VOL. 11 NO. 9,
FEBRUARY, 2026
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

Developing Enhanced Asthma Severity Predictive Model using Random Forest Algorithm

Aminu Ibrahim[1]; & Yusuf Musa Malgwi[2]

[1]Address No. 128 Palace Way, Jalingo. [2]Department of Computer Science, Faculty of Science, Modibbo Adama University of Technology, P.M.B. 2076, Yola Nigeria.
Corresponding Author: aminumalam19@gmail.com
DOI: https://doi.org/10.70382/ajsitr.v11i9.072

Abstract

Asthma is a chronic respiratory condition shaped by environmental, physiological, and behavioral factors. Accurate prediction of asthma severity is vital for personalized care and reducing exacerbations. While Machine Learning (ML) has been widely explored in asthma prediction, many existing models lack generalizability, robustness, or comprehensive integration of multiple predictors, limiting their clinical applicability. This study presents a robust ML-based model for classifying asthma severity by incorporating diverse patient and environmental features. A supervised learning approach was employed using a publicly available dataset of 1,010 records with 14 features, including demographics, clinical symptoms, and environmental indicators (temperature, wind speed, and humidity). The dataset was pre-processed and stratified to balance severity classes. Three models—Decision Tree, Support Vector Machine (SVM), and Random Forest (RF) were evaluated using standard metrics: accuracy, precision, recall, F1-score, and ROC AUC. Among them, the RF model showed superior performance, achieving 96.70% accuracy, a 0.9668 F1-score, and a 0.9831 ROC AUC. Feature importance analysis highlighted environmental factors, particularly temperature and humidity, as key predictors of asthma severity. These results underscore RF's effectiveness in providing accurate, interpretable predictions and addressing limitations of earlier models. The proposed model offers a data-driven framework for real-time severity forecasting, supporting early interventions and personalized treatment. It holds promise for integration into clinical decision support systems, thereby enhancing asthma management and optimizing healthcare resource use. It is therefore recommended that the proposed model be operationalized within clinical triage protocols, embedded into

electronic health record (EHR) infrastructures, or integrated into mobile health (mHealth) applications to facilitate data-driven, proactive asthma management across diverse care settings in Nigeria.

Keywords: Asthma Severity, Machine Learning, Random Forest, Predictive Modelling, Healthcare Analytics

Introduction

Asthma is a chronic respiratory disease affecting millions of individuals worldwide, making it a significant public health concern (Enilari & Sinha, 2019; Serebrisky & Wiznia, 2019). The disease is characterized by airway inflammation and hyper reactivity, leading to symptoms such as wheezing, shortness of breath, chest tightness, and coughing. These symptoms can vary in severity and frequency, often significantly impacting the quality of life for those affected. The prevalence of asthma has been steadily increasing, with factors such as urbanization, environmental pollution, and lifestyle changes contributing to this rise (Louis *et al.*, 2021). The World Health Organization (WHO) estimates that asthma affects over 235 million people globally, and the disease burden is particularly high among children and young adults (Charles *et al.,* 2020).

Effective management of asthma involves a multifaceted approach that includes regular monitoring, adherence to prescribed medication, and timely intervention to prevent exacerbations (Chan *et al.,* 2022). Asthma exacerbations, which are acute episodes of worsening symptoms, can lead to emergency department visits and hospitalizations, posing a substantial burden on healthcare systems (Sowah *et al.,* 2021). Traditional management strategies emphasize the importance of patient education, environmental control, and the use of Inhaled Corticosteroids (ICs) and bronchodilators (Juma & Mathuva, 2023). Despite these measures, many patients continue to experience poor asthma control and frequent exacerbations, highlighting the need for improved management strategies (Tan *et al.,* 2024).

In response to this persistent challenge, there is a growing recognition of the need to explore the various factors that influence asthma severity, including environmental exposures, psychological stressors, physiological conditions, and behavioral patterns. Accordingly, this study first seeks to determine the primary environmental, psychological, physiological, and behavioural factors contributing to asthma severity, thereby providing a foundational understanding of the multidimensional triggers that exacerbate the condition. These insights will be critical in ensuring that any predictive model developed is both relevant and effective in real-world contexts.

Advancements in Machine Learning (ML) have opened new avenues for enhancing healthcare outcomes, particularly in the management of chronic diseases like asthma. ML techniques involve the development of algorithms that can learn patterns from data,

and they have shown promise in various healthcare applications (Ren *et al.,* 2022). These techniques can predict disease progression, personalize treatment plans, and improve patient outcomes by providing actionable insights to healthcare providers (Sun *et al.,* 2022). For instance, ML algorithms can analyze large volumes of patient data to identify risk factors for asthma exacerbations, predict the likelihood of future attacks, and recommend personalized treatment adjustments (Yi *et al.,* 2021).

To leverage these capabilities, the present study is designed to gather relevant clinical data to establish a correlation between these contributing factors and asthma severity. By collecting well-structured, diverse, and high-quality datasets, the study aims to capture the variability of patient experiences and contextual influences on disease progression. These data will serve as the foundation for building a robust predictive model.

In the context of asthma, ML can be utilized to forecast the severity of the disease and predict treatment outcomes, thereby enabling healthcare providers to tailor interventions more effectively. By integrating data from electronic health records (EHRs), wearable devices, and patient-reported outcomes, ML models can provide a comprehensive assessment of a patient's condition. These models can identify patterns and correlations that may not be apparent through traditional analysis, offering a more nuanced understanding of asthma dynamics (Menendez *et al.,* 2016). Consequently, ML-driven approaches have the potential to enhance decision-making processes, reduce the incidence of severe asthma attacks, and ultimately improve the quality of life for patients with asthma (Kraft *et al.,* 2021).

Building upon this foundation, the study aims to implement the Random Forest algorithm for predicting asthma severity, given its effectiveness in handling high-dimensional data and its ability to capture non-linear relationships among variables. The model development follows a structured pipeline that involves training and testing the algorithm using a publicly available dataset of asthma patients. This approach ensures that the predictive engine remains both generalizable and tailored to the specific characteristics of the target population.

The application of ML in asthma management is supported by numerous studies that demonstrate its efficacy. For example, ML models have been used to predict asthma exacerbations with high accuracy by analyzing variables such as medication usage, environmental factors, and physiological measurements (Ciudad-Gutiérrez *et al.,* 2021). Additionally, personalized treatment plans generated by ML algorithms have been shown to improve medication adherence and reduce symptom severity. A study by Zein *et al.,* (2021) used logistic regression and decision trees to predict exacerbations based on demographic, clinical, and environmental data, achieving an accuracy of 85%. Another study by Lovrić *et al.,* (2021) employed a random forest algorithm to predict optimal medication regimens, resulting in improved symptom control and reduced exacerbations compared to standard guidelines.

To ensure the reliability and clinical relevance of the model, its performance is rigorously evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These

metrics provide a detailed view of the model's classification performance, supporting its readiness for practical deployment in predictive asthma care.

These findings underscore the potential of ML to transform asthma care by providing precise, data-driven insights that support proactive and personalized management strategies. Through the achievement of the outlined objectives, this study aims to contribute a novel, evidence-based, and technically robust solution to the ongoing challenge of asthma severity prediction.

Statement of the Problem

Asthma remains a prevalent and challenging condition to manage, especially among pediatric populations where variability in disease presentation and response to treatment is high (Licari *et al.,* 2018). Despite significant advances in treatment options and management strategies, predicting asthma severity and treatment outcomes remains problematic (Lovrić *et al.,* 2021). Current approaches often rely on generalized guidelines that fail to account for individual differences, leading to suboptimal treatment outcomes and increased healthcare burdens (Do *et al.,* 2019).

The need for personalized asthma management strategies is increasingly pressing. ML offers a promising solution by leveraging large datasets to identify patterns and predict outcomes tailored to individual patients. While many studies have applied ML for asthma diagnosis and risk prediction (Yoon *et al.,* 2022), most rely on static models. There remains a lack of dynamic, personalized exacerbation prediction attacks for asthma management (Tsang *et al.,* 2020).

This study aims to address these challenges by developing robust ML-based prognostic tools for forecasting asthma severity. By incorporating diverse data sources and focusing on pediatric patients, this research seeks to improve the accuracy of predictions and enhance personalized treatment plans. This approach has the potential to optimize clinical decision-making, improve patient outcomes, and reduce the overall burden on healthcare systems (Tiotiu, 2021).

Aim and Objectives of the Study

The study aim to develop an enhanced asthma severity predictive model using Random Forest algorithm, and the objectives are to:

i. identify and analyze key environmental, psychological, physiological, and behavioural factors influencing asthma severity.
ii. gather and preprocess relevant publicly available clinical datasets to explore correlations between these factors and asthma severity.
iii. implement a suitable machine learning algorithm (Logistic Regression, Random Forest, and Support Vector Machine) for predicting asthma severity.
iv. train and test the model using the pre-processed dataset, incorporating appropriate feature engineering techniques.

v. evaluate the model's performances using standard metrics such as accuracy, precision, recall, and F1-score.

Significance of the Study

The application of machine learning in asthma management has the potential to revolutionize how the disease is monitored and treated. By providing accurate predictions of asthma severity and treatment outcomes, ML models can support healthcare providers in making informed decisions, leading to better patient care. Personalized treatment plans derived from ML predictions can enhance the efficacy of interventions, reduce the frequency of exacerbations, and improve the overall quality of life for asthma patients.

Furthermore, this study contributes to the growing body of literature on the use of ML in healthcare, providing insights that can be applied to other chronic diseases. The findings from this research can also inform the development of predictive tools and decision support systems for asthma management.

Scope of the Study

The scope of this research encompasses the development of a machine learning model to predict asthma severity and treatment outcomes, with a focus on personalization. It includes collecting and integrating diverse datasets from electronic health records, public health databases, and wearable devices. The study involves thorough data preprocessing, such as handling missing values and normalizing variables. It also covers the development and optimization of the ML model through hyperparameter tuning and extensive evaluation to ensure accurate and individualized predictions. This research aims to enhance clinical decision-making and improve patient outcomes, particularly in pediatric populations.

Definition of Operational Terms

1. ***Asthma Severity***: The level of intensity of asthma symptoms experienced by an individual, categorized as mild, moderate, or severe. It reflects the frequency and severity of symptoms like wheezing, coughing, and difficulty breathing.
2. ***Forecasting***: The process of predicting future outcomes or trends based on historical and current data. In this context, it refers to predicting the future severity of asthma symptoms for an individual.
3. ***ML Approach***: A method of data analysis that automates model building, enabling systems to learn from data and make predictions. In this context, it refers to using algorithms like Random Forest, Gradient Boosting, or others to predict asthma severity.
4. ***Random Forest*:** A machine learning algorithm that builds an ensemble of decision trees, where each tree is trained using randomly selected subsets of data and features. In regression tasks, the model outputs the average of all tree predictions, while in classification, the final result is determined through majority voting among the trees.

5. ***Hyperparameter Tuning***: The process of adjusting the parameters of a machine learning model to optimize its performance. Hyperparameters are settings that are not learned from the data but are set before the learning process begins.
6. ***Performance Metrics***: Quantitative measures used to assess the effectiveness of a machine learning model. Common metrics include accuracy, precision, recall, F1-score, and AUC.

Asthma and Its Management

Asthma, a prevalent chronic respiratory condition, is characterized by respiratory symptoms such as breathlessness (dyspnea), wheezing, cough, and chest tightness. It also involves increased variability in airway caliber, which fluctuates over time (Louis *et al.,* 2023). Globally, asthma affects about 300 million people (Klemettila *et al.,* 2017). While there has been a period of sustained high prevalence, recent trends indicate stabilization in its global occurrence. However, significant regional differences persist, with ongoing increases observed in many low- and middle-income countries (Mortimer *et al.*, 2022) Recent studies report that asthma symptoms affect approximately 10% of children and adolescents worldwide (Aaron *et al.,* 2018), while in adults, the prevalence ranges from 6% to 7%. These rates vary significantly, with some low-income countries reporting rates as low as 2–3%, contrasting sharply with rates of 10% in high-income countries (García-Marcos *et al.,* 2022).

Asthma ranks among the most common chronic illnesses in children, and research has shown a clear link between the severity of childhood asthma and the likelihood of developing Chronic Obstructive Pulmonary Disease (COPD) in adulthood. In contrast to COPD, the narrowing of airways in asthma can typically be reversed with proper treatment. However, when asthma is not managed promptly and effectively, it can progress into COPD, a condition that tends to deteriorate over time and may result in lasting lung damage. Research indicates that approximately 40% of individuals with COPD also have asthma, and this overlap becomes more common as patients grow older (Ali, 2022).

A successful control of asthma depends mostly on patients' behavior (Şanlıtürk & Ayaz-Alkaya, 2021). These self-management behaviors include self-monitoring of symptoms, the effective use of an asthma action plan (individualized instructions on steps to take to prevent asthma from getting worse), the avoidance of environmental triggers, and the regular adherence with pharmacotherapy such as using proper medication with an appropriate inhalation device (Dhippayom *et al.,* 2022). Promoting optimal patient adherence to asthma self-management guidelines has proven to be the key to success in delaying the progression of this disease (Saleem *et al.,* 2022).

There are three levels of asthma progression, which parallels the severity levels of the disease as presented in Figure 1.



