Augmentation**: Data augmentation techniques such as random rotation, flipping, and zooming may be applied to increase the diversity of the dataset and improve model robustness This helps in preventing overfitting and enhances the model's ability to generalize to unseen data.

**iv. Feature Extraction:** Feature extraction involves selecting and transforming relevant features from the dataset to improve the model's predictive performance.

Techniques like normalization, scaling, and encoding categorical variables are applied to preprocess the data effectively.

## 9.2. Model Architecture:

**i. Support Vector Machine (SVM):**

Support Vector Machine (SVM) is an effective supervised device gaining knowledge of set of rules used for class and regression obligations. It works by way of finding the most useful hyperplane that separates unique lessons inside the characteristic area, maximizing the margin between the lessons.

**SVM** is powerful in high-dimensional spaces and when the wide variety of functions exceeds the quantity of samples.it can cope with each linear and non-linear information via the use of various kernel functions. **SVMs** are extensively utilized in diverse fields including image type, text type, and bioinformatics due to their versatility and ability to address complicated datasets successfully.

**SVM** is employed as one of the primary algorithms for asthma prediction due to its effectiveness in handling high-dimensional data and finding optimal hyperplanes for classification. It aims to separate individuals into asthma-positive and asthma-negative classes based on the input features. This model is trained using the preprocessed dataset to learn the underlying patterns and relationships between the features and asthma occurrence.

