

Special Issue – AI-Driven and Multimodal Innovations in Biomedical Imaging and Sensing

Identifying Asthma Risk Factors and Developing Predictive Models for Early Intervention Using Machine Learning

Priyanshi Kotlia^{1*}, Janmejy Pant² and Manoj Chandra Lohani²

¹Research Scholar, School of Computing Graphic Era Hill University, Bhimtal Campus, India.

²Department of Computer Science and Engineering,
Graphic Era Hill University, Bhimtal Campus, India.

*Corresponding author Email: priyanshikotlia585@gmail.com

<https://dx.doi.org/10.13005/bpj/3089>

(Received: 30 December 2024; accepted: 14 February 2025)

The chronic respiratory illness called asthma causes substantial life quality deterioration for countless people across the world. Adequate diagnosis in the early stages of the condition proves essential for effective treatment which benefits the health status of patients while boosting their productivity levels. Asthma diagnosis shows difficulties in practice because of its clinical similarities with other related respiratory conditions. A research project applies machine learning models to environmental and physiological along with lifestyle data with the purpose of improving asthma diagnosis and forecasting capabilities. A combination of age, gender, familial asthma background, BMI, FEV1/FVC ratio, allergen exposure, AQI, smoking exposure, physical activity levels and diet quality indices serves as independent variables throughout the research assessment. The research depends on data mining methods together with machine learning algorithms including Random Forest, Logistic Regression, and XGBoost to reach exact prediction results. The evaluation metrics consist of accuracy and F1-score together with precision and recall as well as ROC curves to assess model performance. The prediction accuracy reaches 99% for Random Forest and XGBoost while their ROC score reaches 98% which demonstrates their competence in asthma classification. The lower performance of Logistic Regression produced an accuracy of 85% along with an ROC score of 94%. The research results demonstrate that machine learning holds remarkable prospects to transform medical practice when applied to asthma diagnosis and treatment. The use of multiple predictive variables through this diagnostic method leads to much improved diagnostic precision which supports appropriate medical care at the proper time. Future research efforts will concentrate on enlarging the available dataset as well as developing advanced transfer learning methods to optimize the model's functionality for low-resource medical environments. The findings from this research create pathways to develop better diagnostic instruments that enhance asthma treatment approaches for improved patient healthcare.

Keywords: Asthma, Clinical decision making, Chronic obstructive pulmonary disease, Machine learning, Respiratory.

The chronic respiratory disease asthma affects numerous individuals across the world. It is now widely known that medical science acknowledges asthma as multiple diseases that

share common clinical expressions. Asthma features the same pulmonary symptoms yet different biological characteristics, which results in various treatment effects.¹ Traditional asthma

categories were established using two main factors: symptom intensity and response to prescribed medications. Traditional asthma classification approaches have shown insufficient ability to represent the diverse aspects of the disease entity. Advances in molecular research have more effective targeted treatments require researchers to focus on exact disease classifications.² The creation of anti-IL-5 biologics represents a major advancement in asthma medical solutions. These treatments did not deliver results in patients with mild or moderate asthma but provided highly effective outcomes for individuals with glucocorticoid dependent eosinophilic asthma.² New targeted therapies gained approval from the healthcare authorities after identifying glucocorticoid dependent eosinophilic asthma. Different patient reactions to treatments increase treatment expenses because of their varying response rates. The expenses involved in developing each new drug amount to \$2.6 billion according to the researchers Mersha and Ray.^{2, 3} Machine the combination of learning and artificial intelligence (AI) has developed a significant contribution toward improving asthma classification methods. The research isolated four principal asthma endotypes through analysis of epithelial cells obtained from nasal passageways.⁴

Research indicates that worldwide asthma prevalence grows due to environmental, genetic and lifestyle reasons. The global asthma numbers currently stand at 300 million people yet experts predict this population will expand to 400 million by 2025.⁴ The Various essential variables contribute to the projected increase in asthma case numbers. Air pollution from industrial each release of industrial emissions along with vehicle exhaust and indoor pollutants worsens respiratory exposure among people. Rapid urban development has caused population growth which in turn creates denser areas where allergens and environmental factors become elevated.⁵ The impact of climate change upon rising temperatures affects the situation together with environmental conditions. The combination of weather extremes enhances the numbers of airborne allergens and pollen in the environment. The obesity epidemic has the relationship between asthma development includes two distinct endotypes which manifest through IL-6 along with IL-17-driven inflammatory responses

more prevalent.⁶ Reduction of breastfeeding along with multiple early life contact events produces adverse effects on immune system development. Early exposure to pollutants together with antibiotic overuse and improper antibiotic use dysregulate the immune system development of children. The development of asthma emerges during adulthood because of these exposures.⁷ The respiratory disease called chronic obstructive pulmonary disease (COPD) represents another significant health challenge with current global prevalence of 330 million people who suffer from it annually. The disease affects 330 million affected people annually and the global mortality reaches about 3 million annual deaths. While these two prevalent respiratory conditions share a common characteristic because they produce persistent breathing complications however their causes and treatment methods remain separate.⁷ COPD represents a progressive condition that rejects treatment initiatives because smoking together with prolonged dangerous air pollutant exposure trigger its main development mechanism. The main contributors to COPD development include cigarette smoking and consistent inhalation of toxic air pollutants. Asthma shows the ability to become reversible whereas the disease processes of COPD remain unalterable. Both genetic and environmental factors serve as the triggering elements for this condition. Some patients present medical signs which match those of both asthma and COPD pathology. The combination of asthma with COPD results in a medical condition called asthma-COPD overlap syndrome which physicians identify as ACOS (Asthma-COPD Overlap Syndrome). The symptoms of asthma result from inflamed airways and swelling together with the shrinking of breathing tubes which causes difficulty in breathing along with coughing and wheezing. The airway inflammation together with swollen and narrowed breathing tubes results in breathing issues which are accompanied by coughing and wheezing events. Multiple risk elements influence the start and development of asthma conditions. Asthma continues to deteriorate due to a combination of inherited risk factors together with surrounding pollution elements and allergic triggers and workplace hazards.⁸ The three main contributors to asthma development include allergens as well as workplace irritants along with

environmental pollutants. Additionally, lifestyle factors like obesity and the severity of asthma typically gets worsened by eating unhealthy foods.⁸ The lungs of young children can undergo development problems from early respiratory infections resulting in future asthma susceptibility. Exposure to smoking and secondhand smoke exposure causes additional risk because it both aggravates respiratory tissues with inflammation and irritates airways.⁸ The development of better asthma treatments requires a full comprehension of both medical characteristics and molecular processes of the condition. AI-based research enables better diagnosis methods and patient treatment response predictions through application of machine learning technology.^{9,10} Improvements in precision medicine will allow doctors to select more effective customized treatments based on how patients react to therapy worldwide.¹⁰

Here's how each of the independent variables listed contributes to the occurrence and severity of asthma:

Age

The immune system development differences between children and adults cause asthma symptoms to appear differently between these age groups. The time of childhood often serves as when asthma develop in those who have it. The visible onset of asthma occurs in children because their immune systems along with their respiratory organs remain in development yet immature more susceptible to inflammation and airway hyperresponsiveness.¹⁰ As a result, the asthma symptoms of children tend to become more significant and occur more frequently than in adult populations. In children, Asthma triggers in children mostly emerge from allergens like dust along with pollen substances and pet dander in addition to respiratory infections affecting children's asthmatic symptoms. The asthma symptoms become worse because children have incomplete immune defense capabilities.¹⁰ In the primary causes for developing asthma as an adult commonly involve smoke related irritants combined with pollution sources from the environment. Adults typically develop asthma because of their exposure to workplace chemicals together with their existing allergic conditions while environmental triggers such as pollution and smoke also play a role. Respiratory infections occasionally function as asthma setters

for individuals of all ages. Childhood triggers affect both age groups yet the weak immune systems of children make their asthmatic reactions worse. The biological makeup of children typically leads to their asthma becoming more severe compared to adults. Adults face different experiences of vulnerability because environmental factors from their past interact with existing allergies.¹¹

Gender

Studies show that asthma presentation and its intensity rates differ among male and female populations because of natural biological features and hormonal changes. Boys show a higher tendency to develop asthma than girls do in early life because their airway structures remain smaller than those of girls.¹² Asthma advances to become worse with increasing severity during adult years in women, particularly after puberty because of hormonal changes, especially fluctuations in the hormone estrogen participates as a key factor that modifies airway inflammation together with immune responses which affects asthma severity. The presence of estrogen in the body enhances both airways sensitivity to irritants and their inflammatory reaction. Scientific studies have suggested that hormonal changes lead to increased asthma severity in women during the adolescent and post-menopausal periods as the hormones move ups and downs during menstruation can activate asthma episodes which results in perimenstrual asthma.¹³ The condition of asthma that emerges during menstruation or within ten days before menstruation is called perimenstrual asthma.