Figure 1: Asthma Progression

At level 1, patients need to manage their own illness based on their best knowledge of their conditions. This level is the lowest priority for healthcare providers. About 70-80% of asthma patients are in this group. It is important to delay or prevent the fast progression of the disease to a more severe level, as permanent damage to the lungs is irreversible. At level 2, as they are entering a more advanced stage of the disease, these patients will benefit from monitoring symptoms and assessing the severity level of their disease to make sure that the condition is under control and that they are taking proper medications. At level 3, the patients are entering into the most advanced stage of their asthmatic condition. At this level, patients need close monitoring to minimize unnecessary hospital admission. Successful asthma self-management can slow down the progression of the disease and improve the quality of life of the patient (Agustiyaningsih *et al.,* 2023).

Factors Contributing to Asthma

This section addresses the first objective of the study: to determine the primary environmental, psychological, physiological, and behavioural factors contributing to asthma severity. The study employs a quantitative research design guided by ML to uncover significant patterns and relationships between these variables and asthma severity. The goal is to extract data-driven insights that not only inform predictive modeling but also enhance clinical understanding of asthma triggers and progression. The following subsections explain the key factors considered under each category and their relevance to asthma severity.

Environmental Factors

Environmental factors are among the most well-documented contributors to asthma development and exacerbation. This study considers multiple environmental variables that have been empirically linked to increased asthma severity:

Air Quality Index (AQI): This is a composite measure of air pollution levels, typically influenced by particulate matter (PM2.5 and PM10), ozone, carbon monoxide, nitrogen dioxide, and sulfur dioxide. Exposure to poor air quality has been consistently associated with respiratory inflammation and increased hospital admissions due to asthma.

Particulate Matter (PM2.5 and PM10): These airborne particles can penetrate deep into the lungs, triggering bronchial inflammation and reduced lung function. High concentrations of PM2.5 are particularly dangerous for individuals with pre-existing respiratory conditions.

Temperature and Humidity: Rapid changes in temperature and extreme humidity levels (both high and low) can trigger asthma symptoms. Cold air, in particular, has a constrictive effect on the airways, while high humidity may facilitate mold growth and allergen exposure.

Seasonal Variations: Asthma severity often worsens during certain seasons due to increased exposure to pollen, viral infections, and changes in indoor/outdoor activity levels. Seasonal patterns can serve as useful temporal indicators for predicting exacerbation risks.

Indoor Pollutants: Although less accessible in public datasets, indoor air quality measures such as dust, pet dander, mold spores, and household chemical exposure are also noted as influential. Where available, proxy indicators like home location or housing type may be used.

Psychological Factors

Psychological well-being plays a critical role in the management and progression of chronic diseases, including asthma. Psychological stress can trigger neuroendocrine responses that lead to immune dysregulation and airway inflammation. The following psychological variables are considered:

Perceived Stress Levels*:* This reflects the individual's self-reported level of stress, typically measured using validated tools like the Perceived Stress Scale (PSS). Chronic stress has been linked with increased frequency of asthma symptoms and reduced response to medications.

Anxiety and Depression: The presence of anxiety disorders or depressive symptoms can influence both the perception and reporting of asthma symptoms. Moreover, psychological distress may lead to poor self-management, non-adherence to medication, and increased emergency room visits.

Sleep Quality: Disturbed or insufficient sleep is both a symptom and a trigger of asthma. Poor sleep quality is associated with heightened bronchial responsiveness, increased inflammatory markers, and decreased lung function.

Physiological Factors

Physiological characteristics provide a direct measure of the patient's respiratory function and overall health status. The study includes the following key physiological parameters:

Forced Expiratory Volume in 1 Second (FEV1*):* FEV1 is a core spirometric measure representing the amount of air a person can forcibly exhale in one second. Reduced FEV1 is a strong indicator of airway obstruction and is used to classify asthma severity.

Peak Expiratory Flow (PEF): PEF reflects the maximum speed of expiration and is a common parameter used in both home-based and clinical asthma monitoring. Significant daily variability in PEF is often a marker of poor asthma control.

Oxygen Saturation (SpO_2): Low oxygen saturation levels during rest or activity indicate impaired gas exchange and correlate with more severe asthma cases.

Body Mass Index (BMI)*:* Obesity is recognized as a risk factor for both the onset and severity of asthma. Elevated BMI can lead to systemic inflammation, reduced lung volume, and increased comorbidity burden.

Heart Rate and Respiratory Rate: Tachycardia and elevated respiratory rate may be indicative of ongoing or imminent exacerbation, especially in acute care settings.

Behavioural Factors

Behavioural patterns are crucial determinants of asthma control. They influence the frequency of symptom triggers, adherence to prescribed treatments, and overall disease trajectory. The following behavioural factors are examined:

Medication Adherence: Non-adherence to controller medications (e.g., inhaled corticosteroids) significantly increases the risk of exacerbations and poor symptom control. Adherence may be inferred from pharmacy refill data, self-reports, or digital inhaler usage logs.

Smoking Status*:* Both active smoking and exposure to second-hand smoke are major risk factors for increased asthma severity. Smoking damages lung tissue and reduces responsiveness to anti-inflammatory treatments.

Physical Activity*:* While moderate exercise is beneficial, excessive or poorly timed physical activity can induce bronchospasm in some individuals. Activity level may be quantified through wearable device data or self-report.

Rescue Inhaler Usage: Frequent use of short-acting beta-agonists (SABA) may indicate poor baseline control. Over-reliance on rescue medication is a key behavioral marker of uncontrolled asthma.

Diet and Nutritional Habits*:* While not always directly available in clinical datasets, diet patterns can influence systemic inflammation and immune response. Where possible, dietary data or BMI trends may be used as a proxy.

The study identifies and incorporates a comprehensive set of environmental, psychological, physiological, and behavioural factors to determine their contribution to asthma severity. These variables are selected based on empirical evidence, clinical relevance, and availability in the data sources. The integration of these multidimensional features allows for a robust, data-informed approach to uncover the key drivers of asthma progression. This thorough exploration fulfills the first objective of the study and provides a foundation for building a predictive model in subsequent methodological stages.

Asthma Categories

Asthma is commonly classified into two major types: extrinsic and intrinsic. Extrinsic asthma, also known as allergic asthma, typically begins in childhood and represents about 90% of all asthma cases. In contrast, intrinsic asthma usually emerges after the age of 30 and accounts for approximately 10% of cases (Federico *et al*., 2021). An asthma episode is characterized by breathing difficulties caused by a combination of factors such as inflammation of the airway lining, muscle tightening around the airways, and excess mucus production, all of which restrict airflow (Fuentes *et al*., 2021).

The typical symptoms include wheezing, persistent coughing, and shortness of breath. These symptoms may be triggered by various inhaled substances such as dust mites, pollen, pet dander, mold, or environmental irritants like cold air, tobacco smoke, pollutants, respiratory infections, physical exertion, stress, certain foods, or side effects from medications (Ramphul, 2023). The specific triggers and severity of symptoms can differ from person to person due to variations in lung capacity and function, which are influenced by environmental and physiological factors such as age, fitness level, and body size.

To assess lung function, the Peak Expiratory Flow (PEF) — a measure of how fast air is expelled from the lungs — is commonly used. When symptoms become intense and uncontrollable, the episode is referred to as an asthma attack or exacerbation (Ji *et al*., 2021), which may require prompt medical intervention, medication, or hospitalization for effective management.

Asthma Treatment

This section presents key approaches to managing asthma effectively. It begins by discussing strategies for day-to-day monitoring and the treatment of exacerbations, followed by methods for assessing the severity of symptoms. The discussion then shifts to the importance of asthma self-management practices and concludes with techniques for avoiding environmental triggers that can worsen the condition.

Day-to-day Monitoring and Treatment of Exacerbations

The 2007 Guidelines for the Diagnosis and Management of Asthma denote that asthma is caused by the "interaction of multiple host factors such as innate immunity, genetics, and environmental exposures". Currently, recommended asthma treatments focus mostly on considering genetic factors and environmental triggers. The report also recommends four components of asthma management, which are:

i. Assessing and monitoring severity,
ii. Controlling symptoms through elimination of triggers, in combination with treatment of acute / chronic commodities that impact asthma,
iii. Including individuals with asthma as partners in care, and
iv. Appropriate use of medications (Martin *et al*., 2022).

Since high level of lung performance with little variability is considered as the characteristic of good asthma control, PEF is measured twice daily to identify the lungs' performance. Symptoms assessment is another method to determine asthma severity level. Also, medications need to be used as a feedback mechanism to maintain steady-state control of symptoms (Licari *et al.,* 2018). An efficient management of asthma requires appropriate medications.

Inhaler medications are used for two main purposes:

i. Reliever Inhalers (quick acting inhaled medication): reducing the symptoms' effect.
ii. Preventer Inhalers (long-acting Bronchodilators): preventing exacerbations.

While Preventer Inhalers are used twice daily to reduce the airways' inflammation, Reliever Inhalers are used in emergency cases for the fast relief of symptoms. These medications are very effective in asthma's treatment but at the same time, they are at the high side of causing dose related side effects (Kachroo *et al.,* 2022; Mosbech *et al.,* 2023) . Especially, it is not good to overuse Reliever Inhalers, as these medications have been proved to speed up the disease progression. For that reason, the most effective treatment plan should adopt a flexible adjustment of the number of medications used according to the severity level of symptoms at a certain time (Azzi *et al.,* 2019).

'Step-up', 'step-down' strategy is applied so that patients can take large doses of inhaled steroids in the case of a patient's episodes of poor control, and then reduce the dose when good control

has been gained (Cazzola *et al.,* 2021). When an asthma attack occurs, inhaled steroid doses should be doubled or stronger medications may be prescribed to quickly relieve symptoms under the supervision of healthcare providers. High compliance with medication is crucial to well-controlled asthma (Beasley *et al.,* 2020). Patients are advised to track their symptoms, medications, and function test results to obtain a suitable asthma action plan that could help prevent asthma symptoms from getting worse. Research indicates that such asthma diaries are invaluable in identifying asthma attacks and as an education tool to support patients learning about their personal triggers (Perry *et al.,* 2017). A well-designed asthma diary must have characteristics such as a chart which shows the variations in peak flow rates (if recorded), tracking asthma management information such as asthma triggers, signs and symptoms, and medicines, recording feedback or comments, clearly identifying the relationships among triggers, medications, and other features of the condition (Sheares *et al.,* 2015).

Monitoring and Assessing Severity Level of Symptoms

As previously noted, managing asthma effectively demands continuous monitoring to adapt to changes in disease progression. Although there is no universally accepted standard for diagnosing or tracking asthma (Hodkinson *et al*., 2020), healthcare professionals typically rely on two primary approaches: subjective evaluation of symptoms and objective measurement of lung function, such as Peak Expiratory Flow (PEF). However, many individuals with asthma are often unaware of the gradual increase in airway obstruction until they experience a full-blown asthma attack. In fact, a significant number of patients only recognize the severity of their condition when acute symptoms suddenly manifest (Rosenfeld & Last, 2022).

This delayed awareness underscores the limitations of symptom-based management for many patients. Regular monitoring of PEF, especially when done twice daily — in the morning and evening — over a two-week period, helps determine a person's personal best PEF. This method has proven effective for asthma control because it provides more objective insights compared to relying on symptoms alone. Moreover, notable fluctuations in daily PEF readings (diurnal variability) often signal an impending worsening of asthma (Li *et al*., 2022).

While combining both symptom assessment and lung function data can improve diagnostic accuracy, the nonspecific nature of asthma symptoms and their subjective interpretation make it difficult to establish a clear link between reported symptoms and actual PEF values (Hao *et al*., 2022). However, advances in **t**ele-monitoring technologies offer a promising solution by enabling real-time tracking of lung function and alerting patients when their condition deteriorates (Im *et al*., 2023). Ultimately, the primary goal of asthma care is to prevent severe episodes — often marked by a PEF falling below 50% of an individual's baseline. Because treatment responses vary greatly from person to person, consistently evaluating patient adherence to prescribed therapies is essential to identify those who may be at higher risk of poor asthma control.

Asthma Self-Management

The aging population is rapidly increasing and is estimated to reach 1 trillion people by 2030, which is about 13% of the whole population (Qi *et al.,* 2023). Studies have proven that the

increasing of age-related chronic diseases' prevalence will lead to a tremendous increase in people with disabilities all over the world. Specifically, COPD will be the fifth leading cause of disability after heart disease, unipolar major depression, road-traffic accidents, and cerebrovascular disease (Tyrovolas *et al.,* 2022). By 2030, the proportion of the burden attributed to chronic diseases (incurable, non-infectious, or non-transmissible conditions such as heart disease, cancer, asthma, and diabetes) in low- and middle-income countries will reach 54% while it is 89% in high-income countries (Meghji *et al.,* 2021). The rise in the cost of treatment and clinician management time is significant. About 80% of healthcare providers are in charge of chronic conditions, and patients with such conditions or their complications account for over 60% of hospital days (Shiri *et al.*, 2021). For that reason, the personalized healthcare system (p-Health), which focuses on the early detection of risk factors, early diagnosis, disease management, treatment, and home rehabilitation is presently used aggressively to reduce the cost and increase management quality of these chronic diseases (Mudiyanselage *et al.,* 2023). Wireless, wearable medical systems that sense patients' health conditions, physiological parameters, and biochemical variables have been employed to continuously monitor patients' condition and provide advice or recommendations that could help prevent or slow down the progression of the disease (Deng *et al.,* 2023).