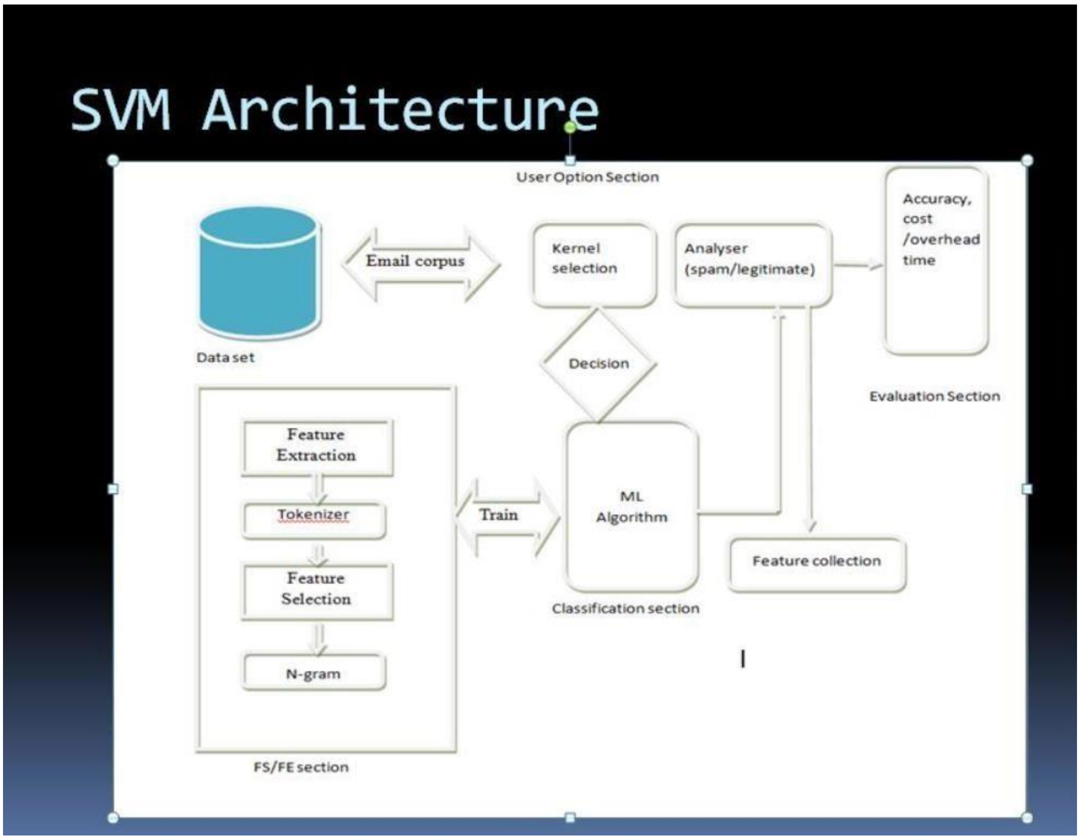


Fig 5: SVM Architecture

**ii. Random Forest:**

Random Forest is a versatile and widely-used ensemble learning algorithm in gadget learning. It operates by using building a mess of decision trees all through the education segment and outputs the elegance this is the mode of the training (type) or imply prediction (regression) of the character timber. Each choice tree inside the Random Forest is skilled on a random subset of the education facts and a random subset of features, making it robust to overfitting and distinctly accurate. Random Forest is effective for each class and regression tasks, and its capacity to deal with big datasets with high dimensionality makes it a popular desire in numerous programs which include finance, healthcare, and bioinformatics. Random Forest is another key algorithm utilized in the asthma prediction system.
It leverages ensemble learning by constructing multiple decision trees during the training phase. each decision tree in the ensemble independently predicts asthma occurrence, and the final prediction is determined by aggregating the results from all trees.

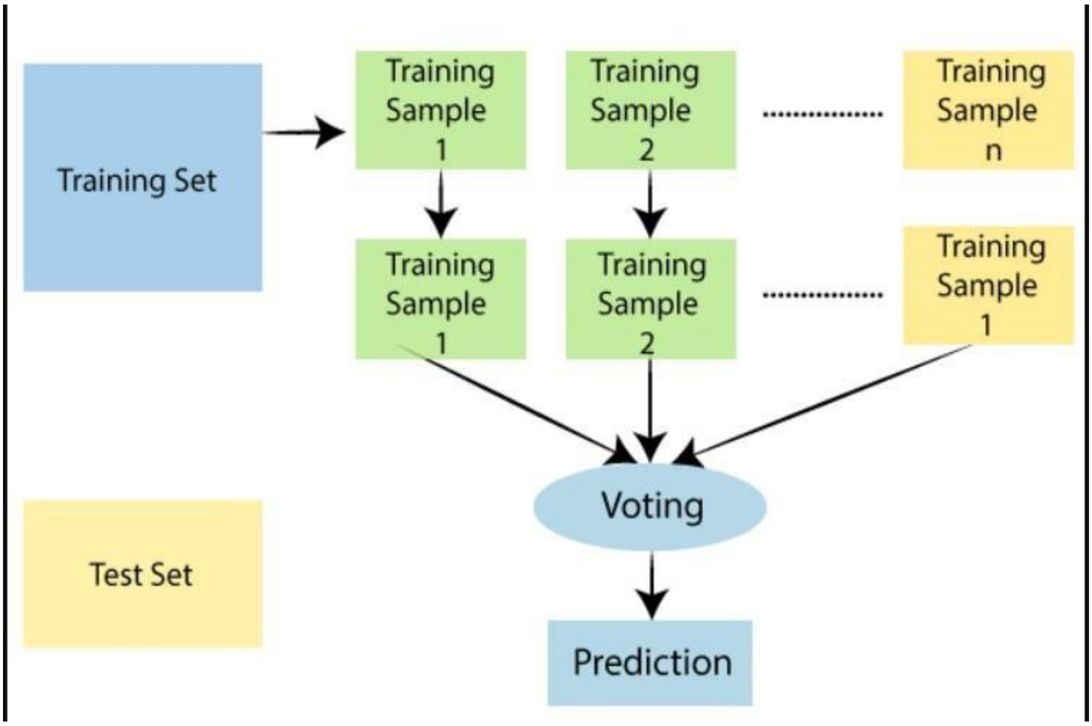


Fig 6: Random Forest Architecture

**9.3. Model Integration**

Both the SVM and Random Forest models are integrated into the prediction system to leverage the strengths of each algorithm.

Ensemble techniques may be employed to combine the predictions from both models for improved accuracy and robustness.

**9.4. Model Evaluation**

The performance of the SVM and Random Forest models is evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score.

Cross-validation techniques may be applied to ensure reliable performance assessment and mitigate overfitting.

**9.5. Model Deployment**

Once trained, evaluated, and integrated, the combined model is deployed into a production environment for real-time asthma prediction.

Deployment involves setting up the model on a server or cloud platform to make predictions accessible via the Python GUI.

### 9.6. User Interface (Python GUI Using TensorFlow):

A user-friendly graphical user interface (GUI) powered by TensorFlow is developed to interact with the prediction system.

Users can input their data, such as demographic and environmental factors, and receive personalized predictions regarding asthma occurrence.

Additionally, the GUI provides recommendations for medication and highlights potential side effects based on the predicted asthma status.

## Implementation of app.py:

I. **Importing Libraries:**

**tkinter:**

import tkinter as tk: This imports the tkinter module and aliases it as tk. Tkinter is the standard GUI (Graphical User Interface) toolkit for Python. It provides classes and functions for creating desktop applications with GUI elements like windows, buttons, labels, etc.