Caring for First Degree Family Members with Asthma

The development of childhood asthma receives significant influence from genetic factors since asthma or allergic conditions affecting family members increase disease risk. A child risk to develop the disease becomes higher when they have allergic conditions including eczema and hay fever. A child with asthmatic parents has a higher chance of asthma development because their genes transmit immune system and inflammatory responses to their children.¹⁴ Additionally genetic predisposition can interact children with inherited asthma risk face the additional threat of developing asthma from environmental allergens and pollutants onset. Certain immune related and

inflammatory system genes create a risk factor for the disease. Exposure to allergens enhances the sensitivity of young airways which leads to continuous inflammation along with asthma manifestation.¹⁵ Children with allergic diseases in their family background along with genetic elements face an elevated asthma hazard in children.¹⁶

BMI (Body Mass Index)

The development and severity of asthma worsens when suffering from overweight or obesity because such weight conditions affect lung function and cause airway inflations. The extra body weight negatively affects lung functioning and simultaneously leads to elevated airway inflammation. Untreated obesity can impede normal lung capacity and worsens asthma symptoms because of reduced breathing capacity. Medical evidence shows obesity induced asthma tends to be more severe besides being harder to control and demonstrating a poor disease prognosis.¹⁷ Bronchoconstriction along with inflammatory responses develops due to obesity which intensifies asthma symptoms through airway narrowing. A long-term inflammatory condition creates worse asthma symptoms that resist treatment effectiveness which highlight the critical link between obesity and asthma severity.¹⁸

FEV1/FVC Ratio (Forced Expiratory Volume/ Forced Vital Capacity Ratio)

Asthma assessment depends on measuring airway blockages through the essential respiratory indicator FEV1/FVC ratio and lung function. The asthma related airway constriction leads to breathing hurdles because individuals with asthma show restricted breathing paths. The asthma diagnosis causes the FEV1/FVC ratio of asthmatic patients to become lower than ratios of non-asthmatic. Less than 0.70 ratio signifies obstructed airways thus establishing asthma as the diagnostic condition. The diagnosis separates asthma from restrictive lung diseases through this test according to some researchers.¹⁹ Better lung function accompanies higher ratios which show milder airway constriction patterns. Symptoms of an asthma episode usually led to a reduction in this ratio because breathing becomes harder. The test provides essential information that helps healthcare providers measure airway obstruction severity for proper decision making about treatment plans.²⁰

Allergen Exposure

The Air Quality Index (AQI) functions as a uniform metric for evaluating the density of airborne contaminants within the environment pollutants and their potential health effects. The monitoring tool tracks main pollutants including PM2.5, PM10, ground-level ozone (*Of*), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and carbon monoxide (CO). Long-term exposure to these pollutants leads to serious degradation of asthma conditions. Susceptible persons who experience exposure to symptoms will develop disease conditions. Pollutants like PM2.5 and when exposed to ozone the airways will become inflamed and more mucus develops while bronchial tubes contract. Multiple asthma attacks occur due to these pollutants according to the researchers.²¹ Additionally, the combination of contaminated air with current asthma patients leads to stronger asthma attacks which increases the risk of developing severe symptoms over time. The respiratory health of patients with allergic asthma suffers mainly due to their sensitivity to airborne substances including pollen and mold. Pollution exposures augment respiratory discomfort when pollutants connect with air components inside bronchial tissues. People must track air pollution levels as well as minimize their contact with polluted areas. People with asthma need to be exposed to reduced pollution zones to handle their symptoms and stop long-term health deterioration.²¹

AQI (Air Quality Index)

The Air Quality Index (AQI) used to evaluate pollution combines several airborne elements which include particulate matter (PM) along with nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO) and ground-level ozone (*Of*). Nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO), and ground level ozone (*Of*) in the atmosphere.²¹ Human air intake suffers degradation particularly in urban environments because of air pollution therefore creating lung damage, irritation and increased inflammation in individuals with asthma. People with asthma experience increased sensitivity toward pollution in the air because airborne contaminants lead to asthma attacks and provoke breathing difficulties and worsen their asthma condition.²² Long-term exposure to

PM2.5 and *Of* pollutants causes health problems which affect human bodies negatively. Chronic airway inflammation together with bronchial tube narrowing results from this condition while simultaneously causing more mucus production. The presence of asthma becomes more unmanageable combined with increased disease severity. Additionally long-term exposure people with genetic asthma risk may develop the condition after exposure to pollution. The safeguarding of respiratory health requires both monitoring Air Quality Index values and minimizing the exposure to pollutants.²³

Smoking Exposure

The combination of active smoking together with exposure to cigarette smoke in the environment deteriorates asthma symptoms while simultaneously damaging lung function and increasing asthma attack frequency.²³ Exposure to tobacco smoke along with smoking substantially worsens asthma symptoms and degrades lung function and triggers more asthma attacks. The chemicals present in tobacco smoke prove hazardous to health. The harmful chemicals in tobacco smoke produce irritation that triggers an increase of mucus production combined with airway inflammation and lung tissue destruction.²² Smoking causes an increase in asthma symptom intensity among patients who suffer from asthma. Smoking also causes bronchoconstriction while decreasing respiratory function so breathing becomes more difficult for patients. Secondly children face the highest danger from smoke exposure since it generates an increased likelihood of asthma development whereas it worsens asthma symptoms for those already diagnosed.²⁴ The condition becomes more severe in patients who already have a diagnosis. Asthmatic smokers encounter extreme attacks of the disease at higher rates than patients who do not smoke. The respiratory health suffers a destructive impact from smoking behaviors. Complete avoidance of smoking activities whether inhaled or exhaled should be a priority.²⁵

Physical Activity

Physical exercise causes dual effects on asthma control. People who have asthma should engage in moderate exercise because it helps their lungs grow stronger and makes their cardiovascular system healthier and their airways less inflamed

while simultaneously improving their symptom management. As a result of intense or prolonged exercise certain people experience asthma attacks more often in cold dry conditions. The medical term exercise-induced bronchoconstriction (EIB) refers to a respiratory condition which results in airway closure followed by breathing troubles following intense workouts.²⁵ An asthmatic individual can safely participate in exercise by warming up first and making their selections for activities and using prescribed inhalers to reduce the risk of attacks during or after physical activity.²⁶

Diet Quality Indices

A diet containing fruits along with vegetables and omega-3 fatty acids functions as an essential component to control asthma because it lowers inflammation and strengthens immunity. The antioxidants together with vitamins and minerals found in these foods both reduce airway inflammation and boost lung function which controls asthma better.²⁷ The combination of processed food consumption together with excess sugar and unhealthy fats in poor diet quality results in higher inflammation levels and worsened asthma symptoms and reduced respiratory disease resistance. The condition of obesity caused by unhealthful eating results in worsened asthma symptoms when it increases respiratory tract inflammation and makes asthma treatment complexities greater. Healthcare professionals use Geographic Information Systems (GIS) and Artificial Intelligence (AI) to locate asthma risk zones as well as speed up diagnostic processes for asthma patients. Research and healthcare providers utilize GIS to study asthma frequency patterns through air pollution and pollen readings and geographical element analysis thus identifying regions with elevated asthma incidence.²⁸ The subfields Machine Learning (ML) and Deep Learning (DL) within artificial intelligence allow researchers to perform advanced big data analysis which predicts asthma trends while simultaneously evaluating patient risk situations and uses combined environmental and physiological information for earlier disease diagnosis. AI and GIS integration produces more specific asthma forecasting and management capabilities that support the creation of focused intervention strategies to enhance patient results.²⁹

Literature Review

Asthma endotypes function as distinct asthma subtypes due to their unique biological mechanisms and they serve as fundamental elements for the development of individualized treatment plans. Phenotypes evaluate visible symptoms yet endotypes examine the particular pathophysiological mechanisms that cause asthma. Researchers together with clinicians use this identification to establish new therapeutic plans which demonstrate enhanced outcomes for individual patient populations. Applied endotype knowledge enables the creation of prevention methods that match each person's specific inflammatory pathophysiology along with hereditary risk factors. Defining asthma endotypes entails different interpretations that reduce the accuracy of asthma classification and prediction models developed for asthma research. The machine learning algorithm Random Forest provides excellent environmental modeling performance through its broad adoption for predicting groundwater potential and flood-prone areas and air pollution hotspot locations.⁸ Research on using the technique for asthma-prone area prediction stands at an early developmental phase. The relationship between asthma frequency and air quality together with weather conditions has been confirmed through studies which took place in New York and Italy and Japan. The analysis of environmental asthma triggers and asthma prevalence mapping has benefitted immensely from Geographic Information Systems (GIS). It was the researcher Peled who studied asthma behavior among children in Israel and the researcher Ahmad Khan who examined asthma in relation to vegetation patterns in Pakistan. The current research achieves its goal of estimating asthma-prone areas by uniting RF (Random Forest) modeling with GIS while analyzing environmental and physiological data metrics. The approach improves the control and prognostic aspects of asthma through its provision of spatial information which healthcare providers use to direct their intervention strategies.⁹

Deep learning through machine learning models enables autonomous pattern extraction from raw data without human intervention in choosing relevant features for disease diagnosis enhancement. Recent testing results show that DL

models achieve high diagnostic accuracy when identifying pulmonary tuberculosis as well as diabetic retinopathy and fibrotic lung disease.¹¹ Healthcare professionals believe that AI-based imaging solutions deliver significant value to medical facilities located in regions which have few radiologists.^{10,11,12} The field of respiratory medicine has benefitted from artificial intelligence which now supports assessments of lung cancer as well as pulmonary function test reading and detection of fibrotic lung diseases. AI diagnostic systems which earn FDA regulatory approval help doctors make faster and more precise medical diagnoses while shortening the timeframe needed for therapeutic decisions.¹³ AI faces several difficulties in healthcare that block its development through data bias and inadequate training datasets and problems related to model selection. The achievement of reliable safe AI systems depends on getting high quality data as well as effective computational infrastructure and close collaboration between clinicians and AI researchers and developers. Medical directors should lead the development of validation models to confirm AI systems operate with precision alongside human experts while remaining human independent. Advanced regulations along with ethical considerations serve as essential elements for mainstream healthcare AI integration which supports high standard diagnosis and maintains patient safety.¹⁴ The combination of AI and aerial data in respiratory medicine and air quality assessment will produce innovative asthma management approaches for Uttarakhand regions.