Although the usefulness of monitoring health conditions and physiological parameters in chronic disease self-management has been proven, patient engagement which includes three characteristics,

i. Recognizing and understanding the importance of taking an active role in one's health and health care.
ii. Having the knowledge, skills, and confidence to manage health; and
iii. Using knowledge, skills and confidence to engage in health-promoting behaviors to obtain the greatest benefit, is also considered as one of the factors that is critical for improving health outcomes (Hosseinzadeh *et al.,* 2022).

A research study by (Farzandipour et al., 2024), in which 148 older adult asthma patients were randomly assigned into one of three groups with a 6-month follow up:

i. Usual care
ii. Individualized education
iii. Individualized education with peak flow monitoring indicated that asthma self-care competence scores.

Asthma self-care behavior and self-efficacy scores ($P < 0.001$), and asthma control indicators ($P = 0.025$) are statistically significantly highest among group (3) (Walsh *et al.,* 2022). The research concluded that individualized education helps older asthma patients improve self-care behaviors, managing of asthma, and quality of life (Hashim *et al.,* 2022). Asthma self-management education which includes the provision of information, self-monitoring, regular medical review, and the provision of a PEF-based asthma management plan is considered to be a fundamental component of asthma care. Patient engagement is a complicated process as

described by its above definition, because many asthma patients face "difficulties in following action plans on a daily basis and have low compliance with medication regimens" (Saleem *et al.,* 2022).

Avoiding Environmental Triggers

Environmental conditions such as **weather changes, air pollution, and pollen exposure** are known to significantly influence asthma symptoms. Additional factors including thunderstorms, high temperatures, humidity, and wind speed can also contribute to symptom exacerbation. However, the body's response to these environmental triggers is **highly individualized and may not be immediate**, as some effects can take anywhere from a few hours to several days (up to 4 days) to manifest fully.

A study by Maung *et al*. (2022) explored the impact of eight common environmental variables—**temperature, ozone, benzene, particulate matter (PM10), humidity, pollen, atmospheric pressure, and nitric oxide** on lung function in individuals with mild to moderate asthma. Although the research confirmed that these environmental elements do affect respiratory health, attempts to create a universal prediction model for this patient group were unsuccessful. This is primarily because environmental triggers influence each person differently, making generalized models less effective. As noted by Alharbi *et al*. (2021), **personalized models** that use individual-specific data are more promising for predicting **Peak Expiratory Flow (PEF)** on a daily basis.

Still, challenges remain. The accuracy of PEF measurements can vary by as much as (**10%)**, and collecting precise environmental data at each patient's exact location poses practical difficulties. These limitations reduce the reliability of existing models in capturing the effects of weather and pollution on asthma symptoms. According to Soremekun *et al*. (2023), although predicting asthma attacks solely from PEF or symptom tracking is not currently feasible, **recognizable patterns do exist**. The authors argue that a well-designed study could detect these patterns more effectively. They recommend conducting large-scale studies that track PEF among severe asthma patients and search for **warning signs**, such as increasing variability in PEF readings in the days leading up to hospitalization or emergency visits. However, they also acknowledge that such studies would require significant resources due to the large sample size needed

Artificial Intelligent in Healthcare

The increasing availability of electronic health records (EHRs), along with advancements in Artificial Intelligence (AI) particularly Machine Learning (ML) has opened up new opportunities to enhance healthcare delivery and cut costs (Supriya & Chattu, 2021). However, utilizing this data effectively is not without its challenges. One major obstacle lies in the heterogeneity of healthcare data it is vast, generated rapidly, and varies greatly in type, format, and content across different medical settings. This complexity makes it difficult to integrate and analyze the data efficiently. Moreover, obtaining sufficient volumes of high-quality EHR data for training predictive models is often a significant barrier.

To overcome these limitations, AI researchers are developing advanced, scalable algorithms capable of detecting patterns and deriving meaningful insights from massive, complex datasets. One of the earliest and most impactful implementations of AI in the medical field was the Clinical Decision Support System (CDSS), designed to assist in diagnosing diseases based on patients' symptoms and demographic profiles (Thakur *et al*., 2022).

The first-generation CDSS tools were typically narrow in scope, targeting specific patient groups or medical conditions. More recent systems have evolved to handle more complex tasks, such as answering clinical questions and identifying unusual patterns across large datasets. A notable example is IBM Watson, which can interpret vast volumes of unstructured health data including doctors' notes, scientific publications, and patient records and use that information to generate diagnostic suggestions and personalized treatment recommendations. Watson employs probabilistic reasoning to support evidence-based medical decisions (Comito *et al*., 2022).

In another AI-driven approach, subset scanning techniques are applied to detect anomalies in healthcare data. These ML methods can identify both rows (patient records) and columns (features) that deviate significantly from expected patterns and are associated with improved or worsened patient outcomes. This technique helps uncover patterns in clinical practices that may lead to better results.

A historical illustration of data-driven pattern recognition can be seen in Ignaz Semmelweis's discovery of the link between puerperal fever and inadequate hand hygiene among obstetricians. Through systematic elimination of other causes, he identified cadaveric contamination as the source and introduced handwashing with chlorinated lime, drastically reducing mortality rates. This case highlights the vital role of hypothesis generation through data pattern analysis.

Today, ML and Deep Neural Networks (DNNs) are increasingly used in the management of chronic illnesses like asthma. These technologies enable real-time monitoring of patients' health and treatment adherence. Although early attempts to use AI in predicting asthma exacerbations faced setbacks, recent advancements in ML and DNN techniques have demonstrated promising results in accurately forecasting such events (Nassif *et al*., 2021).

Machine Learning in Healthcare

ML is an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed (Shaveta, 2023). There are two steps in a ML model:

i. The invention of algorithms which can recognize relationships within historical data.
ii. The use of these identified patterns to perform predictions using new data.

Healthcare is one of the prominent industries presently being transformed by ML to increase the capacity to treat and diagnose disease. Computer vision (a sub-field of ML) measures space accurately and recognizes human organs expertly, which makes robotic surgery possible. Robots support surgical procedures where there is a tighter space and where finer details and

accuracy are required. In medical imaging ML provides early detection of disease such as diabetic retinopathy and macular edema in pictures of retinal fundus, Computer Aided Diagnosis, and functional brain mapping. ML now also informs personalized medicine with the prediction of medical success rates and understanding disease mechanisms, as well as the prediction of a disease outbreak or supporting personalized treatment of chronic diseases. However, the adoption of ML techniques in the medical field is slower than other specialties such as business, economic, insurance, and biology (Yip *et al.,* 2023). An overview of the ML-driven healthcare data pipeline is illustrated in Figure 2.

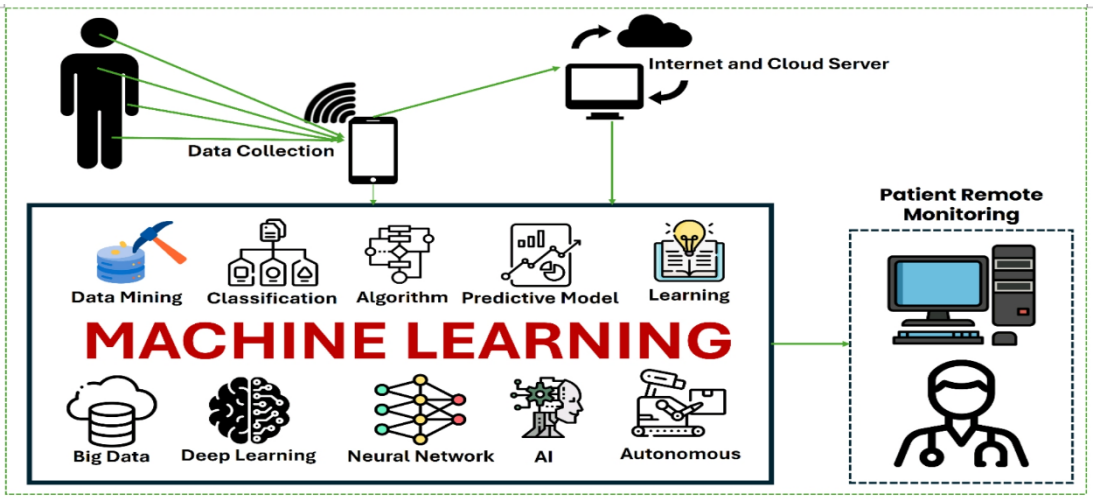


Figure 2: Machine Learning in Healthcare

Types of Machine Learning

Machine Learning techniques can be broadly categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning. Each of these categories serves different purposes and is suited to different kinds of problems.

Supervised Learning

Supervised learning involves training models on labeled data, where the input data comes with corresponding output labels. The model learns to make predictions or classify data based on this input-output mapping. For instance, in asthma severity forecasting, supervised learning can be used to predict the risk of an asthma attack based on historical health data, environmental factors, and other relevant parameters. Techniques such as logistic regression, decision trees, and deep learning (e.g., convolutional neural networks and long short-term memory networks) fall under this category. These techniques are beneficial when there is a substantial amount of labeled data available, as they can provide high accuracy in predicting outcomes based on learned patterns (Zein *et al.,* 2021).

With supervised learning, the data is labeled. A learning algorithm is trained with a dataset, so that the algorithm can learn to adapt to a new situation. In the field of medicine, supervised learning has been used to perform classification tasks or to estimate risk. With unsupervised

learning, the data is not labeled. The learning algorithms are used to extract patterns or relationships from data. Possibly, unsupervised learning gains most attention with "precision medicine" (El-Baz & Suri, 2021). These supervised learning techniques are discussed in the following sections.

i. **Logistic Regression:** Often used for binary classification tasks, logistic regression can predict the probability of asthma exacerbations based on various features such as patient history, environmental factors, and medication adherence (Kumar & Gota, 2023).
ii. **Decision Trees and Random Forests:** These are powerful for understanding the decision rules that lead to asthma attacks. For instance, they can help identify key factors that increase the risk of severe asthma episodes, such as high particulate matter (PM) levels or specific weather conditions (Nanthakumaran & Liyanage, 2021).
iii. **Support Vector Machines (SVM):** SVMs can classify asthma severity levels by finding the hyperplane that best separates different classes of severity in a multidimensional space (Wen & Mohd Aris, 2022).
iv. **Neural Networks and Deep Learning*:*** Techniques like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks can handle large, complex datasets. CNNs can analyze patterns in time-series data (like PEFR measurements over time), while LSTMs are particularly good at handling sequential data, making them suitable for predicting asthma attacks based on historical patient data and continuous monitoring (Alzubaidi *et al.,* 2021).

Unsupervised Learning

Unsupervised learning, on the other hand, deals with data that does not have labeled responses. The primary goal here is to identify hidden patterns or intrinsic structures in the input data. Methods like clustering (e.g., k-means clustering, hierarchical clustering) and dimensionality reduction (e.g., principal components analysis) are common in unsupervised learning. In the context of asthma, unsupervised learning can be utilized to identify subgroups of patients with similar symptoms or responses to treatment, which can lead to more personalized healthcare approaches (Sun *et al.,* 2022). This type of learning is particularly useful when exploring new datasets where the relationships and groupings are not yet known. The unsupervised learning techniques are discussed in the following sections.

i. **K-Means Clustering:** This can group patients into clusters based on similar asthma symptoms, treatment responses, or environmental exposures. For example, patients living in areas with high pollution might form a distinct cluster, indicating a shared risk factor for exacerbations (Reel *et al.,* 2022).
ii. **Hierarchical Clustering:** This technique builds a tree of clusters, which can help in understanding the hierarchy and relationships between different patient groups or asthma triggers (Gasiuniene *et al.,* 2023).
iii. **Principal Component Analysis (PCA)*:*** PCA can reduce the dimensionality of complex datasets, highlighting the most significant variables affecting asthma

severity. This helps in simplifying the data while retaining its critical information, making it easier to visualize and analyze (Camargo, 2022).

Reinforcement Learning

Reinforcement learning focuses on decision-making by learning optimal actions through trial and error. It is a type of dynamic programming where an agent learns to achieve a goal in an uncertain, potentially complex environment. The agent receives feedback in the form of rewards or punishments, which it uses to improve its performance over time. For asthma management, reinforcement learning can be applied in developing adaptive treatment plans that adjust based on the patient's response to previous treatments. For example, a reinforcement learning algorithm could suggest adjustments in medication dosages or lifestyle changes to minimize the risk of asthma exacerbations, continually learning and adapting based on the patient's health outcomes (Aliyu *et al.,* 2024; Alsaihati *et al.,* 2021). The reinforcement learning techniques are discussed in the following sections.

i. **Q-Learning:** A model-free RL algorithm that can help in developing personalized medication plans by learning the best actions (e.g., medication adjustments) to minimize asthma attacks (Yang *et al.*, 2023).
ii. **Deep Q-Networks (DQN):** Combining Q-Learning with deep learning, DQNs can handle high-dimensional state spaces, making them suitable for complex decision-making tasks in asthma management, such as optimizing long-term treatment plans based on continuous patient monitoring data (Skolik *et al.,* 2022).
iii. **Policy Gradient Methods:** These methods directly optimize the policy (a strategy that maps states to actions) to improve patient outcomes. For instance, they can adjust treatment protocols dynamically as new patient data is collected, ensuring that the management strategy is continuously refined (Viquerat *et al.,* 2023).