**pandas:**

import pandas as pd: This imports the Pandas library and aliases it as pd. Pandas is a powerful data manipulation and analysis library for Python. It provides data structures like Data Frame for handling structured data and operations to manipulate and analyze the data efficiently.

**scikit-learn:**

from sklearn.preprocessing import Standard Scaler: This imports the Standard Scaler class from the preprocessing module of scikit-learn. Standard Scaler is used for standardizing features by removing the mean and scaling to unit variance.

from sklearn.svm import SVC: This imports the Support Vector Classifier (SVC) class from the svm module of scikit-learn. SVC is a supervised learning model used for classification tasks.

from sklearn.ensemble import Random Forest Classifier: This imports the Random Forest Classifier class from the ensemble module of scikit-learn. Random Forest Classifier is an ensemble learning method for classification tasks.

from sklearn.model_selection import train_test_split: This imports the train_test_split function from the model_selection module of scikit-learn. It's used for splitting data into training and testing sets for model evaluation.

At the start of the script, essential libraries and modules are imported to facilitate various functionalities. The "app.py" script relies on the Flask framework for web application development. Additionally, it integrates the functionalities of the "file" module, which is responsible for the image upload and prediction.

```
from flask import Flask, request, render_template
from tensorflow import keras
import numpy as np
from PIL import Image
import os
from tensorflow.keras.preprocessing import image
```

## II. Functioning to Train Models:

The **train_models** function is responsible for training two types of classifiers: Support Vector Machine (SVM) and Random Forest.

**SVM Classifier Initialization:**

It initializes an SVM classifier (svm_clf) using the SVC class from scikit-learn.

The SVM classifier is configured with a linear kernel and a specified random state for reproducibility (random_state=42).

**Random Forest Classifier Initialization:**

It initializes a Random Forest classifier (rf_clf) using the RandomForestClassifier class from scikit-learn.

The Random Forest classifier is configured with 100 estimators (trees) and a specified random state for reproducibility (n_estimators=100, random_state=42).

**Model Training:**

Both classifiers (svm_clf and rf_clf) are trained using the scaled training data (X_train_scaled and y_train) obtained from the data splitting step.

**Return Trained Models:**

Finally, the function returns the trained SVM and Random Forest classifiers (svm_clf, rf_clf) as output.

## III. Prediction of asthma:

**Input Data Preparation:**

It converts the input values provided by the user into a pandas Data Frame with appropriate column labels (labels) to match the features used during model training.

**Data Scaling:**

It scales the input data using the same scaler that was fitted during model training. This ensures that the input data is transformed in the same way as the training data.

**Model Prediction:**

It predicts asthma using both the SVM and Random Forest classifiers on the scaled input data. The predictions are obtained separately from both classifiers (svm_pred and rf_pred).

**Return Predictions:**

Finally, the function returns the predictions made by both classifiers (svm_pred and rf_pred).

The **predict_asthma** function plays a crucial role in the application by utilizing the trained machine learning models to make predictions about asthma severity, allowing users to gain insights into potential asthma diagnoses based on their input data.

## IV. GUI creation:

This segment of code is responsible for a Tkinter window titled "Asthma Detection System" is created. Inside this window, a frame (main_frame) is constructed to hold the main content of the application. This frame is configured with padding to provide spacing around its contents.

```
# Create the main window root = tk.Tk() root.title("Asthma Risk Prediction System")
```

Next, labels and entry fields for input features are created dynamically based on the list **labels** which contain the names of the input features such as age, gender, outdoor job, etc.

```
# Create a frame for the main content main_frame = ttk.Frame(root, padding="20")
main_frame.grid(row=0, column=0, sticky="ns
```

A loop iterates over each label in the labels list. For each label, a **Tkinter** label widget **(ttk.Label)** is created and placed within the **main_frame.**

The label is positioned in the grid layout using the grid method with a row index (i) corresponding to its position in the list and a column index of 0 to align it to the left side of the frame. The sticky parameter is set to "w" to ensure the label sticks.

# Create labels and entry fields for the input features labels = ['Age', 'Gender', 'OutdoorJob', 'OutdoorActivities', 'SmokingHabit', 'Humidity', 'Pressure', 'Temperature', 'UVIndex', 'WindSpeed'] entries = {} for i, label in enumerate(labels): ttk.Label(main_frame, text=label).grid(row=i, column=0, sticky="w") entries[label] = ttk.Entry(main_frame) entries[label].grid(row=i, column=1, sticky="ew")

### V. Display Prediction:

The **display_prediction** function retrieves input values from the GUI entry fields, predicts asthma severity using trained models, updates the GUI with prediction results, and display relevant medical information based on the predicted severity. It achieves this in just a few lines of code, making the application efficient and user-friendly.