Asthma a prevalent non-communicable disease poses a significant health risk worldwide particularly in low and middle income countries. The low predictive accuracy of conventional methods drives researchers to employ Machine learning (ML) techniques for asthma diagnosis along with prediction and management functions.¹⁵ A variety of research projects utilized ML classifier systems to enhance both asthma identification and foretelling capabilities. The research found that XGBoost provided superior performance than Random Forest (RF) classifiers in adult asthma diagnosis reaching 81% accuracy and 85% AUC. The model performance reached new heights through Bayesian optimization because it optimized hyperparameters.¹⁶ Another study developed an interactive model for asthma health

assessment in African nations using XGBoost which delivered 90.6% accuracy superior to other methods including Decision Tree (90.49%) and KNN (90.28%).¹⁷ Researchers utilized XGBoost and LightGBM together with RF to detect asthma triggers and XGBoost proved most effective with 90.2% accuracy and 90.4% recall and 83.5% precision. The research data demonstrates that mobile health technology has the capability to detect asthma early while managing its course.¹⁸ The under sampling of non-asthma cases in youth asthma prediction using Logistic Regression produced an AUC of 0.7654 and achieved an F1-score of 0.3452. The research identified significant risk factors which included both low family poverty ratio and asthma background among parents.²¹ The prediction system for severe asthma exacerbations through home monitoring utilized XGBoost and One-Class SVM ML models together with Logistic Regression on daily peak expiratory flow data. The XGBoost (AUC 0.85–0.87) and Logistic Regression (AUC 0.86–0.90) delivered superior performance than standard clinical rules in this setting but still encountered issues because of insufficient event occurrences and excessive incorrect warnings.²²

Advances in medical science through artificial intelligence (AI) research have gained strong interest between medical practitioners and researchers for predicting asthma exacerbation outcomes. Medical expertise when combined with AI techniques demonstrates potential to generate cost effective methods for asthma diagnosis while also helping to predict patient risks for early intervention to improve results. The predictive models incorporate machine learning (ML) methods which increase both diagnostic precisions along with prediction reliability. Researchers currently demonstrate the effectiveness of ensemble machine learning methods when identifying asthma patients alongside evaluating their risk levels.²³ Ensemble methods constitute essential AI technology which joins predictive models into robust and accurate prediction solutions. The CatBoost classifier has established itself as a modern method which produces better results for asthma occurrence predictions and asthma risk grade evaluations. The proposed model achieved better results than standard methods Support Vector Machine (SVM)

and K-Nearest Neighbors (KNN) in combination with Logistic Regression and Adaboost classifier as well as Gradient Boosting classifier and Random Forest and Decision Trees. Amongst the model optimization techniques, the CatBoost model reached 93% accuracy by working with only 20% of features from a minimized feature set.²⁴

The diagnostic application of machine learning on individual asthma cases remains very scarce despite the recent research progress. Excellent opportunities exist for improving asthma diagnosis through machine learning because these methods allow better diagnostic procedures and more precise detection while decreasing human mistakes. Predicting asthma exacerbations is a fundamental research area within the field because it investigates the impact of ML techniques on this process.²⁵ ML methods enable researchers to examine environmental element relationships with asthma prevalence allowing better understanding of asthmatic disease triggers from environmental sources. ML algorithms help determine the essential factors that doctors need to diagnose asthma which include age characteristics plus gender factors together with BMI measurements and smoking histories and both indoor air quality and outdoor influences in the environment.²⁶ Large datasets can become the basis for ML methods to recognize patterns and detect relationships which standard clinical evaluations generally miss. The random forest classifier demonstrates better accuracy along with superior specificity and sensitivity and higher F1-score compared to standard logistic regression methods as other ML approaches. Research depicted the ML enhanced random forest model surpassing the diagnostic capabilities of the CAPE & CAPP models making machine learning algorithms a promising tool for surpassing traditional assessment methods according to research results.²⁸ The application of machine learning for asthma diagnosis and prediction shows promise but requires advanced research to add new learning methods and improve AI system diagnostic capabilities. Additional studies in this field need to enhance feature selection methods as well as optimize model parameters and increase dataset dimensions to construct more accurate diagnostic models for asthma and risk assessment.³⁰

MATERIALS AND METHODS

Data set description

To develop our model for asthma prediction we have mobilized a huge number of data consisting of a large number of variables from districts of Uttarakhand covering all the basic social, physiological and lifestyle factors. The independent variables to be employed in this study are age, gender, history of relation to first degree family members with asthma, BMI, FEV1/FVC ratio, allergens to which the individual has been exposed to AQI, smoking exposure, physical activity level and diet quality. These are some of the variables which have been considered to forecast the probability of asthma ailment.

Primary documents from public health and censuses were used to collect age and gender data in order to compare patient demographics. BMI and FEV1/FVC were obtained to explore the capacity of them to indicate the diagnosis of asthma and the severity of asthma. To account for heredity factors, family history of asthma was allowed as a candidate covariate. Smoking exposure information and physical activity and diet quality indices of participants were obtained to determine their overall impact on risk of developing asthma.

Air Quality Index (AQI) containing pollutant concentration data such as PM10, PM2.5, SO₂, CO, O₃ and NO₂, was collected from Uttaranchal Pollution Control Board. Information on allergen exposure along with other environmental parameters were collected and compared with demographic and health data. These parameters were collected on an hourly and daily basis to capture temporal and seasonality of water quality.

Table 1 using these variables we compiled a stable sample data that we used in training and testing machine learning algorithms, Random Forest, Log regression and XGBoost on the status of asthma. These models were selected because they can accommodate the interactions as well as the nonlinear relations between the variables. The diversity and precision of the dataset increase the robustness and generalizability of our results, enabling the discovery of asthma at risk populations and improved prediction capacity.

Machine Learning Algorithms

ML are useful in the diagnosis and

management of respiratory diseases such as asthma by providing tools to increase test reliability, better patients care delivery and address asthma's multifaceted character. This review is concerned with the percentage of accuracy in diagnosis, prediction, and control of asthma using machine learning models: the Random Forest, XGBoost, and Linear Regression models strengths and weaknesses.¹⁷

Machine Learning Algorithms in Asthma Diagnosis and Prediction

Asthma is a polygenic disorder that depends on multiple genes, environment, and clinical characteristics. The disorder usually necessitates the gathering of health history, clinical assessment, pulmonary function tests, and imaging. However, there are diagnostic difficulties which are as follows: Asthma symptoms are vague and it can mimic other respiratory disease and it is not clinically homogeneous.¹⁸ The application of machine learning provides a way of addressing these problems from a perspective of increased and correlated data dimensions with the resultant more accurate diagnostics and asthma risks predictions.

Random Forest (RF) in Asthma Prediction

RF is an ensemble learning method and its application to asthma prediction has been tried successfully because of its flexibility in unstructured data with nonlinear interactions and low sensitivity to overfitting. In special RF has been used to model asthma risk from various environment and physiological measures. Through the construction of several trees and combining the predictions RF is capable of providing better categorization of asthma patients as compared to existing models. The data sources used can include age, gender, BMI, air quality, smoking history, family history of asthma and the above makes RF models more accurate in their predictions. Therefore, the high tolerance of RF in large numbers of samples and many attributes can be useful for predicting asthma in various populations.²³

XGBoost in Asthma Risk Prediction

Another powerful algorithm is known as XGBoost or Extreme Gradient Boosting which has been used when predicting asthma. It is the most sophisticated gradient boosting algorithm which aims at handling decision trees and reducing errors tackled by previous models. XGBoost is known to perform well and again because of this aspect

it is suitable when dealing with a large dataset such as that of asthma. The algorithm operates where a strong model is made from a compilation of many weak models in this case decision trees. XGBoost has applied in the Asthma Study used to predict the risk of asthma attacks patients with asthma and assessment on how factors such as Pollution and Temperatures affect asthma. Based on the characteristics of asthma data XGBoost can effectively solve the problem of non-linear marginal of variables for asthma prediction.²⁴

Linear Regression in Asthma Prediction

Linear regression remains the easiest statistical estimation technique used in asthma prediction models. It can also be used to estimate the determining extent of the interdependent variable by one or many independent variables like age, weight, and environment and so on where the dependent variable is a continuous variable like severity of asthma or lung capacity. While Linear Regression is relatively simplistic compared to RF and XGBoost it gives models that are easy to explain and can open up insights into the linear connection between asthma predictors. Linear regression is also widely applied for quantifying effects of certain risk factors (such as BMI, smoking history or exposure to certain pollutants) on asthma.²⁵ Even though Linear Regression is still quite useful when the links between factors and responses are not complex and interpretability is an important factor.