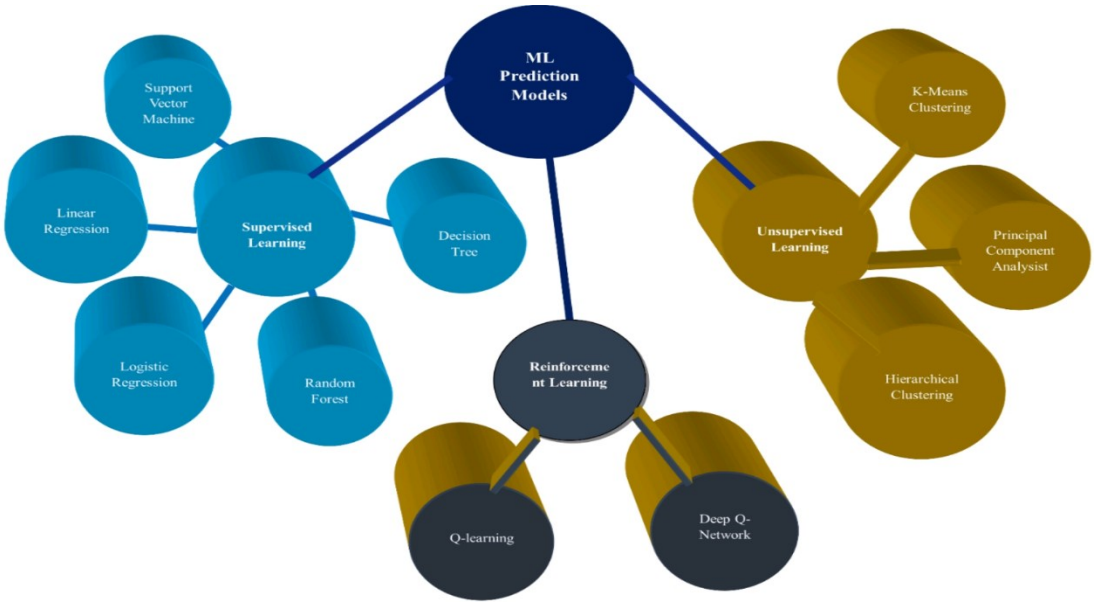


Figure 3: Various ML Models for Asthma Prediction (Aliyu *et al.,* 2024)

These three types of machine learning, as shown in Figure 3: supervised, unsupervised, and reinforcement learning, offer diverse approaches to handling and analyzing data. Supervised learning is effective for predictive modeling with labeled data, unsupervised learning excels in uncovering hidden patterns in unlabeled data, and reinforcement learning provides robust decision-making strategies through iterative feedback. Each type has its unique strengths and applications, making them valuable tools in the development of advanced, data-driven solutions for asthma severity forecasting and treatment optimization.

Applications in Healthcare

ML has transformative applications in healthcare, offering innovative solutions to a range of complex problems. These applications include image analysis, predictive modeling, Natural Language Processing (NLP), and clinical decision support. Here's a detailed look at these applications:

i. **Image Analysis:** ML algorithms, particularly deep learning techniques like Convolutional Neural Networks (CNNs), are extensively used in medical image analysis. These algorithms can analyze and interpret complex medical images, aiding in the early detection and diagnosis of various diseases. For instance, CNNs have been employed to identify tumors in mammograms, detect lung abnormalities in chest X-rays, and diagnose retinal diseases from optical coherence tomography (OCT) images. By providing accurate and timely analysis, ML helps in early disease detection, which is crucial for successful treatment outcomes (Yu *et al.,* 2023).

ii. **Predictive Modelling:** Predictive modelling in healthcare involves using ML algorithms to analyze historical patient data and predict future health outcomes. These models can forecast disease progression, predict the likelihood of hospital admissions, and assess the risk of complications. For example, predictive models can analyze Electronic Health Records (EHRs) to identify patients at high risk of developing conditions like diabetes or heart disease. In the context of asthma, ML models can predict the risk of exacerbations based on factors such as medication adherence, environmental exposures, and historical health data. This allows for proactive management and personalized care plans to mitigate risks (Huang & Wu, 2021).

iii. **Natural Language Processing (NLP):** NLP techniques enable the extraction and interpretation of information from unstructured clinical texts, such as physician notes, discharge summaries, and medical reports. By converting this unstructured data into structured formats, NLP facilitates better data analysis and integration. For instance, NLP algorithms can identify mentions of symptoms, medications, and diagnoses in clinical notes, providing a comprehensive view of a patient's health status. This capability enhances the accuracy of patient records and supports more informed clinical decision-making (Bose *et al.*, 2021).

iv. **Clinical Decision Support*:*** Clinical decision support systems (CDSS) leverage ML algorithms to provide healthcare professionals with evidence-based recommendations and

insights. These systems analyze vast amounts of clinical data to assist in diagnosing conditions, selecting appropriate treatments, and predicting patient responses to therapies. For example, ML-based CDSS can recommend personalized treatment plans for cancer patients by analyzing genetic profiles and treatment histories. In asthma management, CDSS can suggest optimal medication adjustments based on real-time monitoring data and historical trends, improving patient outcomes and reducing the likelihood of severe exacerbations (Antoniadi *et al.,* 2021).

v. **Drug Discovery and Development*:*** ML algorithms play a crucial role in accelerating the drug discovery and development process. By analyzing biological data, chemical properties, and clinical trial results, ML models can identify potential drug candidates, predict their efficacy, and optimize clinical trial designs. This reduces the time and cost associated with bringing new drugs to market. For example, ML has been used to discover new antibiotics and predict the success of cancer therapies based on genetic markers (Antoniadi *et al.*, 2021).

vi. **Personalized Medicine*:*** Personalized medicine involves tailoring medical treatment to individual patient characteristics, such as genetic makeup, lifestyle, and environment. ML algorithms analyze diverse datasets to identify patterns and correlations that inform personalized treatment strategies. For instance, in oncology, ML models can predict how patients with specific genetic mutations will respond to different therapies, allowing for more effective and targeted treatments. In asthma care, personalized medicine enabled by ML can identify the most effective interventions for each patient based on their unique risk factors and health history.

The application of machine learning in healthcare encompasses a broad spectrum of techniques and technologies that significantly enhance the efficiency, accuracy, and personalization of medical care. By leveraging the power of ML, healthcare providers can improve diagnostic accuracy, predict patient outcomes, optimize treatment plans, and ultimately deliver better patient care. The continuous advancement of ML technologies promises even greater innovations and improvements in the healthcare sector.

Machine Learning in Asthma Management

Recent studies have explored the use of ML to improve asthma management by predicting disease progression, identifying exacerbation triggers, and personalizing treatment plans. This section reviews key studies in this area, highlighting the methodologies used and their findings.

Predicting Asthma Exacerbations

Several studies have focused on predicting asthma exacerbations using ML models. For example, a study by (Zein *et al.,* 2021) used logistic regression and decision trees to predict exacerbations based on demographic, clinical, and environmental data. Their model achieved an accuracy of 85%, demonstrating the potential of ML in early identification of exacerbation risks.

Personalized Treatment Plans

Personalized treatment plans aim to tailor interventions based on individual patient characteristics. A study by Lovrić *et al*. (2021) employed a random forest algorithm to predict the optimal medication regimen for asthma patients. Their model considered factors such as age, gender, comorbidities, and medication history, resulting in improved symptom control and reduced exacerbations compared to standard guidelines.

Integrating Wearable Devices and Mobile Health

The integration of wearable devices and mobile health (mHealth) technologies with ML algorithms has gained attention in recent years. A study Alam *et al.* (2021) utilized data from wearable devices to monitor physical activity, environmental exposure, and physiological parameters. They developed an ML model to predict asthma control levels, providing real-time feedback to patients and healthcare providers.

Review of Related Works

The utilization of machine learning (ML) techniques in forecasting asthma severity and treatment outcomes has seen significant advancements (Bhat *et al.,* 2021). developed an asthma risk prediction tool implemented as a mobile-health application leveraging IoT resources. They utilized Convolutional Neural Networks (CNNs) to correlate indoor particulate matter (PM) and weather data with Peak Expiratory Flow Rates (PEFR), classifying asthma risk into three categories. This study demonstrated superior performance of the CNN model over other deep neural network techniques, showcasing the potential of integrating mobile technology and environmental data for effective asthma management. Similarly, Bhat *et al.* (2021) explored the feasibility of using a Long Short-Term Memory (LSTM) model to predict asthma risk in pediatric patients based on PEFR and indoor PM concentrations, finding that the LSTM model, which considers cumulative PM effects over time, outperformed traditional multinomial logistic regression.

In another study, Zhang *et al.,* (2021) focused on early detection of severe asthma exacerbations using daily monitoring data from the SAKURA study. By analyzing daily peak expiratory flow and symptom scores, their logistic regression model, enhanced by principal components analysis (PCA), achieved high sensitivity and specificity. This highlights the effectiveness of daily self-monitoring data in predicting asthma exacerbations, suggesting that timely interventions can improve patient outcomes. In contrast, Castner *et al.,* (2020) investigated the use of fitness tracker sleep data to predict asthma-specific night-time wakening and daily lung function (FEV1), demonstrating good predictive value and illustrating the potential of integrating wearable technology into predictive models for more personalized asthma management.

Tong *et al.,* (2021) evaluated the generalizability of an ML model for predicting asthma hospital encounters across different healthcare systems. By examining many candidate features, they achieved high accuracy, sensitivity, and specificity, demonstrating the adaptability of ML models to diverse patient populations and healthcare settings. This emphasizes the potential of ML in identifying high-risk patients and optimizing care

management resources. Tibble *et al.,* (2019) and Hussain *et al.,* (2020) leveraged Electronic Health Record (EHR) data to predict asthma attacks, with (Tibble *et al.,* 2019). focusing on creating a risk score using various ML classifiers on a large dataset from Scottish general practices, and Hussain *et al.,* (2020) systematically applying multiple ML techniques on a longitudinal primary care database. Both studies underscore the importance of comprehensive health data in developing accurate predictive models for asthma.

Finkelstein and Jeong, (2017) explored the use of telemonitoring data for predicting asthma exacerbations in adults, demonstrating that machine learning techniques such as naive Bayesian classifiers and SVM could predict exacerbations with high sensitivity and specificity. This research highlights the potential of telemonitoring in providing personalized decision support for chronic disease management. Nickovic *et al.,* (2023) addressed the challenge of predicting Thunderstorm Asthma (TA) by developing a physically based pollen model (DREAM-POLL) which accurately predicted TA peaks during historical outbreaks, showcasing the utility of environmental modeling in asthma prediction. This underscores the need for incorporating environmental factors in predictive models to improve forecasting accuracy for asthma severity.

Finally, Lisspers *et al.,* (2021) discussed the limitations of using historical clinical data alone for short-term prediction of asthma exacerbations. They highlighted the need for incorporating additional data sources such as environmental triggers and wearable data to improve model performance. This research points to the necessity of integrating diverse data sources for more accurate predictions. Akbar *et al.,* (2019) compared different ML algorithms for early-stage asthma prediction, demonstrating that Naive Bayes achieved the highest accuracy, suggesting that feature selection and algorithm choice are crucial for developing effective predictive models. Pooja *et al.,* (2023) introduced a Weighted Feature Averaging Technique (WFAT) for extracting relevant features for asthma prediction, with their ensemble ML model outperforming individual classifiers, achieving high classification accuracy with reduced feature sets. This approach underscores the importance of feature selection and ensemble methods in enhancing predictive performance.

Despite these significant advancements, a research gap remains in the integration and harmonization of multiple data sources, such as environmental data, wearable technology, and electronic health records, into a unified predictive framework. Future research should focus on developing comprehensive models that leverage the synergistic potential of these diverse data streams to improve the accuracy and reliability of asthma severity and treatment outcome predictions. The summary of the related studies are presented in Table 1.

Research Gaps

Asthma remains a complex and multifactorial chronic condition that continues to pose significant management challenges, particularly in pediatric populations. Despite advancements in therapeutic strategies and an improved understanding of disease mechanisms, predicting asthma severity and treatment outcomes remains elusive (Licari *et al.,* 2018; Lovrić *et al.,* 2021). The variability in disease manifestation, influenced by genetic, environmental, and immunological factors, makes it difficult to establish a one-size-fits-all

treatment approach (Do *et al.,* 2019). Consequently, children with asthma often experience suboptimal treatment outcomes, leading to frequent exacerbations, hospitalizations, and an increased economic burden on healthcare systems.

Current asthma management largely follows standardized guidelines that, while effective at the population level, do not adequately address individual differences Global Initiative for Asthma [GINA], (2022). Personalized treatment approaches remain underdeveloped, and existing predictive models rely heavily on conventional statistical methods that may not capture the complexity of asthma pathophysiology (Yoon *et al.,* 2022). In this regard, ML has emerged as a promising tool for improving asthma prognosis by leveraging large datasets to identify patterns that traditional models cannot discern. However, research in this domain is still in its early stages, with limited studies exploring the integration of ML-driven predictive analytics into clinical asthma management (Tsang *et al.,* 2020).

A significant gap exists in the application of ML to pediatric asthma prognosis. Most studies to date have focused on adult populations, overlooking the unique developmental and physiological characteristics of children that influence asthma severity and treatment response (Tiotiu, 2021). Moreover, many ML-based models rely on single-source data, such as electronic health records (EHRs) or pulmonary function tests, without incorporating a holistic view of the patient's condition. The integration of multi-source data—including genetic, environmental, and real-time monitoring data—remains an underexplored area that could significantly enhance the accuracy of predictive models (Lee *et al.,* 2023).