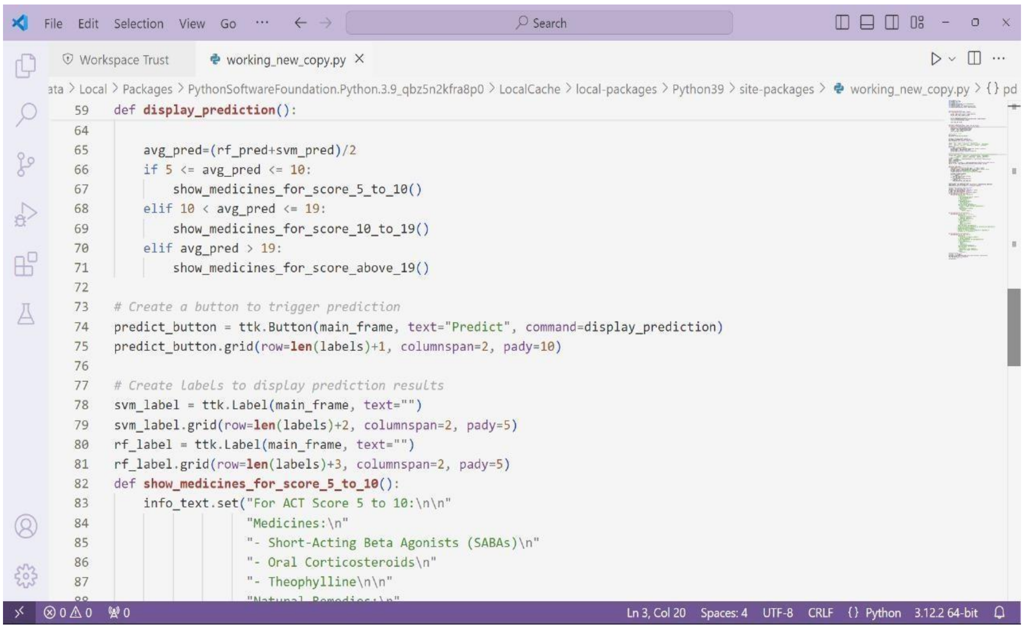

Fig.7: Display Prediction of Asthma risk prediction

### V. Medicine Information:

The functions show_medicines_for_score_5_to_10, show_medicines_for_score_10_to_19, and show_medicines_for_score_above_19 designed to provide users with medicine information tailored to specific ranges of predicted ACT scores. Each function offers guidance on medication options, natural remedies, and potential side effects corresponding to different severity levels of asthma.