Machine Learning in Childhood Asthma Prediction

The diagnosis of childhood asthma is even more complicated because the criteria used to diagnose the disease and specific biomarkers are not clearly defined. Most of the earlier statistical methods including logistic regression models that have been used earlier do not capture the interactions between genetic, environmental and clinical factors adequately. Criticisms such as Random Forest and XGBoost models are more flexible and provide higher accuracy in dealing with the maximum numbers of features like clinical, genetic, and exposome data. Some of these ML algorithms might reveal patterns that cannot be looked at by normal methods of asthma prediction models. For instance they can determine pre-pregnancy or lifestyle factors or constitutional

characteristics, which make the individual candidate for asthma and therefore more targeted early diagnosis and early management.²⁷

Proposed methodology

The main goal of this research is to classify asthma risk using different machine learning techniques including Random Forest, logistic regression and XG boost. The data being gather includes demographics and health data this information is then used to create models and assess their accuracy. In Table 1 is presented the research model and the step-by-step procedure for developing a risk prediction model for asthma.

Since the data collected from various organization which has variety of features of the data, preprocessing became inevitable to structure the data for analysis using Python. The methodology consists of the following steps:

Data Preprocessing

Preprocessing of the data simply make sure that the set data base is free from any form of errors or inconsistencies. The independent variables include: These factors include Age, Gender, Body Mass Index, Self-reported history of first- degree relative with Asthma, FEV1/FVC, Exposure to Allergens, Air quality Index, Exposure to cigarette Smoking, Physical Activity Level, and Diet Quality indicators.

Steps in preprocessing

Handling Missing Data

For categorical data set with numeric values, missing value were imputed with the mean or median equivalent of the concerned columns.

Outlier Detection

On the outliers, the standard statistical approaches such as Z-score or the IQR techniques were applied in order to reduce their effect on the model.

Normalization

For checkpoint 2 continuous variables including BMI, AQI, and FEV1/FVC Ratio were standardized through standard normalization in an attempt to increase model convergence.

Encoding Categorical Variables

Variables such as the gender and the relation of an individual to family members with asthma were converted to numerical forms using one-hot encoding or label encoding.

Data Splitting

The preprocessed dataset was towed into training and testing sets in order to have an independent validation of the proposed model.

Building the Automobile of the Prediction Model

To develop asthma risk prognostication models, the researchers employed ensemble learning approaches, as well as conventional methods. The models utilized the features mentioned above and followed these steps:

Feature Selection

Pearson correlation test p-values were used to determine highly significant predictors including FEV1/FVC Ratio, BMI, AQI and smoking exposure while a pair plot depicted the Feat-Fig movement of FEV1/FVC Ratio, BMI, AQI and smoking exposure.

Model Implementation

The three data mining techniques that was developed were Random Forest, Logistic Regression, and XGBoost. From these models different trends in the data were found including the effects of allergens in asthmatic risk or the role of diet quality indices in asthmatic vulnerability.

Model Evaluation

The applicability of the models was evaluated by comparing real asthma incidence data with the predictions made by the models. Multiple evaluation metrics were used:

Precision

The percentage of true positive cases among all positive predictions regarding asthma cases.

Recall (Sensitivity)

The possibility to recognize all real asthma patients.

F1-Score

The underline of harmonic mean between precision as well as the recall by covering both of them.

Accuracy

The ratio of correctly classified cases over the given set of data on a particular problem.

ROC-AUC Curve

Sensitivity and specificity are measures that are used because they can be compared against each other in different modality trade off studies.

Visualization and correlation analysis

Therefore, Figure 1 is a pair plot, which shows the correlation of all variables, including

BMI, FEV1/FVC Ratio, AQI, allergen exposure, and smoking exposure, to asthma severity. Higher BMI is negatively correlated with the FEV1/FVC ratio, indicating reduced lung capacity, airway narrowing, and increased asthma risk due to excess weight exerting pressure on the respiratory system. Explaining the models and how these relationships helped develop the models and give further understanding of other factors that lead to risk of asthma. By merging demographic, health and lifestyle-related variables into machine learning models this work presents a novel solution for identifying potential asthma risk.

The post regression coefficients of various independent variables and their interaction with asthma related characteristics: age, gender, relation to first-degree family members with asthma BMI, FEV1/FVC ratio, allergen exposure, AQI, smoking exposure, physical activity, and diet quality indices are presented in Figure 2 in the form of correlation coefficients quantifying the strengths and directions of linear relationships between the features. Physical activity helps reduce asthma risk by lowering BMI, improving lung function, and reducing systemic inflammation. Regular exercise strengthens respiratory muscles enhancing overall lung capacity and asthma control.

Each cell in the correlation matrix represents a correlation coefficient value ranging between -1.0 and 1.0:

- +1.0: Perfect positive relationship suggesting that if one variable goes up the other variable also goes up.
- -1.0: Shows right hand negative one, meaning that, there is perfect negative correlation, which means that as one variable goes up, the other goes down.
- 0: Shows non-directional relationship between two variables.

By analyzing this matrix, we identified significant associations among key predictors of asthma:

- BMI and FEV1/FVC Ratio: This was however noted to portray a very strong negative relationship where high BMI is likely to reduce lung capacity therefore increasing the chances of having asthma.
- Smoking Exposure and FEV1/FVC Ratio: The cross sectional exposure to smoking showed a negative association with the FEV1/FVC ratio supporting the primary emphasis on exposure to

cigarettes as a prominent determinant of respiratory health degree.

- AQI and Allergen Exposure: We found a moderate positive relationship between the two variables which pointed out that people being exposed to higher AQI levels are likely to come across allergens that trigger asthma.
- Physical Activity and BMI: Weight status was described with moderate negative correlation approaching significance higher physical activity may counteract asthma threat through lower BMI.
- Relation to First-Degree Family Members with Asthma and Asthma Risk: A positive association was compiled, hence, genetic factors being a potential cause of asthma.

Analyzing correlation coefficients between factors like BMI, FEV1/FVC ratio, and smoking exposure helps identify key predictors of asthma, understand their impact on lung function and develop targeted prevention strategies. In the study BMI and FEV1/FVC ratio showed the strongest negative correlation while smoking exposure and FEV1/FVC ratio also exhibited a significant negative association indicating their major role in worsening asthma severity.

This was done using Quantile-Quantile (Q-Q) plots as presented by Figure 3 to determine the distribution of independent variables; age, gender, relation with first-degree family members with asthma BMI, FEV1/FVC ratio allergen