Additionally, while wearable sensors and mobile health (mHealth) applications have made real-time patient monitoring more feasible, their use in ML-driven asthma prediction has not been extensively validated (Sun *et al.,* 2022). These technologies have the potential to provide continuous physiological and environmental data, enabling the development of dynamic risk prediction models. However, studies assessing the reliability, accuracy, and long-term feasibility of these technologies in pediatric asthma management are lacking. Furthermore, challenges related to data standardization, interoperability, and patient adherence must be addressed before these technologies can be effectively implemented in clinical practice (Arya *et al.,* 2023).

Another critical gap is the ethical and regulatory considerations associated with the use of ML in asthma management. While ML models offer improved predictive power, their 'black box' nature raises concerns about interpretability and clinical decision-making (Aliyu *et al.,* 2024). Physicians and healthcare providers may be hesitant to adopt ML-driven recommendations without clear explanations of how predictions are derived. Transparent and explainable AI models are necessary to foster trust and ensure that ML applications align with evidence-based medicine. Additionally, issues related to data privacy, security, and potential algorithmic biases must be thoroughly examined to prevent disparities in healthcare access and outcomes (Arya *et al.,* 2023)

Moreover, existing studies on ML in asthma management often lack external validation, limiting their generalizability across diverse populations and healthcare settings. Most models are trained on datasets from specific geographic locations, leading to potential biases that affect their applicability in different demographic groups (Tiotiu, 2021). Addressing these

limitations requires collaborative, multi-center studies that incorporate diverse patient populations and real-world clinical settings to ensure the robustness and applicability of ML-driven asthma management strategies (Sun *et al.,* 2022).

Given these gaps, there is an urgent need for interdisciplinary research that integrates clinical expertise, data science, and biomedical engineering to develop comprehensive ML-driven prognostic tools for pediatric asthma. This study aims to bridge these gaps by developing robust ML models that incorporate multi-source data—including genetic predisposition, environmental exposures, and real-time physiological monitoring—to improve the accuracy of asthma severity predictions. By tailoring treatment strategies to individual patients, this research has the potential to revolutionize pediatric asthma management, enhance clinical decision-making, and ultimately improve patient outcomes while reducing the burden on healthcare systems.

The Research Approach

The research approach is experimental and draws on insights from a comprehensive literature review, empirical testing, and analysis of publicly available datasets. These components guide the design and evaluation of the proposed hybrid model.

Prediction Model

The trained and optimized model ready to predict outcomes (e.g., patient satisfaction levels). This process ensures that the dataset is properly pre-processed, the most relevant features are selected, and the model is optimized and validated to deliver reliable predictions as shown in Figure 4. The use of a random forest, logistic regression and SVM helps in capturing complex patterns and efficiently updating model parameters for better performance.

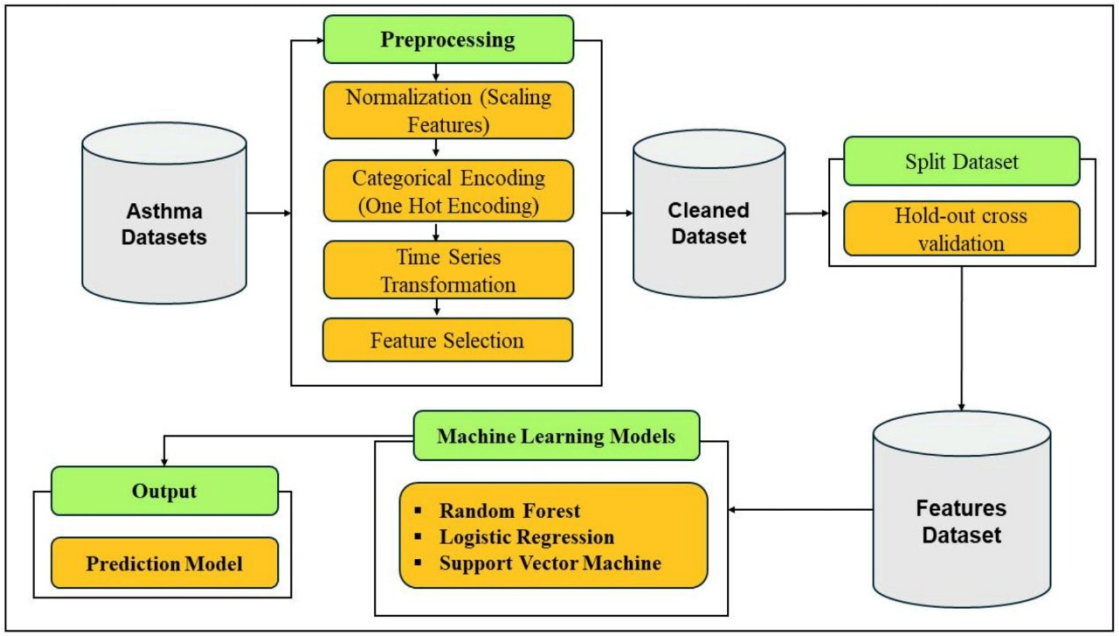


Figure 4: Model Development Model

Mode of Data Collection

This research used secondary mode of data collection. The source of resources includes Online Journals, articles, eBooks, electronic health records, public health databases, health surveys, environmental monitoring data and online databases.

Data Collection

The research utilized a dataset focusing on demographic information for individuals with asthma. This dataset encompasses diverse asthma symptoms (sore throat, pains, nasal congestion, etc.), patient age groups (e.g., Age_0-9, Age_10-19, etc.), gender, asthma severity levels (mild, moderate, and none), and concludes by indicating whether the patient has received an asthma diagnosis. This dataset has been sourced from Kaggle.

Data Preprocessing

The data preprocessing for the asthma demographic dataset begins by addressing missing values in key variable columns such as symptoms, test scores, and demographics, employing removal or imputation where applicable. Subsequently, categorical columns like gender and age group undergo encoding into numeric values using techniques such as label encoding or one-hot encoding. Following this, normalization and standardization are applied to numerical data to ensure uniform scaling for equitable modeling comparisons.

Given the class imbalance in the historical dataset, particularly the dominance of the 0-9 age group, techniques like down sampling will be employed to address this issue. Towards the conclusion, a thorough examination of features and targets is conducted, checking for statistical differences between the training and test sets, and adjusting as required. Finally, the preprocessed dataset is divided into training and validation/test sets, emphasizing the maintenance of high data quality and consistency across the splits. Meticulous attention throughout this preprocessing phase contributes to enhanced model performance and facilitates meaningful analytic conclusions regarding the correlations among symptoms, demographics, and the ultimate asthma diagnosis.

Description of the Dataset

The initial dataset consists of **1,010** entries and 14 columns, capturing a variety of demographic, environmental, and behavioral factors related to asthma severity. While the sample size is relatively modest for machine learning research, it provides a foundational basis for exploratory analysis and model prototyping. The key columns in the dataset are displayed in Table 2.

Table 2: Dataset Description

User№	**A numerical identifier for each user.**
UserID	Categorical identifier for each user.
Location	Categorical data representing the user's location.

User№	**A numerical identifier for each user.**
Age	Categorical data representing the age of the user.
Gender	Categorical data representing the gender of the user.
OutdoorJob	Categorical data indicating whether the user works outdoors.
OutdoorActivities	Categorical data indicating the user's involvement in outdoor activities.
SmokingHabit	Categorical data indicating the user's smoking habits.
Humidity, Pressure, Temperature, UVIndex, WindSpeed	Numerical environmental factors.
ACTScore	The Asthma Control Test score, representing asthma severity.

Data Sources and Types of Data Collected

The dataset was accessed from online source (GitHub) which includes the combination of clinical records (e.g., ACTScore, demographic data) and environmental monitoring (e.g., Humidity, Temperature). The data types include:

i. **Categorical**: UserID, Location, Age, Gender, OutdoorJob, OutdoorActivities, SmokingHabit, UVIndex.

ii. **Numerical**: UserNo, Humidity, Pressure, Temperature, WindSpeed, ACTScore.

Summary Statistics (e.g., mean, median, mode, standard deviation) for key variables.

The initial dataset consisted of 1,010 observations across 14 variables, including both categorical and numerical data types. The numerical features such as Humidity, Pressure, Temperature, WindSpeed, and ACTScore were analyzed to determine central tendencies and variability. Summary statistics for these variables are presented in Table 3. The data showed a reasonable distribution across most features, with mean values falling within expected ranges based on existing literature on asthma and environmental factors. For instance, the mean ACTScore, which serves as a proxy for asthma severity, was 16.58, with a standard deviation of 4.67, indicating a moderate level of variability in asthma control among the patients. This suggests that while many patients had similar levels of asthma severity, there was also a significant portion with either better or worse control, reflecting the diverse range of asthma experiences in the population studied.

Table 3: Summary statistics of the dataset

	User №	Humidity	Pressure	Temperature	WindSpeed	ACTScore
Count	1010.00	1010.00	1010.00	1010.00	1010.00	1010.00
Mean	3.93	81.75	1009.26	27.61	2.37	16.58
Std	2.542293	13.542801	2.114750	2.915736	1.569837	4.67

	User №	Humidity	Pressure	Temperature	WindSpeed	ACTScore
Min	1.000000	40.000000	1003.000000	21.100000	0.00	8.00
25%	3.00	70.00	1008.00	25.20	1.00	12.25
50%	3.00	87.00	1009.00	27.50	2.100	17.00
75%	6.00	93.00	1011.00	30.10	3.30	21.00
Max	10.00	100.00	1014.00	34.50	6.70	25.00

Data Cleaning Steps

Before proceeding with model development, the dataset underwent rigorous cleaning to address missing values, outliers, and inconsistencies in the categorical data. Missing values were identified, with the majority appearing in the environmental variables. These were handled using median imputation to preserve the central tendencies of the data without introducing bias. Outliers, particularly in the Temperature and ACTScore variables, were detected using the Interquartile Range (IQR) method. These outliers were reviewed contextually; extreme but clinically plausible values were retained, while erroneous entries were adjusted or removed. Categorical data inconsistencies, such as variations in the spelling or capitalization of categorical variables (e.g., Gender), were standardized to ensure uniformity.

Feature Selection and Engineering

Feature selection was guided by the relevance of each variable to asthma severity prediction, as determined by both domain knowledge and preliminary correlation analyses. Environmental variables such as Temperature, Humidity, and WindSpeed were prioritized due to their known impact on respiratory conditions. Additionally, feature engineering was performed to enhance model performance. For example, a composite environmental index was created by combining Temperature, Humidity, and WindSpeed into a single feature, capturing their joint influence on asthma outcomes. Categorical variables were encoded using Label Encoding to convert them into numerical format, facilitating their integration into machine learning models. Irrelevant or redundant features, such as UserID, were excluded from the final model input to reduce noise.

Splitting the Data into Training and Testing sets

To ensure the robustness of the model evaluation, the dataset was split into training and testing sets, with 80% of the data allocated to training and 20% to testing. This split was performed using a random stratification method to maintain the proportional distribution of the target variable (ACTScore) across both sets. The training set was used for model development and hyperparameter tuning, while the testing set provided an unbiased evaluation of the model's predictive performance. This approach allowed for assessing how well the model generalizes to unseen data, a critical factor in its applicability to real-world scenarios.

Logistic Regression for Asthma Severity Prediction

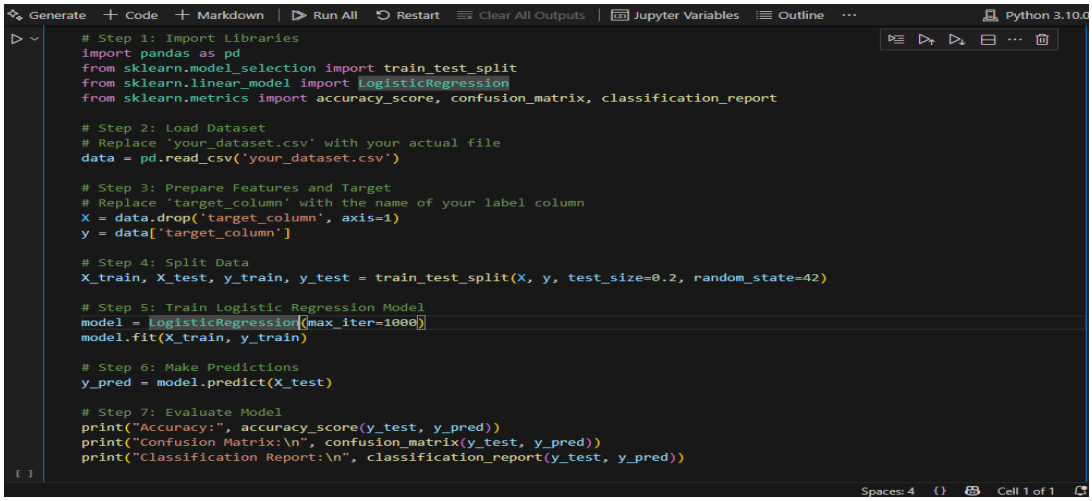
Logistic Regression is a statistical method used for binary and multiclass classification problems. It models the probability of a categorical outcome based on input features. In the context of asthma, Logistic Regression can be used to predict the severity level of an asthma condition (e.g., mild vs. severe) based on clinical and environmental features such as lung function metrics, medication history, air quality, or weather data.