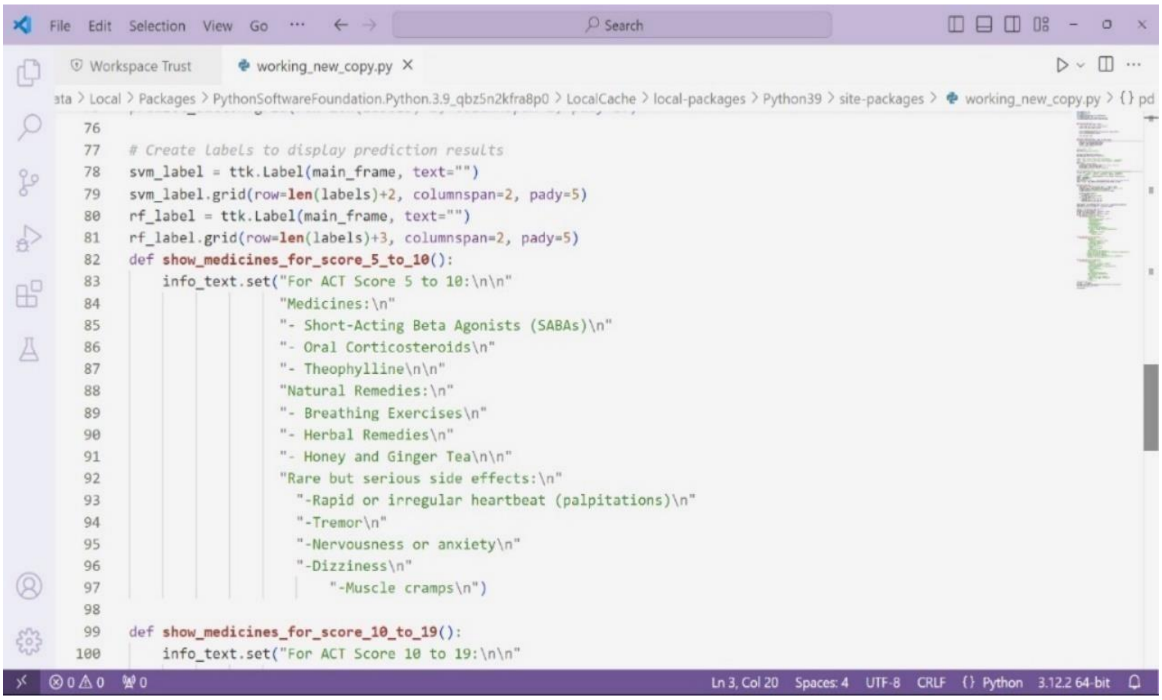


Fig8: Medication Support for range 5 to 10 ,10 to 19 ACT.

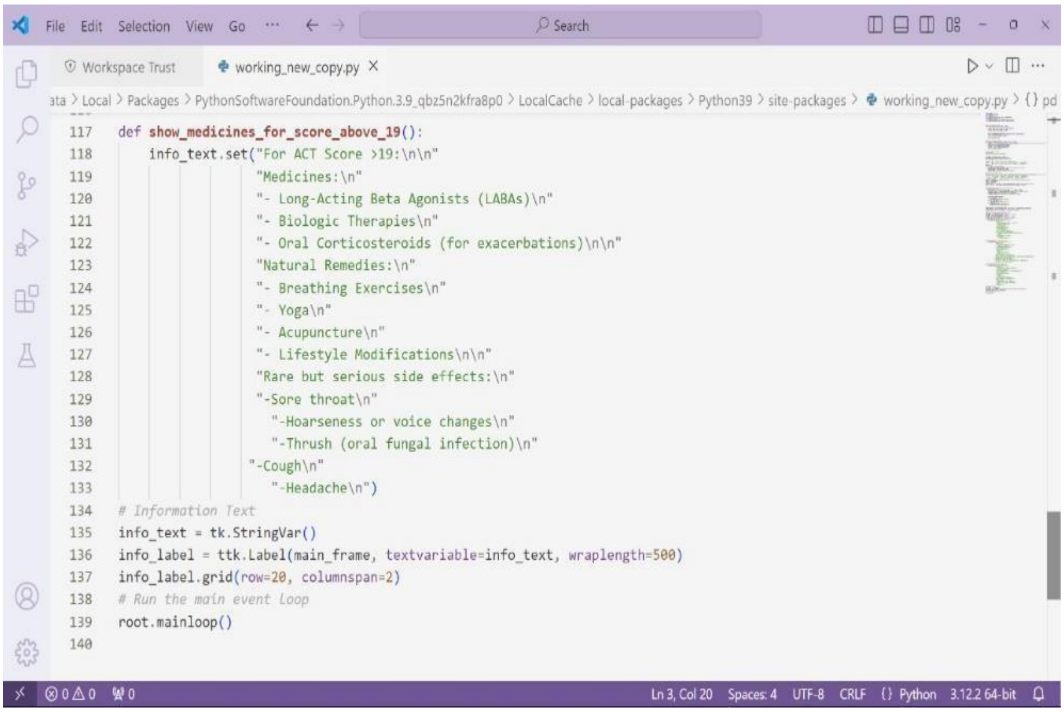


Fig9: Medication Support for range Above 19 ACT.

## VI. Information Text:

Information Text label is a crucial component of the application's GUI, providing users with essential information about medicines and potential side effects corresponding to the predicted ACT score range. By dynamically updating based on the severity level of asthma predicted by the machine learning models, it ensures users receive tailored guidance regarding medication options and associated risks. This feature enhances user understanding and decision-making by offering comprehensive insights into treatment recommendations and potential implications for different asthma severity levels. Additionally, the Information Text label contributes to the application's usability and transparency, promoting informed healthcare decisions and fostering a sense of empowerment among users.