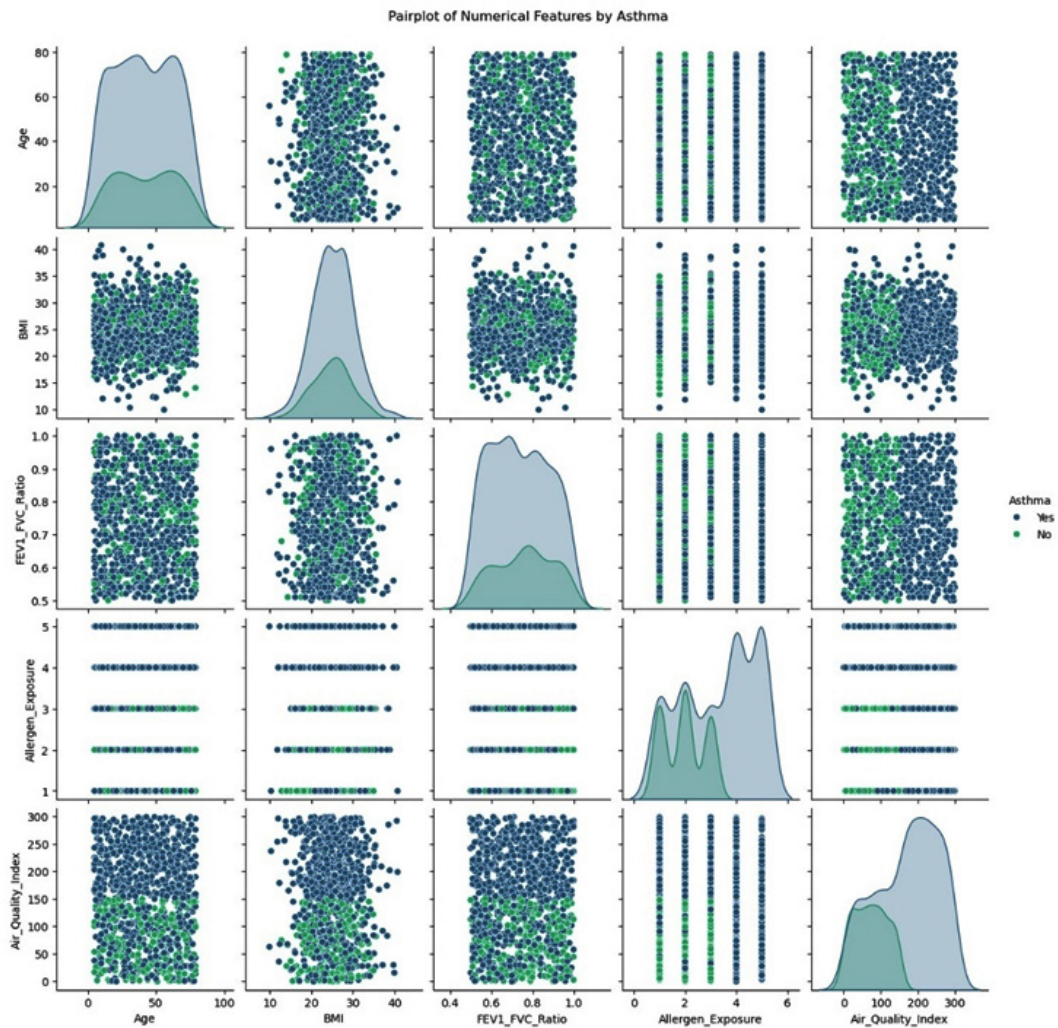


Fig. 1. Visualization of the various features of the dataset

Table 1. Data Set Description

	Age	Gender	Family_ History	BMI	FEV1_ FVC_ Ratio	Allergen_ Exposure	Air_ Quality_ Index	Smoking_ Exposure	Physical_ Activity	Diet_ Quality	Asthma
0	56	Male	No	23.9	0.60	4	108	No	Moderate	Moderate	Yes
1	19	Male	No	21.7	0.74	5	128	No	Moderate	Moderate	Yes
2	76	Female	No	25.8	0.87	5	9	No	High	Moderate	Yes
3	65	Male	No	27.2	0.75	2	46	No	Moderate	Moderate	No
4	25	Male	No	19.5	0.74	3	95	No	Moderate	Good	No
...
995	67	Male	No	26.8	0.84	5	229	Yes	Low	Moderate	Yes
996	26	Male	Yes	27.0	0.62	4	155	No	Moderate	Good	Yes
997	30	Male	No	15.4	0.88	4	196	Yes	Moderate	Poor	Yes
998	32	Female	No	23.6	0.88	1	222	No	Moderate	Moderate	Yes
999	53	Male	Yes	29.2	0.66	5	249	No	Low	Moderate	Yes

exposure, AQI, smoking exposure, physical activity and diet quality indices. Quantile-Quantile (Q-Q) plots help assess whether a variable follows a specific distribution by comparing its quantiles to those of a theoretical distribution (e.g., normal distribution). This is crucial for statistical and machine learning methods as many models assume normality, affecting accuracy and performance. This step was necessary in order to understand if the data on each variable is normally distributed which is often a strict prerequisite when working with a number of methods in statistics and machine learning.

For other continuous variables including age, BMI, and FEV1/FVC, the Q-Q plots suggested the existence of asymmetry or outliers at the ends of the lines. Continuous variables like age, BMI, and FEV1/FVC were transformed using normalization or log transformations to address asymmetry and outliers ensuring a more normal distribution. Categorical variables like gender and family history of asthma were encoded using one-hot encoding or label encoding to facilitate their inclusion in machine learning models. These observations indicated that there is a need for either transformations or high powerful statistical methods to deal with non-normality. Allergen exposure and AQI seemed not to need further transformation because they fitted better in the normal distribution of the variable.

As for the gender and relation to first-degree family members with asthma as well as other categorical variables Q-Q plots could not be performed because of the discrete nature of such variables. However these variables were treated accordingly to be used in the machine learning models that follow later.

Figure 4 shows that box plots were used to display the reflection and visualize the probability of outliers in different categories of some selected independent variables which are age, gender, first-degree family members with asthma, BMI, FEV1/FVC ratio, Allergen exposure, AQI, smoking exposure, physical activity and diet quality indices. Box plots helped identify outliers and skewness by visually displaying the distribution of variables like age, BMI, and smoking exposure, highlighting extreme values beyond the whiskers. Skewed distributions were detected through asymmetric box shapes and longer whiskers, indicating

deviations from normality. In this method, it was easier to notice the distribution of the data, mean, range and variance, which enabled the identification of outliers if the data was skewed or if the range of values shuttled between narrow or wider intervals.

The distribution of ages and BMI numbers was skew to the right and box plots showed that

the distribution was overall pretty normal with only few extreme outliers in terms of age and extreme BMI values. Some of these outliers were explored to find out whether the differences were arising from data entry errors or from observations of truly extraordinary events. Likewise, the FVC and smoking exposure data also showed a few outliers which could be due to unusual readings and other possibilities that may deserve different consideration in the course of modeling.

The distribution of the AQI allergen exposure and physical activity variables was slightly more symmetrical with relatively few outliers beyond the interquartile range. Although, diet quality indices of SCCM participants had only slight skewness toward the higher values and higher values were skewed towards the middle of distribution in quartile distribution of the diet quality scores.

As expected, the distributions of the gender and relation to first-degree family members variables were clear with gender being two-group (male, female) and the relation to family members with asthmatic parents (1), without asthmatic parents (0).

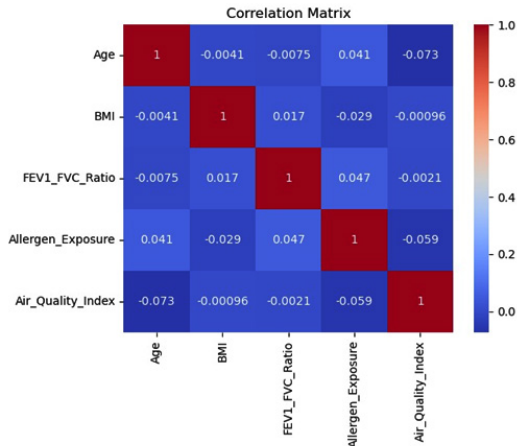


Fig. 2. Correlation Matrix between various features of data set

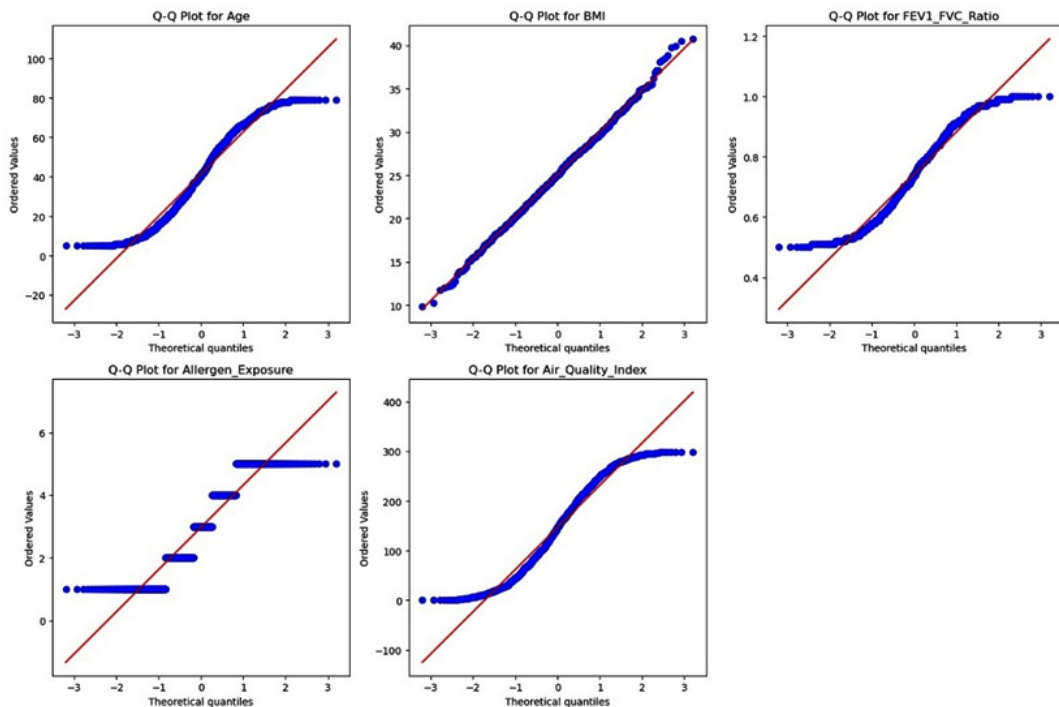


Fig. 3. Quantile-Quantile (Q-Q) plots for independent variables of the data set

RESULTS

The descriptive plots in Figures 1–4 was based on independent variables such as age, gender, family history of asthma, BMI, FEV1/FVC ratio, allergens, AQI, smoking exposure, physical activity, and diet quality indices. The dataset was split into training and testing segments with a 70:30 ratio allowing the majority of the data to be used for model training. Logistic Regression,

Random Forest, and XGBoost algorithms were implemented to detect asthma status using these features. Random Forest and XGBoost emerged as the most effective models, achieving 99% accuracy and 98% ROC scores. Logistic Regression demonstrated relatively lower performance with 85% accuracy and a 94% ROC score. Confusion matrices and classification reports in Figures 5, 6, and 7 highlighted the model performances across these algorithms.