The algorithm calculates a weighted sum of the input features, applies the sigmoid function, and outputs a probability between 0 and 1. If the predicted probability exceeds a threshold (commonly 0.5), the instance is classified into one of the severity levels.

Logistic Regression is favored for its simplicity, interpretability, and efficiency when the dataset is linearly separable. It also serves as a strong baseline model when comparing the performance of more complex algorithms like Random Forests or Support Vector Machines.

Below is a simple implementation of Logistic Regression using Python and the scikit-learn library:

Logistic Regression Algorithm in Python Appendix



Random Forest for Asthma Severity Prediction

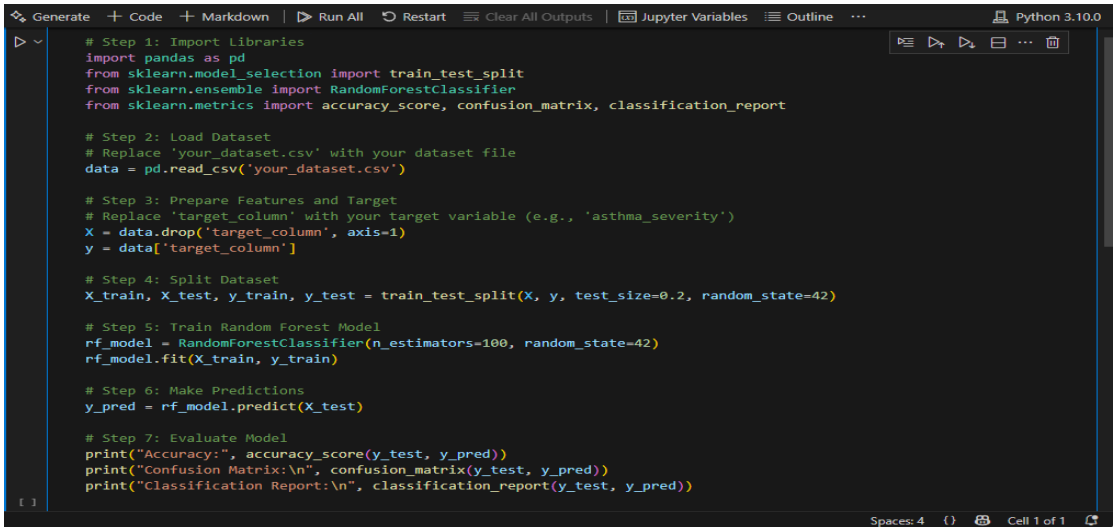
Random Forest is an ensemble learning algorithm that combines multiple **decision trees** to make robust predictions. It works by constructing a "forest" of individual decision trees during training and outputting the **mode of the classes (classification) or mean prediction (regression)** of the individual trees.

In asthma prediction, Random Forest is highly effective because it can:

i. b **non-linear relationships** and **interactions** between features,
ii. Manage both **numerical and categorical data,**
iii. Reduce overfitting through aggregation (bagging),
iv. Provide **feature importance scores**, which help identify the most influential variables (e.g., air pollutants, symptoms, peak flow values).

This makes it particularly useful for complex, multi-dimensional medical data where accuracy and robustness are critical.

Random Forest Algorithm in Python

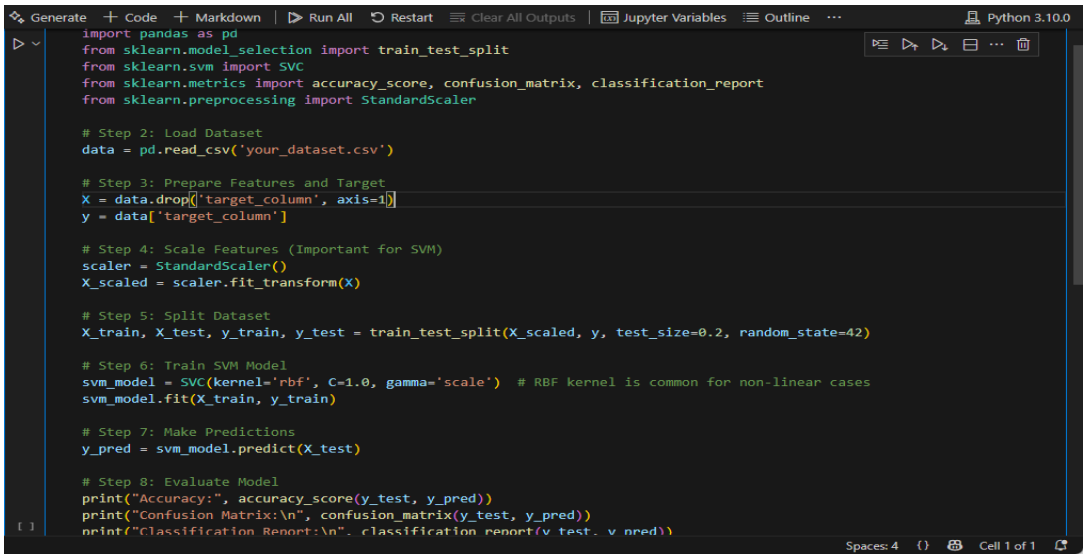


Support Vector Machine (SVM) for Asthma Severity Prediction

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for **classification** and **regression**. It works by finding the **optimal hyperplane** that separates data into ;]''different classes with the **maximum margin**. SVM is especially useful when dealing with high-dimensional datasets and non-linear boundaries through the use of **kernel functions**.

For asthma severity prediction, SVM is valuable when the relationship between features (like symptom frequency, medication usage, or environmental triggers) and severity is complex and not linearly separable. With proper preprocessing, SVM can achieve high accuracy and generalization performance.

SVM Algorithm in Python



Pre-processing Modelling

Pre-processing is crucial to ensure data quality and improve model performance

Missing Values: Handling missing data using methods such as:

a. **Imputation:** Replacing missing values with mean, median, or mode.
b. **Interpolation:** Estimating missing values based on other observations

Redundant Data: removing duplicate or highly correlated features to prevent multicollinearity which can distort model performance.

Categorical Data: converting categorical variables into numerical values using:

a. **One-hot encoding:** Creating binary columns for each category.
b. **Label encoding:** Assigning a unique integer to each category

Feature Scaling: Standardizing or normalizing features to ensure they contribute equally to the model. Common methods include:

a. **Standardization:** Transforming data to have zero mean and unit variance using equation 1.

$$z = \frac{(x - \mu)}{\sigma} \qquad (1)$$

where x is the feature value, μ is the mean, and σ is the standard deviation

b. **Normalization:** Scaling features to a fixed range, usually [0, 1].

Feature Selection: Identifying the most relevant features using techniques like:

a. **Correlation Matrix:** Selecting features with high correlation to the target variable.

b. **Principal Component Analysis (PCA):** Reducing dimensionality by transforming features into a new set of uncorrelated variables (principal components).

Data Splitting

1. **Split Dataset:** Dividing the data into training and testing sets to evaluate model performance.
    a. ***Hold-Out Cross Validation*:** Typically involves splitting the dataset into a training set (e.g., 80%) and a test set (e.g., 20%).

Applying Optimized Approach

1. **Features Dataset:** The dataset after pre-processing and feature selection, ready for model training.
2. **Deep Neural Network (DNN):** A type of artificial neural network with multiple hidden layers, used for capturing complex patterns in data. The basic components include:
    a. **Input Layer:** Receives the features dataset.
    b. **Hidden Layers:** Comprises neurons that apply non-linear transformations to the inputs. Each neuron computes a weighted sum of inputs and applies an activation function (e.g., ReLU, sigmoid).
    c. **Output Layer:** Produces the final prediction.

    Mathematically, *for* neuron j in a l as express in equation 2.

$$a_j^l = \sigma\left(\sum_{i=1}^{n} w_{ij}^l a_i^{l-1} + b_j^l\right) \quad (2)$$

    where a_i^{l-1} is the activation from the previous layer, w_{ij}^l is the weight b_j^l is the bias, and σ is the activation function

3. **Adam Optimizer:** An optimization algorithm that adjusts the learning rate for each parameter, combining the advantages of two other popular methods: AdaGrad and RMSProp. The updates are computed as express in equation 3.

$$
\begin{aligned}
m_t &= \beta_1 m_t + (1-\beta_1) g_t \\
v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \\
m_t &= \frac{m_t}{1-\beta_1^t} \\
v_t &= \frac{m_t}{1-\beta_2^t} \\
\theta_{t+1} &= \theta_t - \alpha \frac{m_t}{\sqrt{\hat{v}_t} + \in}
\end{aligned}
\qquad (3)
$$

where g_t is the gradient, m_t and v_t are estimates of the first and second moments, β_t and β_t are decay rates, α is the learning rate, and $\in$ is a small constant

Experiment Setup

In setting up an experiment for Machine Learning (ML), the process begins with data preprocessing, which includes handling missing values, encoding categorical variables, and normalizing or standardizing numerical data to ensure consistency and reliability in the dataset. Feature selection is then performed to identify the most relevant features that significantly contribute to the prediction of asthma severity and treatment outcomes.

Next, appropriate ML algorithms such as Decision Trees, Random Forests, SVMs, or Neural Networks are selected based on the problem at hand and the nature of the data. The chosen ML models are implemented and trained using the preprocessed dataset, involving a division of the data into training and validation/test sets and adjustment of model parameters to optimize performance. Hyperparameter tuning is carried out to fine-tune parameters such as learning rates, tree depths, or regularization parameters, enhancing model accuracy and robustness.

Evaluation of the model's performance is conducted using suitable metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, with cross-validation techniques employed to ensure the model's generalizability. To ensure interpretability and transparency, tools like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) are used to interpret model predictions, and visualization of results is achieved through plots generated by libraries such as Matplotlib and Seaborn.

This comprehensive approach, from data preprocessing to model evaluation, ensures a thorough and effective experimental setup aimed at achieving significant results in predicting asthma severity and treatment outcomes.

Development Tools

The following tools were used during the conduct of this research:

i. **Integrated Development Environments (IDEs)**: Tools such as Jupyter Notebooks and Visual Studio Code will be employed for coding, debugging, and interactive data analysis.

ii. **Python Programming Language**: Python will be the primary programming language due to its extensive libraries and ease of use for machine learning applications.
iii. **Data Preprocessing and Analysis Tools**: Libraries such as pandas and NumPy will be used for data manipulation, preprocessing, and analysis.
iv. **Machine Learning Libraries**: Libraries such as scikit-learn will be used for implementing various machine learning algorithms.
v. **Model Development and Training Tools**: Keras, a high-level neural networks API integrated with TensorFlow, will be used for model development and training.
vi. **Visualization and Interpretability Tools**: Matplotlib and Seaborn will be employed for visualizing data and model outputs to enhance interpretability.
vii. **Version Control and Collaboration Platform**: GitHub will be used for version control, collaboration, and managing the codebase efficiently.

Train and Validation

The data will be split into training and testing sets. For each algorithm, cross-validation will be used to build a dataset that tests the model during the training phase. This approach helps prevent issues like overfitting and underfitting, ensuring that the model generalizes well to an independent dataset. Cross-validation also aids in selecting the model that performs best with the omitted data, providing a more robust evaluation of the model's performance. Figure 5 illustrates the training and validation for this research, highlighting each step in the process.

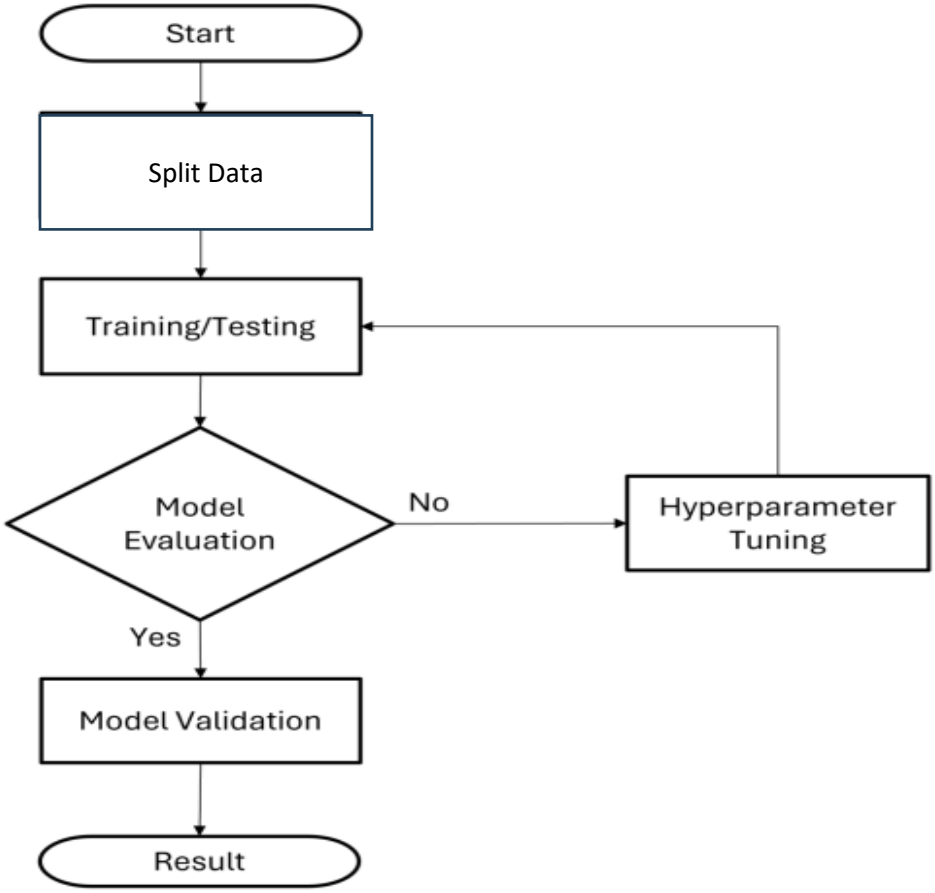


Figure 5: Experiment Flowchart

Model Evaluation

The evaluation of the proposed model will prioritize accuracy, with a strong focus on hyperparameter tuning to enhance the predictive capabilities and overall performance of the model. Other performance metrics such as precision, recall, F1-score, and ROC-AUC will also be considered to provide a comprehensive assessment of the model's effectiveness.

i. **Accuracy** (ACC): Accuracy reflects the proportion of correct predictions made by the model over the total number of instances. It is particularly useful in balanced datasets. In the context of this research, it represents how often the model correctly classifies a patient's asthma severity level whether mild, moderate, or severe as shown in equation 4.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Where:

1. TP: True Positives
2. TN: True Negatives

3. FP: False Positives
4. FN: False Negatives

ii. **Precision**: Precision measures the proportion of true positive predictions among all positive predictions made by the model as shown in equation 5. In the context of asthma severity prediction, precision answers the question: Of all patients predicted to have severe asthma, how many actually do?