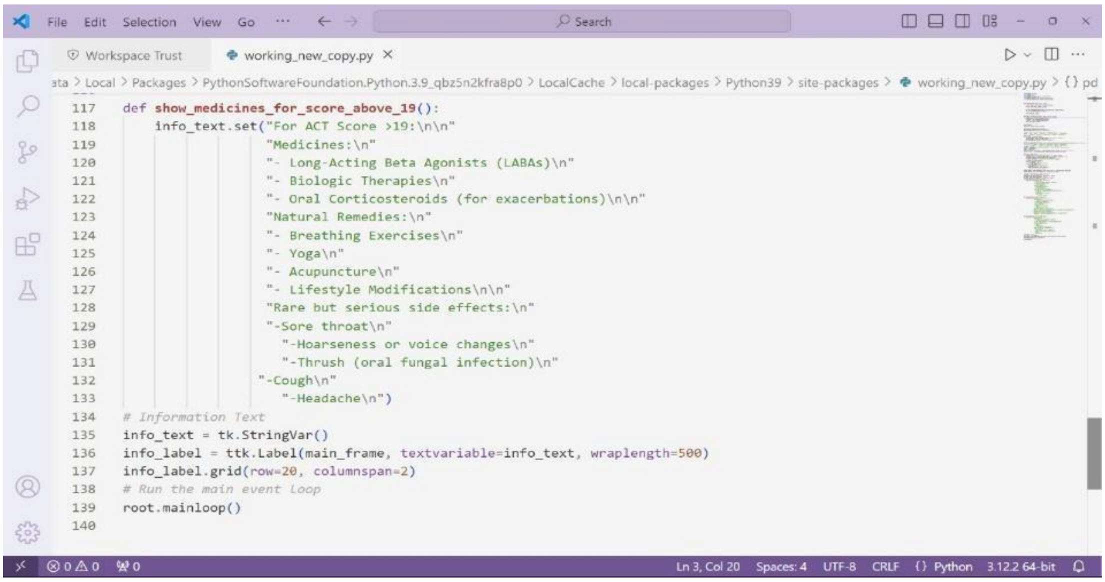


Fig 10: Information Text label

# 10. RESULTS AND DISCUSSION

The asthma risk prediction project is built upon the utilization of Support Vector Machine (SVM) and Random Forest models, which are employed to evaluate the likelihood of asthma occurrence based on user-provided data. Upon submission of input data through the intuitive Python Graphical User Interface (GUI), the information undergoes meticulous preprocessing steps, which include the crucial task of encoding categorical variables to ensure compatibility with the models. Subsequently, the SVM and Random Forest models swing into action, leveraging the processed data to predict the asthma control status for the individual. The outcomes of these predictions, coupled with recommendations for suitable medications and a thorough exploration of potential side effects, are then promptly relayed to the user via the web interface. this project utilizes SVM and Random Forest models to assess asthma likelihood based on user-input data. After input submission through the Python GUI, the data undergoes preprocessing, including encoding categorical variables. The models predict asthma control status, and the results, along with medication suggestions and potential side effects, are displayed on the web interface. Users receive clear indications of asthma prediction ranges, medication recommendations, and rare but serious side effects to enable informed decision-making about their health management. This streamlined process enhances asthma care by providing personalized insights and actionable recommendations.


.

Fig 11: Website interface of Asthma Risk Prediction.

Initially, users interact with a Python GUI application to provide the necessary asthma diagnosis information, such as age, gender, smoking habit, and humidity This input information this implementation forms the cornerstone of personalized health assessments in the process, providing customized insights and recommendations.

Then, the Python GUI application assumes the role of data manager, carefully processing user-provided information. This section includes actions aimed at ensuring the accuracy and relevance of the data, thus improving its suitability for subsequent analysis After completing data processing, the Python GUI application configures sophisticated data to be sent to sophisticated machine learning models specially designed for asthma prediction Powered by advanced algorithms, this model carefully examines input data beyond rigorous analysis The crucial of the model on the probability of manifesting asthma in the wood user It is predictive. Having completed its analysis quickly, the model quickly feeds back findings to the Python GUI application, triggering users interact with the application to review predictions and make decisions appropriate in the early stages of asthma control.