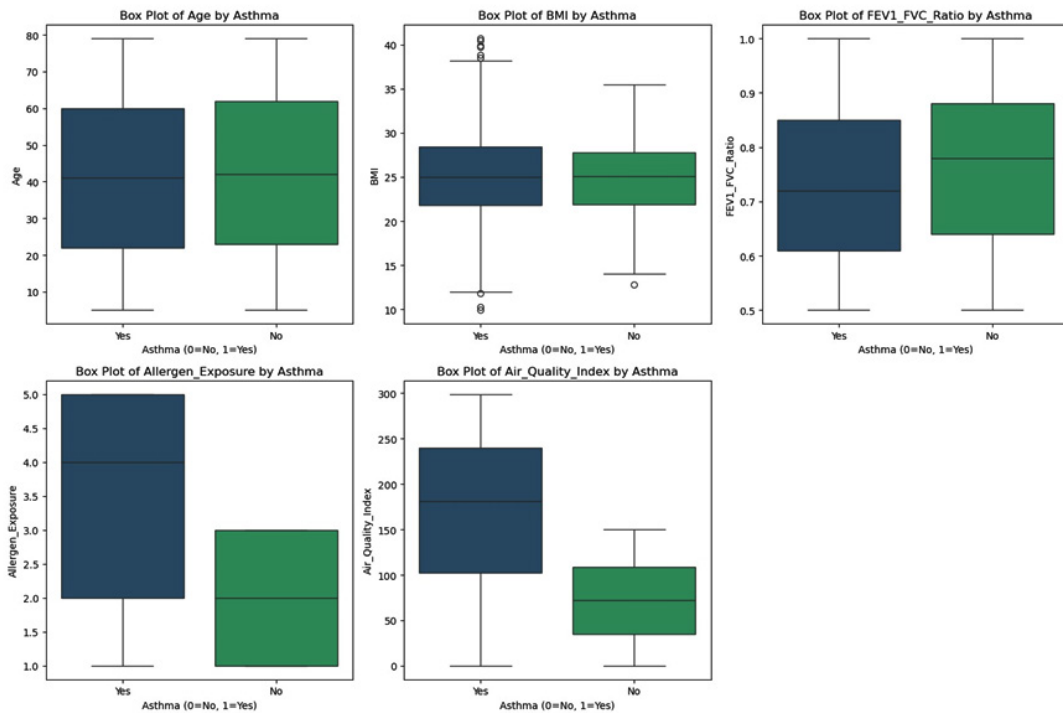


Fig. 4. Box plots for independent variables of the data set

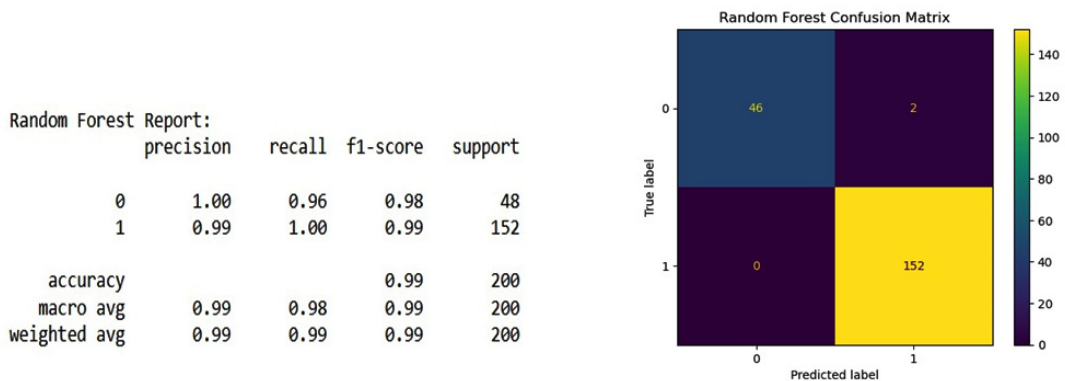


Fig. 5. Classification Report and Confusion Matrix of Random Forest

ROC curves in Figure 8 provided insights into the predictive performance of each variable demonstrating the sensitivity and specificity of the models. Variables such as BMI, FEV1/FVC ratio, allergens, and AQI showed significant predictive

capabilities. Among categorical variables age displayed a complex curvilinear relationship with asthma susceptibility while gender showed moderate predictive power. Smoking exposure and family history of asthma yielded high

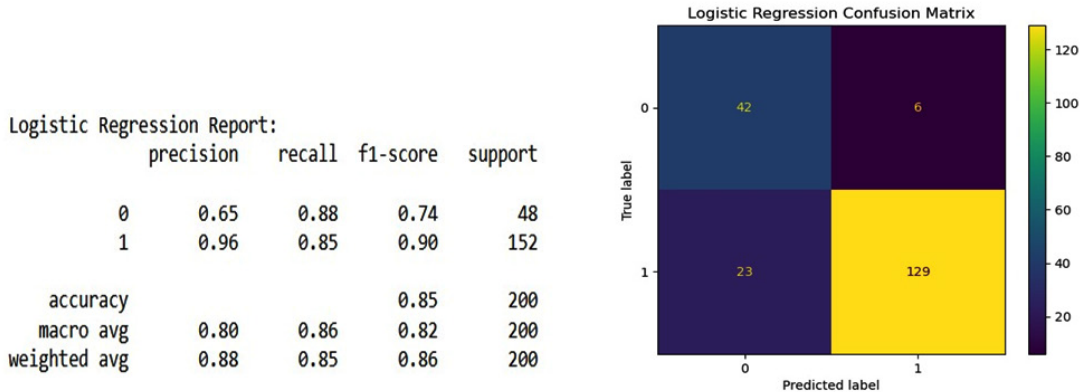


Fig. 6. Classification Report and Confusion Matrix of Logistic Regression

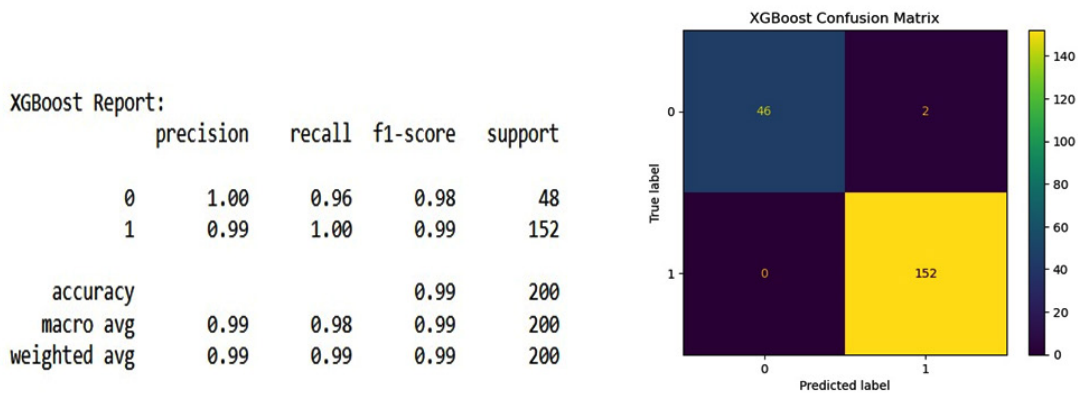


Fig. 7. Classification Report and Confusion Matrix of XGBoost

Table 2. Performance matrices

Model	class	Precision	Recall	F1-score	Support
Logistic Regression	0	0.65	0.88	0.74	48
	1	0.96	0.85	0.90	152
	Accuracy- 85%				
Random Forest	class	Precision	Recall	F1-score	Support
	0	1.00	0.96	0.98	48
	1	0.99	1.00	0.99	152
Accuracy- 99%					
XGBoost	class	Precision	Recall	F1-score	Support
	0	1.00	0.96	0.98	48
	1	0.99	1.00	0.99	152
Accuracy- 99%					

sensitivity indicating their strong influence on asthma prediction. Physical activity demonstrated a protective effect, and higher diet quality was linked to reduced asthma risk.

Our data set has 2 different target classes. Table 2 describes the comparison of Logistic Regression, Random Forest and XGBoost.

Histograms in Figure 9 depicted the distribution of variables including age, gender, BMI, FEV1/FVC ratio, allergens, AQI, smoking exposure, physical activity, and diet quality indices providing a clear visual representation of their trends and outliers. AUC values calculated for individual features revealed their predictive power with values closer to 1 indicating stronger predictive ability.

DISCUSSION

The findings highlighted the importance of Random Forest and XGBoost in developing

reliable predictive models for asthma outperforming Logistic Regression. The confusion matrices demonstrated true and false predictions across asthma severity classes while classification reports evaluated accuracy, recall, and F1 scores. The ROC curves provided a deeper understanding of each variable's contribution to asthma prediction revealing significant roles for environmental, genetic, and lifestyle factors.