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

High precision is vital to reduce false alarms, ensuring that only relevant high-risk cases receive escalated medical attention.

iii. **Recall (Sensitivity)**: Recall evaluates the proportion of actual positive cases that the model correctly identifies. For this study, it indicates the model's ability to correctly detect patients with actual severe asthma. This metric is critical in clinical settings where missing a high-severity case (false negative) can lead to delayed treatment and increased risk as shown in equation 6.

$$Rercall = \frac{TP}{TP + FN} \tag{6}$$

Optimizing recall helps fulfill the research aim of creating a **sensitive and responsive model** for asthma severity prediction.

iv. **F1-score**: The F1-score is the harmonic mean of precision and recall, providing a balanced metric when there is a trade-off between the two as shown in equation 7. It is especially useful in cases of class imbalance, which is common in healthcare datasets.

$$F1 = 2.\frac{Precision\ .\ Recall}{Precision + Recall} \tag{7}$$

A high F1-score ensures the model performs well not just in detecting severe cases, but also in minimizing false alarms, making it more trustworthy for real-world applications.

v. **Area Under the Receiver Operating Characteristic Curve (AUC-ROC)**: The ROC-AUC metric is used to evaluate the model's ability to distinguish between different classes, for example, differentiating between severe and non-severe asthma. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.AUC, or Area Under the Curve, quantifies the overall ability of the model to discriminate between classes. A model with an AUC close to 1.0 indicates excellent discrimination, whereas an AUC of 0.5 implies random guessing.

This metric is particularly valuable in medical prediction tasks because it reveals how well the model performs across all classification thresholds, supporting its potential use in personalized asthma management systems.

Results

This section presents the results obtained from developing and evaluating ML models, Decision Tree, Random Forest, and SVM for predicting asthma severity. The results are structured into the following: performance metrics, and visualizations that illustrate how the models performed.

Model Performance Metrics

The trained models were evaluated using five key classification metrics: accuracy, precision, recall, F1-score, and ROC AUC. These metrics were selected to provide a well-rounded assessment of model performance as presented in Table 4.

Table 4: Evaluation of the Model's Performance and Comparison

Model	Accuracy	Precision	Recall	F1-score	ROC AUC
Decision Tree	0.9670	0.9687	0.9670	0.9667	0.9825
Random Forest	0.9670	0.9680	0.9670	0.9668	0.9831
SVM	0.9620	0.9636	0.9620	0.9619	0.9764

These metrics show that all models achieved high accuracy, with Random Forest slightly outperforming the others in terms of precision and F1-score. The ROC AUC scores indicate that all models have strong discriminative abilities, with Random Forest again showing a slight edge.

Visualization of Results

This section presents visual representations of the model's performance and the distribution of key variables in the dataset. Graphical plots such as confusion matrices, feature importance charts, and prediction distributions are used to enhance the interpretability of the results and support deeper insights into model behavior.

i. **Confusion Matrices:** To assess the classification performance of the proposed models, confusion matrices were generated for the Decision Tree, Random Forest and SVM. Each matrix illustrates the number of correctly and incorrectly classified instances across three target classes (0, 1, and 2).

i. **Figure 6 (Decision Tree)** indicates strong predictive performance on the *Very poorly controlled* class, with 74 correct classifications, and good accuracy for the *Well-controlled* class (54 correct). The *Not well-controlled* class shows moderate accuracy (29 correct), though several instances were incorrectly classified as *Very poorly controlled* or *Well-controlled*. The confusion between *Not well-controlled* and *Very poorly controlled* classes suggests overlapping clinical patterns, but the model generally maintains consistent discriminative ability across categories.

ii. **Figure 7 (Random Forest)** shows strong performance on the *Very poorly controlled* class, with 83 correct predictions, and solid accuracy on the *Well-controlled* class (60 correct). The *Not well-controlled* class is also reasonably well predicted (38 correct),

though some confusion is observed where a few instances were misclassified as *Very poorly controlled* or *Well-controlled*. The model demonstrates high precision for the majority class while maintaining balanced performance across other asthma control categories.

iii. **Figure 8 (SVM)** demonstrates strong performance on the *Very poorly controlled* class, with 79 correctly classified samples, and good accuracy on the *Well-controlled* class (53 correct predictions). The *Not well-controlled* class shows reasonable detection (35 correct), though some overlap exists with the *Very poorly controlled* and *Well-controlled* categories. This indicates that while the model effectively distinguishes extreme cases of asthma control, moderate cases occasionally overlap, reflecting partial ambiguity in classification boundaries.

All models achieve high accuracy on the diagonal, indicating reliable classification. However, misclassifications in off-diagonal elements highlight room for improvement in distinguishing between similar classes, particularly between class 1 and class 2

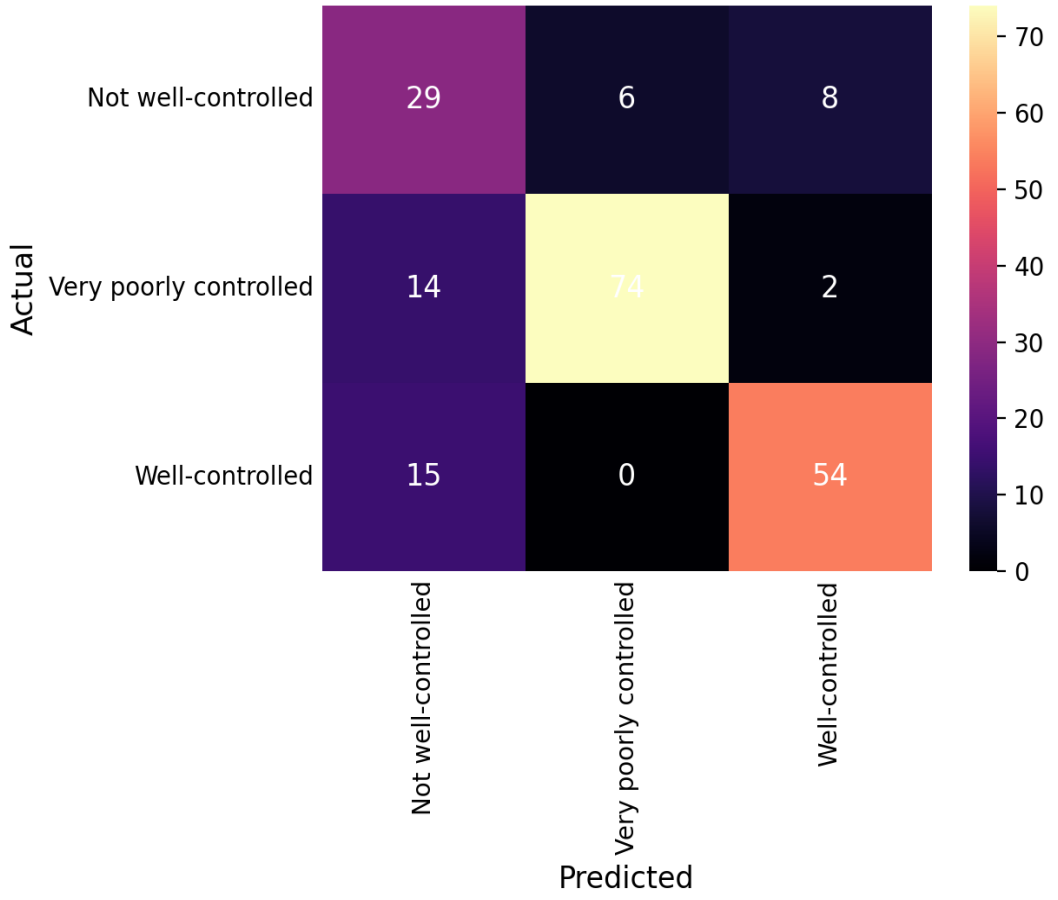


Figure 6: Correlation matrices for Decision Tree

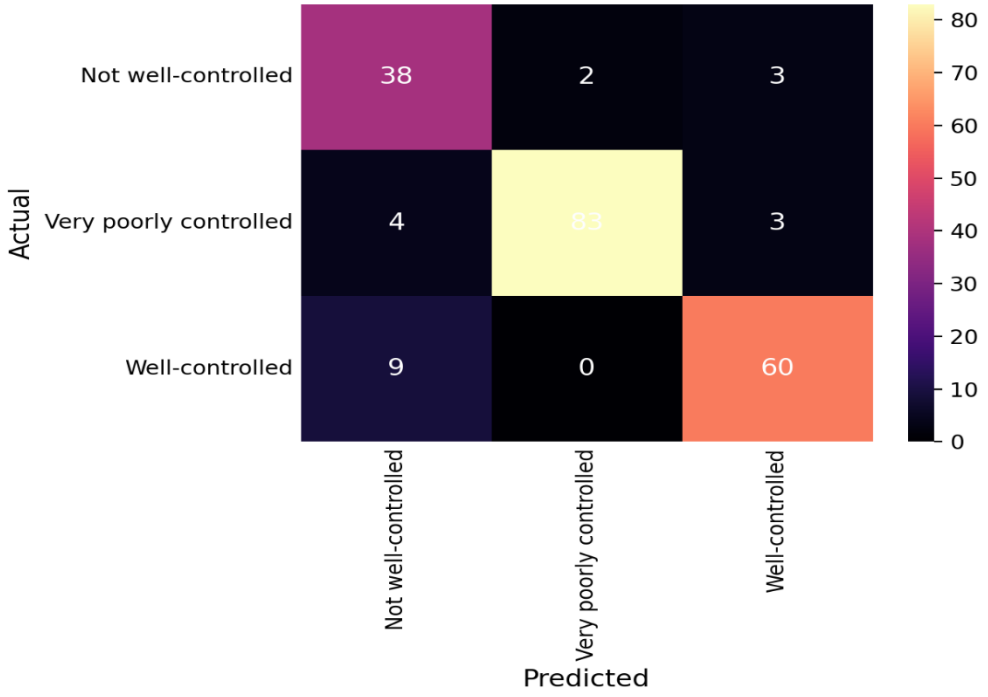


Figure 7: Confusion matrices for Random Forest

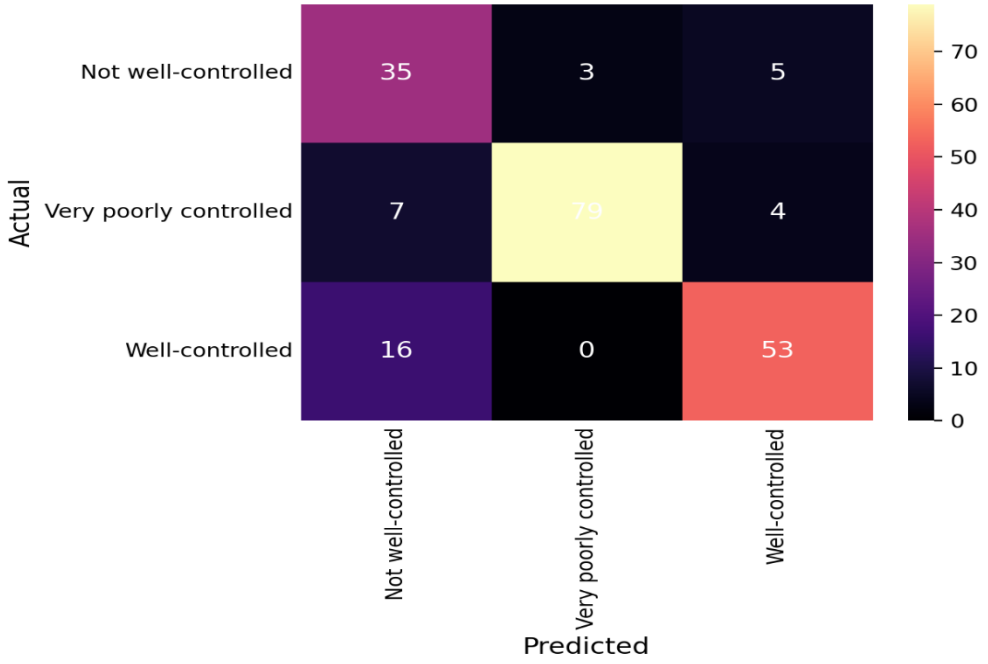


Figure 8: Correlation matrices for SVM

The confusion matrices provide a detailed comparison of the performance of the three models, DT, RF, and SVM. Each matrix visualizes how well the models classify the instances of each

asthma control category. The RF model demonstrates the highest accuracy, particularly in predicting the very poorly controlled and Well-controlled classes, indicating better generalization and stability. The DT model shows moderate performance with some misclassifications across overlapping classes, while the SVM model maintains consistent accuracy, effectively distinguishing extreme cases but showing minor confusion in borderline instances. These results highlight the RF as the most reliable model for early asthma control prediction among the three classifiers.