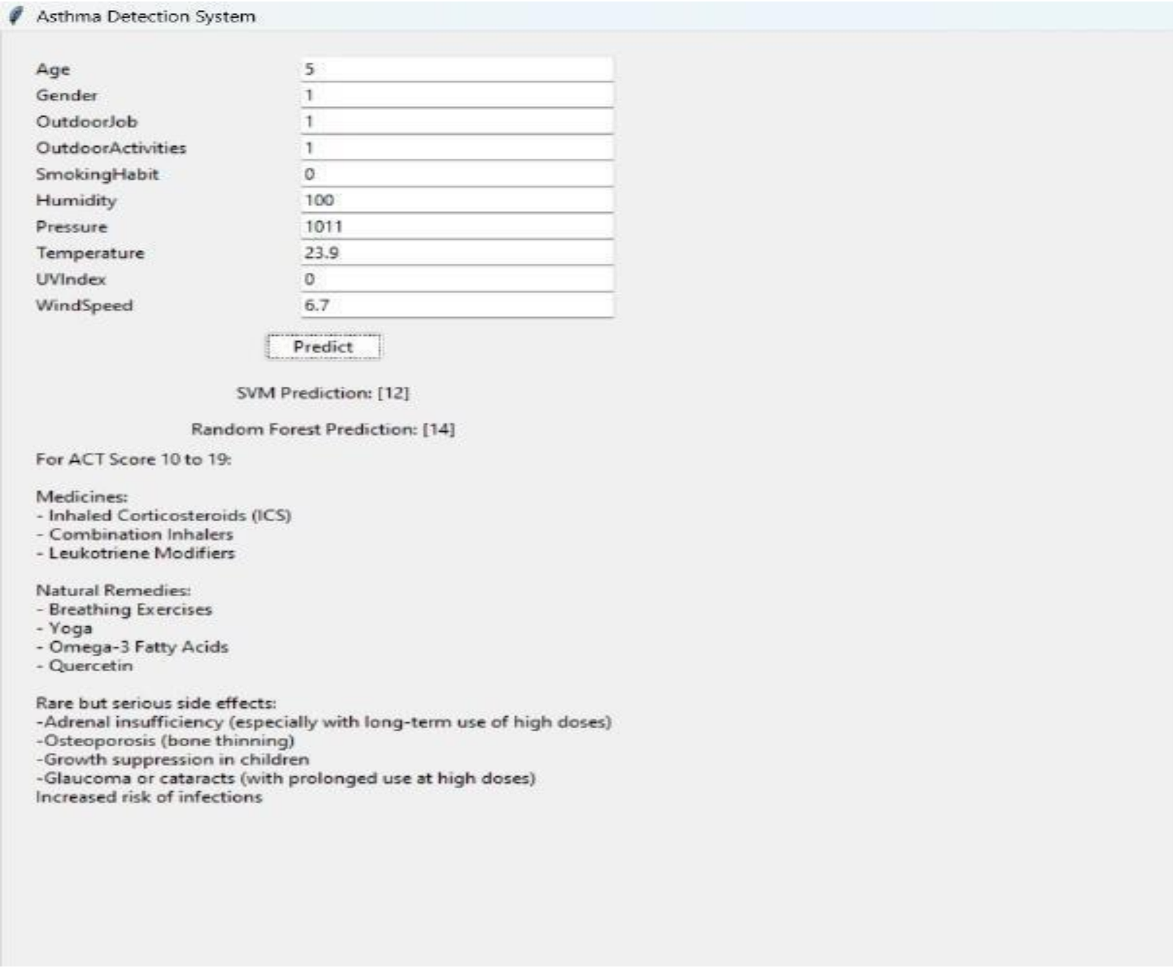


Fig 12: Asthma risk prediction Along with medications Interface.

# 11. CONCLUSION AND FUTURESCOPE

In conclusion, the integration of machine learning algorithms such as Support Vector Machine (SVM) and Random Forest into asthma risk prediction represents a pivotal advancement in healthcare innovation. Through the seamless fusion of user-provided data with sophisticated predictive models, this project heralds a new era of personalized assessments for asthma susceptibility. However, its impact extends far beyond mere diagnosis, as the system goes the extra mile by offering tailored recommendations for medication, exploring natural remedies, and shedding light on potential side effects associated with treatment options. This holistic approach not only equips individuals with the knowledge to make informed decisions about their health but also empowers them to take proactive measures in managing their asthma effectively.

By putting the tools for asthma management directly into the hands of individuals, this initiative facilitates better health outcomes and ultimately enhances the overall quality of life for those affected by asthma. Furthermore, it underscores the transformative potential of technology in revolutionizing healthcare delivery, shifting the paradigm from reactive to proactive care. Through the synergy of machine learning algorithms and personalized health interventions, this project exemplifies how technology can empower individuals to take charge of their well-being, ushering in a future where healthcare is not only more effective but also more personalized and accessible to all.