BMI, allergens, and AQI emerged as strong predictors showcasing their direct association with asthma. Age demonstrated a nuanced relationship with specific age brackets being more susceptible while gender provided moderate predictive capability. Smoking exposure was statistically significant with higher exposure correlating to increased asthma probability. Similarly, family history of asthma played a crucial role in prediction confirming its genetic influence.

Lifestyle factors such as physical activity and diet quality showed protective

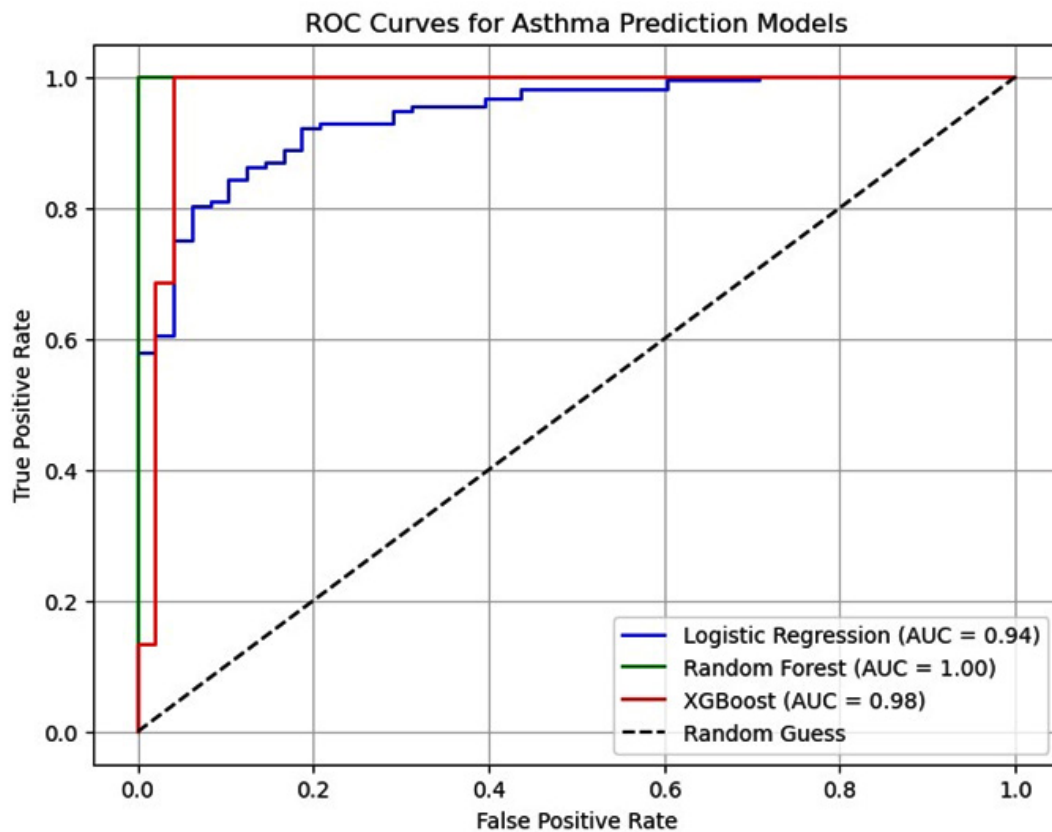


Fig. 8. Receiver Operating Characteristic (ROC) to evaluate asthma risk

effects against asthma. The study found that higher physical activity was negatively correlated with BMI suggesting that maintaining an active lifestyle may help lower asthma risk by reducing obesity-related inflammation and improving lung function. Additionally, a high quality diet rich in antioxidants, vitamins, and omega-3 fatty acids was associated with reduced airway inflammation potentially mitigating asthma symptoms and severity. Active individuals were at lower risk while healthier diets reduced asthma probability. Despite Random Forest and XGBoost demonstrating superior performance, the study suggested refining these models through hyperparameter tuning and feature importance analysis to enhance predictive accuracy further.

The ROC analysis and histograms provided a comprehensive understanding of the variables influence on asthma risk. The AUC values supported the development of a composite ensemble model that integrated the strengths of

individual features to improve asthma prediction. The study emphasized the importance of these independent variables in creating robust models and offered insights for future research in asthma prediction and management in Uttarakhand.

Challenges in Using Machine Learning for Asthma Prediction

Despite the promising results from machine learning models challenges remain in their widespread adoption for asthma prediction:

Data Quality and Availability

It is sometimes possible to achieve a high accuracy using a trained ML model but the ability to generate such a model requires a large high-quality dataset. Imperfect, noisy, and more or less biased data can decrease the efficiency and credibility of asthma models.

Algorithm Interpretability

Algorithms such as Random Forest and XGBoost are very strong for classification but they are very hard to interpret and understand how the

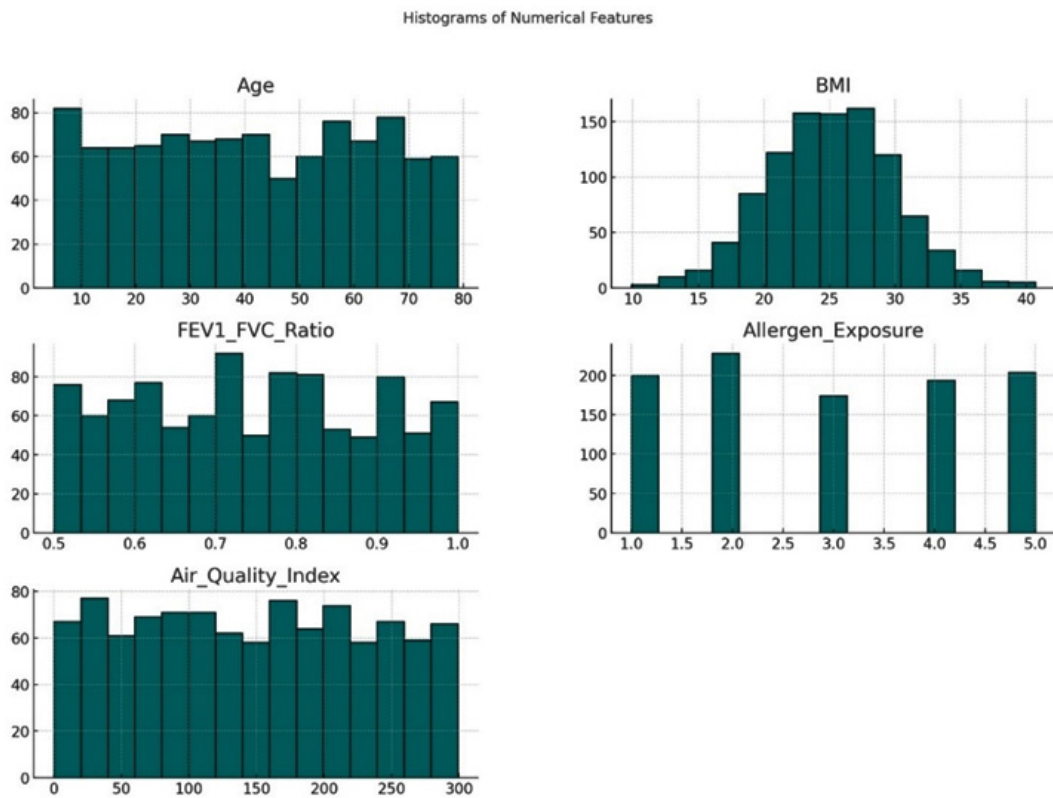


Fig. 9. Histograms to visualize the impact of variable for asthma prediction

inputs are affecting the model. Due to this absence of interpretability these models cannot be used in clinical settings with many applications where they need to be interpreted to arrive at certain decisions.

Computational Requirements

Several of the more sophisticated algorithms for machine learning for instance XGBoost demand a lot of computational power which can be hard to come by in a healthcare facility.

Integration of Multiple Data Sources

The environmental, genetic, as well as clinical data were used for the evaluation of asthma which has been studied widely. Despite this these disparate sources of data need to be combined into a single model for more accurate prediction work while presenting some difficulties.

Apart from diagnostic capability machine learning also enables disease management in asthma. With regard to asthma patient tracking AI-based tools can predict frequent relapses by using environmental and physiological data to facilitate intervention. For instance RF and XG Boost are used in making predictions of asthma attacks considering data on the quality of air the use of prescription and other characteristics regarding the patient. Such predictions assist physicians in individualized therapy decision-making drugs dosage ritual and general diseases management.⁵

Challenges in data quality such as missing values inconsistent records and biases can reduce the accuracy and reliability of machine learning (ML) models for asthma prediction. Algorithm interpretability is another major concern as complex models like deep learning often function as “black boxes” making it difficult for clinicians to understand how predictions are made. Additionally, high computational requirements can limit the accessibility of ML models in resource constrained healthcare settings slowing real time decision making and widespread adoption in clinical practice. Addressing these issues requires robust data preprocessing the use of explainable AI techniques and optimizing computational efficiency.

CONCLUSION

This study demonstrates the significant potential of machine learning techniques in

improving asthma diagnosis and prognosis by incorporating various environmental, physiological and lifestyle factors. The analysis using independent variables such as age, gender, family history of asthma, BMI, lung function, allergen exposure, AQI, smoking, physical activity, and diet quality revealed the effectiveness of machine learning models in accurately predicting asthma status. Among the models evaluated Random Forest and XGBoost demonstrated exceptional performance with accuracy rates of 99% and ROC scores of 98%, highlighting their strong predictive capability. In contrast Logistic Regression exhibited relatively lower accuracy (85%) and ROC scores (94%), indicating that more complex models provide better results in asthma prediction. These findings underline the importance of integrating multiple factors and leveraging machine learning algorithms for more accurate and timely asthma diagnosis ultimately contributing to better patient outcomes and management.

Future research can explore expanding the dataset to include more diverse and comprehensive samples, enhancing the generalizability of the models. Additionally, advanced techniques such as transfer learning could be incorporated to improve the accuracy of predictions especially in low-resource settings where access to large datasets may be limited. Further work can also investigate the integration of real-time environmental data wearable health monitors and more granular lifestyle information to create dynamic models that can offer personalized asthma management. By refining these models and incorporating more advanced methodologies it is possible to develop even more effective strategies for asthma prevention, diagnosis, and treatment ultimately improving the quality of life for individuals affected by this chronic condition.

Acknowledgement

The author would like to thank Graphic Era Hill University, Bhimtal campus, for providing the necessary resources, facilities, and a conducive environment for completing the research work. The first author is also thankful to her coauthor, Dr. Janmejay Pant, for his support and guidance in framing this research.

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The author(s) do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article.

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials.

Permission to reproduce material from other sources

Not Applicable

Authors' Contribution

Priyanshi Kotlia: Conceived and designed the research study, developed the methodology, Writing – Original Draft; Manoj Chandra: Contributed to the interpretation of the results, ensured the technical rigor of the analysis; Janmejay Pant: Worked on machine learning models and assisted in fine-tuning the algorithms, Data Collection, conducted data preprocessing, assisted with visualization

REFERENCES

- Bateman ED, Hurd SS, Barnes PJ, et al. Global strategy for asthma management and prevention: GINA executive summary. *Eur Respir J.* 2008;31(1):143-178. doi:10.1183/09031936.00138707.
- Jackson N, Sajuthi S, Rios C, et al. Machine learning analysis of airway transcriptomic data identifies distinct childhood asthma endotypes. In: *C12. C012 OMICS IN ASTHMA AND COPD*. American Thoracic Society; 2021:A1151-A1151.
- Mersha TB, Afanador Y, Johansson E, et al. Resolving clinical phenotypes into endotypes in allergy: Molecular and omics approaches. *Clin Rev Allergy Immunol.* 2021;60(2):200-219. doi:10.1007/s12016-020-08787-5
- Swain S, Bhushan B, Dhiman G, Viriyasitavat W. Appositeness of optimized and reliable machine learning for healthcare: A survey. *Arch Comput Methods Eng.* 2022;29(6):3981-4003. doi:10.1007/s11831-022-09733-8.
- Hoeksema LJ, Bazzzy-Asaad A, Lomotan EA, et al. Accuracy of a computerized clinical decision-support system for asthma assessment and management. *J Am Med Inform Assoc.* 2011;18(3):243-250. doi:10.1136/amiajnl-2010-000063.
- Fathima M, Peiris D, Naik-Panvelkar P, Saini B, Armour CL. Effectiveness of computerized clinical decision support systems for asthma and chronic obstructive pulmonary disease in primary care: a systematic review. *BMC Pulm Med.* 2014;14(1):189. doi:10.1186/1471-2466-14-189.
- Prasad BDCN, Prasad PESNK, Sagar Y. A comparative study of machine learning algorithms as expert systems in medical diagnosis (asthma). In: *Advances in Computer Science and Information Technology*. Springer Berlin Heidelberg; 2011:570-576.
- Badnjevic A, Cifrek M, Koruga D, Osmankovic D. Neuro-fuzzy classification of asthma and chronic obstructive pulmonary disease. *BMC Med Inform Decis Mak.* 2015;15 Suppl 3(S3):S1. doi:10.1186/1472-6947-15-S3-S1.
- Zanolin ME, Pattaro C, Corsico A, et al. The role of climate on the geographic variability of asthma, allergic rhinitis and respiratory symptoms: results from the Italian study of asthma in young adults. *Allergy.* 2004;59(3):306-314. doi:10.1046/j.1398-9995.2003.00391.x.
- Diwakar M, Singh P, Shankar A. Multi-modal medical image fusion framework using co-occurrence filter and local extrema in NSST domain. *Biomed Signal Process Control.* 2021;68(102788):102788. doi:10.1016/j.bspc.2021.102788.
- Maantay J. Asthma and air pollution in the Bronx: methodological and data considerations in using GIS for environmental justice and health research. *Health Place.* 2007;13(1):32-56. doi:10.1016/j.healthplace.2005.09.009.
- Gorai A, Tuluri F, Tchounwou P. A GIS based approach for assessing the association between air pollution and asthma in New York State, USA. *Int J Environ Res Public Health.* 2014;11(5):4845-4869. doi:10.3390/ijerph110504845.
- Khan I. Spatial association of asthma and vegetation in Karachi: a GIS perspective. *Pakistan Journal of Botany.* 2010;42:3547-3554.
- Škarková P, Kadlubiec R, Fischer M, Kratinová

- J, Zapletal M, Vrabel J. Refining of asthma prevalence spatial distribution and visualization of outdoor environment factors using gis and its application for identification of mutual associations. *Cent Eur J Public Health*. 2015;23(3):258-266. doi:10.21101/cejph.a4193.
15. Howard R, Rattray M, Prospero M, Custovic A. Distinguishing asthma phenotypes using machine learning approaches. *Curr Allergy Asthma Rep*. 2015;15(7):38. doi:10.1007/s11882-015-0542-0.
 16. Spathis D, Vlamos P. Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health Informatics J*. 2019;25(3):811-827. doi:10.1177/1460458217723169.
 17. Razavi-Termeh SV, Sadeghi-Niaraki A, Choi SM. Effects of air pollution in Spatio-temporal modeling of asthma-prone areas using a machine learning model. *Environ Res*. 2021;200(111344):111344. doi:10.1016/j.envres.2021.111344.
 18. Azeta AA, Edward CS, Ekpo RH. Interactive machine learning framework for predicting asthma health conditions using XGBoost. In: *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*. IEEE; 2024:849-856.
 19. Yadav S, Sehrawat H, Jaglan V, et al. A novel effective forecasting model developed using ensemble machine learning for early prognosis of asthma attack and risk grade analysis. *Scalable Comput Pract Exp*. 2025;26(1):398-414. doi:10.12694/scpe.v26i1.3758.
 20. de Hond AAH, Kant IMJ, Honkoop PJ, Smith AD, Steyerberg EW, Sont JK. Machine learning did not beat logistic regression in time series prediction for severe asthma exacerbations. *Sci Rep*. 2022;12(1):20363. doi:10.1038/s41598-022-24909-9.
 21. Ray A, Das J, Wenzel SE. Determining asthma endotypes and outcomes: Complementing existing clinical practice with modern machine learning. *Cell Rep Med*. 2022;3(12):100857. doi:10.1016/j.xcrm.2022.100857.
 22. Xie M, Xu C. Predicting the risk of asthma development in youth using machine learning models (preprint). *JMIR Preprints*. Published online 2024. doi:10.2196/preprints.66948.
 23. Zein JG, Wu CP, Attaway AH, Zhang P, Nazha A. Novel machine learning can predict acute asthma exacerbation. *Chest*. 2021;159(5):1747-1757. doi:10.1016/j.chest.2020.12.051.
 24. Rani A, Sehrawat H. Role of machine learning and random forest in accuracy enhancement during asthma prediction. In: *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. IEEE; 2022.
 25. Prospero MC, Marinho S, Simpson A, Custovic A, Buchan IE. Predicting phenotypes of asthma and eczema with machine learning. *BMC Med Genomics*. 2014;7 Suppl 1(Suppl 1):S7. doi:10.1186/1755-8794-7-S1-S7.
 26. Wasylewicz ATM, Scheepers-Hoeks AMJW. Clinical decision support systems. In: *Fundamentals of Clinical Data Science*. Springer International Publishing; 2019:153-169.
 27. Khalid MS, Haque MA, Hossain MS. A belief rule-based decision support system for assessing clinical asthma suspicion. In: *Proceedings of Scandinavian Conference on Health Informatics. Grimstad*: Linköping University Electronic Press; 2014:83-89.
 28. Peled R, Reuveni H, Pliskin JS, Benenson I, Hatna E, Tal A. Defining localities of inadequate treatment for childhood asthma: a GIS approach. *Int J Health Geogr*. 2006;5:3. doi:10.1186/1476-072X-5-3.
 29. Tomita K, Yamasaki A, Katou R, et al. Construction of a diagnostic algorithm for diagnosis of adult asthma using machine learning with random forest and XGBoost. *Diagnostics (Basel)*. 2023;13(19). doi:10.3390/diagnostics13193069
 30. Wong A. Machine learning models for preventative mobile health asthma control. *J Asthma*. Published online 2025:1-9. doi:10.1080/02770903.2025.2453812