## Future Scope:

The future scope for asthma risk prediction using machine learning is promising, with the potential to significantly enhance early intervention strategies and improve patient outcomes. Advancements in machine learning algorithms, coupled with the availability of large-scale healthcare data, offer opportunities to develop more accurate and personalized predictive models for asthma risk assessment. Future research may focus on integrating diverse sources of data, including genetic information, environmental exposures, and lifestyle factors, to refine predictive models and provide tailored risk assessments for individual patients. By leveraging advanced machine learning techniques such as deep learning, ensemble methods, and federated learning, researchers can extract meaningful insights from complex datasets and uncover novel predictive biomarkers for asthma risk prediction.

Moreover, the future of asthma risk prediction using machine learning lies in its integration with clinical decision support systems and digital health technologies. By embedding predictive models into electronic health record systems, wearable devices, and mobile applications, healthcare providers can receive real-time alerts and personalized recommendations to guide asthma management decisions. Future developments may focus on enhancing the usability and accessibility of predictive tools, enabling seamless integration into clinical workflows and empowering patients to actively participate in their asthma care. Additionally, the integration of predictive analytics with telemedicine platforms and remote monitoring solutions holds promise for expanding access to preventive care and improving patient engagement in asthma management.

Furthermore, the future of asthma risk prediction using machine learning extends beyond individual patient care to population health management and public health initiatives. By estimating on large-scale healthcare datasets and population-level risk factors, predictive models can support targeted interventions, resource allocation, and policy planning efforts to mitigate the burden of asthma on communities. Future research may explore the use of machine learning for identifying high-risk populations, predicting disease outbreaks, and optimizing preventive measures to reduce asthma-related morbidity and mortality. Additionally, interdisciplinary collaboration between healthcare providers, researchers, policymakers, and technology developers will be essential to address ethical, regulatory, and implementation challenges and realize the full potential of machine learning in transforming asthma care delivery and improving population health outcomes.

# 12. REFERENCES


**1.**Yao Tong Beilei Lin, Gang Chen, and Zhenxiang Zhang "Predicting Continuity of Asthma Care Using a Machine Learning Model". Environmental and Public Health (2022) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8835449/.

**2.** Nestor A. Molfino, Gianluca Turcatel & Daniel Riskin "Deep Learning for Asthma Prediction With Environmental Integration" Advances in therapy 41: 534-552 (2024) https://link.springer.com/article/10.1007/s12325-023-02743-3.

**3**. Holly Tibble,Athanasios Tsanas,Elsie Horne "Predicting asthma attacks in primary care: protocol for developing a machine learning-based prediction model" (2019): A Scientific Literature https://www.ncbi.nlm.nih.gov/pmc/articles/PMC66240/.

**4.** Alexander Cobian, Madeline Abbott, BS, Akshay Sood "Asthma prediction using transfer Learning & Electronic health records." 2020: 98–107. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233095/.

**5**. Dilini M Kothalawala ,Clare S Murray,Angela Simpson "Development of childhood asthma prediction models using machine learning" https://pubmed.ncbi.nlm.nih.gov/34841728/.

**6.** Ittai Dayan, Holger R. Roth, Aoxiao Zhong Asthma Prediction Using Federated Learning Across Healthcare Institutions: A Journal 27: 1735–1743 (2021) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7842859/.

**7.** Mohammed Tawfik, Nasser M. Al-Zidi "Asthma Detection System using Deep & Machine learning (2021). https://www.researchgate.net/profile/Mohammed-Tawfik/publication/357310948_Asthma_Detection_System_Machine_and_Deep_Learning-Based_Techniques/links/6377a54154eb5f547ce2ef08/Asthma-Detection-System-Machine-and-Deep- Learning-Based-Techniques.pdf.

**8.**Sunil Kumar; Riya; Sanat Rampal " Advanced Ensemble Learning Approach for Asthma Prediction Vatsal "(2024) https://ieeexplore.ieee.org/abstract/document/10486189.

**9.** Nestor A. Molfino, Gianluca Turcatel, Daniel Riskin "Machine Learning Approaches to Predict Asthma Exacerbations": Advances in Therapy : A Narrative Review (2023) https://www.researchgate.net/publication/376644367_Machine_Learning_Approaches_to_Predict_Asthma_Exacerbations.

**10.** Chamberlain MS, James M. Chamberlain MD " Machine Learning Approach to Predicting Asthma Hospitalizations": Academic Emergency 25: 463-1470A https://onlinelibrary.wiley.com/doi/full/10.1111/acem.13655.

**11.** https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7045443