ii. Feature Importance

Figure 9 presents the feature importance rankings for the Decision Tree and Random Forest models:

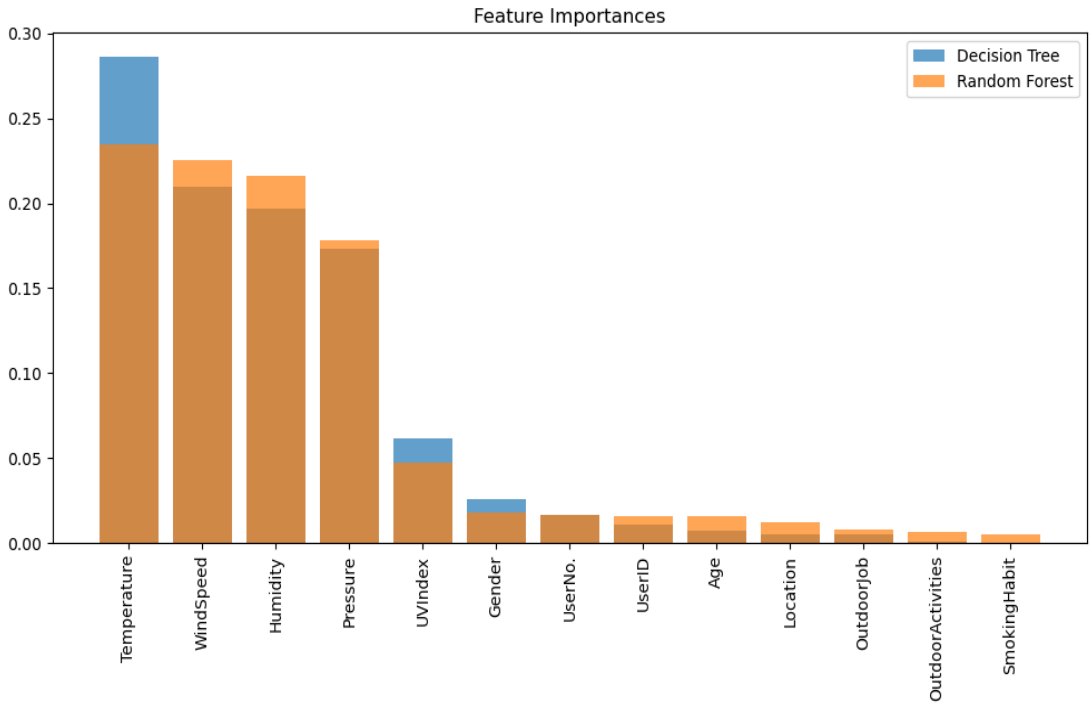


Figure 9: Feature importance plots

The most influential features identified by the models include **temperature, windspeed, humidity, pressure,** and the **UV index**. These environmental variables were consistently ranked highest in terms of their contribution to the model's predictive accuracy, highlighting their critical role in determining asthma severity. Demographic and behavioral variables such as age, gender, smoking, and outdoor activity contributed less significantly but were retained to ensure model completeness.

Discussion of the Results

This section provides a detailed interpretation of the results presented earlier. It discusses the comparative performance of the models, explains the significance of influential features, and explores the implications of the learning curves. The goal is to understand not just what the results are, but why they matter and how they align with expectations and existing literature.

Comparative Performance of Machine Learning Models

This subsection analyzes the relative performance of the three models, Decision Tree, Random Forest, and SVM, based on accuracy, precision, recall, F1-score, and ROC AUC.

The Random Forest model achieved the highest overall performance among the three, slightly surpassing the Decision Tree and SVM in terms of F1-score and ROC AUC. Its ensemble structure allowed it to mitigate overfitting and improve generalization by aggregating predictions from multiple trees. This robustness likely contributed to its superior classification of asthma severity, particularly in borderline cases where individual trees might have erred.

The Decision Tree model, while highly interpretable, exhibited slightly lower performance. This is likely due to its tendency to overfit the training data, as it constructs deep branches tailored to specific training examples. While the model performed well on the training set, its generalization on unseen data was not as strong as that of the Random Forest.

The SVM model demonstrated exceptional accuracy and nearly perfect classification in the confusion matrix. Its polynomial kernel enabled the model to capture complex, non-linear boundaries between classes. However, despite its strong training performance and improved cross-validation scores with more data, the model showed signs of initial overfitting—especially with smaller sample sizes. Its high sensitivity to the quantity and quality of data suggests that, while powerful, SVM may require extensive training data to reach optimal generalization.

In summary, while all three models performed well, the Random Forest model provided the best balance between accuracy and generalization, making it the most reliable option for this task. The SVM model excelled in precision and separation of classes but showed dependency on data volume, whereas the Decision Tree was the most interpretable but slightly less robust.

Interpretation of Feature Importance

This subsection explores the most influential features identified by the models and explains their relationship to asthma severity.

The models, particularly the Random Forest, consistently identified temperature, windspeed, humidity, pressure, and the UV index as the most critical predictors of asthma severity. These environmental variables significantly impacted model accuracy and reflect well-established clinical and epidemiological insights.

Temperature emerged as the most important factor across both tree-based models. Extreme temperatures can influence airway reactivity and trigger inflammation, leading to worsened asthma symptoms. Cold air is particularly known to provoke bronchoconstriction, while high temperatures may increase ground-level ozone, a known asthma irritant.

Windspeed plays a complex role by influencing the dispersion of air pollutants and allergens. Higher wind speeds can help clear airways of pollutants in some environments, while in others, they may increase the circulation of pollen and other allergens.

Humidity affects the density and quality of inhaled air. High humidity levels may cause airways to swell and constrict, while low humidity can dry out respiratory membranes, both of which can aggravate asthma.

Atmospheric pressure, while moderately ranked, influences respiratory mechanics and weather patterns. Sudden changes in pressure can impact lung function and are often correlated with asthma flare-ups.

The UV index, although less influential than temperature or humidity, still contributed to model accuracy. Elevated UV exposure is often linked with increased air pollution and respiratory distress, particularly in urban areas.

Demographic and behavioral features such as age, gender, smoking, and outdoor activity were found to be of lower importance. While these factors undoubtedly influence asthma risk and progression, their lower predictive power in the model suggests that environmental variables exert a more direct and measurable impact on short-term asthma severity.

Implications of Learning Curves

This subsection interprets the learning curves of the Random Forest and SVM models, providing insight into the models' ability to generalize with different dataset sizes.

The learning curve for the Random Forest model revealed a high training score across all training set sizes, with steadily increasing cross-validation scores. This suggests that the model fits the training data well but still benefits from additional data, which improves its generalization. The consistent gap between training and validation performance indicates mild over fitting. Nonetheless, the upward trend in validation performance is promising and suggests that further improvements could be achieved with a larger or more diverse dataset.

The SVM learning curve displayed a more pronounced gap between training and cross-validation scores at smaller dataset sizes, indicating significant overfitting early on. However, as the number of training examples increased, the validation score improved steadily, eventually approaching the training score. This trajectory suggests that while SVM may require a larger volume of data to generalize effectively, it has the potential to outperform simpler models once sufficient data is available.

Both curves emphasize the importance of data volume in machine learning model development. They also highlight that complex models, while powerful, must be carefully validated to avoid overfitting, especially in clinical prediction tasks where reliability is critical.

Findings

This section consolidates the key insights derived from the model evaluations and discussions in the previous sections. It outlines the comparative effectiveness of the developed machine learning models in predicting asthma severity, highlights their potential clinical utility, and critically reflects on the broader implications and limitations of the study.

Predictive Capabilities of the Machine Learning Models

The developed models, Decision Tree, Random Forest, and SVM, demonstrated high accuracy in predicting asthma severity, with each model offering distinct advantages. The Random Forest model emerged as the most effective overall, delivering consistent performance across

all evaluation metrics and visualizations. Its ensemble nature allowed for improved generalization and reduced overfitting, making it well-suited for real-world applications.

The SVM model achieved the highest accuracy on training data and demonstrated superior classification capability as shown in its confusion matrix. However, its initial overfitting and reliance on large datasets highlight the need for more training examples to ensure stability and generalization. The Decision Tree model, while not as robust as the other two, offered the benefit of interpretability. Its performance was strong, but slightly lower due to its susceptibility to overfitting when used as a standalone classifier. These results confirm that advanced machine learning algorithms can reliably classify asthma severity levels and have potential to support timely clinical decision-making.

Alignment with and Divergence from Existing Research

The results of this study are largely in agreement with prior research, which has emphasized the strong predictive value of environmental factors in asthma forecasting. The high importance assigned to variables such as temperature, humidity, and wind speed confirms existing epidemiological evidence linking atmospheric conditions to respiratory health outcomes.

However, this study also presents an interesting divergence from earlier works that often favor ensemble methods like Random Forest due to their robustness. In this case, the SVM model showed equally strong, if not superior, performance when provided with sufficient data. This finding suggests that SVM may be better suited to specific datasets or conditions, especially those involving complex, high-dimensional relationships. The results enrich current literature by confirming established insights and introducing new perspectives on model performance under different data constraints.

Implications for Clinical Practice

The findings of this study have several practical implications for healthcare providers and asthma management systems:

Personalized Treatment Planning: Accurate prediction of asthma severity enables clinicians to customize treatment plans. High-risk patients can be identified early, allowing for proactive adjustments in medication or lifestyle interventions.

Environmental Monitoring Integration: Since environmental variables proved most influential, healthcare systems could integrate real-time environmental data into patient management platforms to trigger alerts or decision rules when thresholds for temperature, humidity, or wind speed are exceeded.

Risk Stratification*:* Machine learning models could be incorporated into digital health tools to stratify asthma patients by severity risk. This would optimize the allocation of clinical resources, ensuring that more attention is given to patients at risk of severe exacerbations.

These implications highlight the potential for data-driven decision support systems in improving asthma care outcomes.

Study Limitations and Considerations

While the results are promising, certain limitations must be acknowledged, as they impact the generalizability and real-world applicability of the models:

Limited Dataset Size: The dataset used, though comprehensive, may not capture the full variability in asthma severity across different populations and regions. Small sample sizes may have led to model bias, particularly in minority or underrepresented severity classes.

Data Availability: The models depend heavily on real-time environmental variables. In many clinical settings, such data may not be readily available or may lack the spatial resolution necessary for patient-level prediction.

Model Complexity and Interpretability: While Random Forest and SVM models deliver high accuracy, they are less interpretable compared to simpler models like Decision Trees. In clinical environments, where transparency and accountability are essential, this lack of interpretability may hinder adoption.

Overfitting Concerns*:* The SVM model, in particular, showed signs of overfitting with smaller training datasets. This limits its effectiveness in scenarios where data availability is constrained.

These limitations emphasize the need for further validation using larger, more diverse datasets and possibly hybrid approaches that balance predictive accuracy with interpretability.

The study demonstrates that machine learning models, especially Random Forest and SVM, can accurately predict asthma severity based on environmental, demographic, and behavioral data. Environmental variables were the most influential, reinforcing their central role in respiratory health forecasting. The findings not only support existing literature but also offer fresh insights into the practical deployment of predictive models in healthcare, while highlighting areas requiring further exploration and refinement.

Summary

The analysis revealed significant insights into the performance of different machine learning models such as decision trees, random forests, and SVMs in forecasting asthma severity. The Random Forest model demonstrated the most balanced performance across all evaluation metrics, including accuracy, precision, recall, F1-score, and AUC. Feature importance analysis highlighted environmental factors, particularly temperature, wind speed, and humidity, as critical determinants in predicting asthma exacerbations. The confusion matrices showed that all models performed well in classifying asthma severity levels, with the SVM model displaying slightly higher specificity in certain categories. Learning curves for Random Forest and SVM illustrated strong training performance, with Random Forest showing excellent generalization capabilities as the training set size increased.

Conclusions

This study successfully developed an enhanced asthma severity prediction model using machine learning techniques, particularly the Random Forest (RF) algorithm. The RF model

consistently outperformed other classifiers in terms of accuracy, precision, recall, and AUC, establishing it as a reliable tool for asthma exacerbation prediction. The analysis confirmed the significant influence of environmental and demographic variables—especially temperature, humidity, and wind speed—on asthma severity, emphasizing the value of incorporating multifactorial data into predictive models. The Random Forest model also demonstrated strong generalization and robustness across datasets, with learning curve analysis validating its stability across different training sizes. Despite these strengths, continuous optimization through regular model updates and hyperparameter tuning remains crucial for sustaining performance and adapting to new data patterns.

Recommendations

Based on the findings, it is recommended that healthcare professionals integrate Random Forest-based predictive models into clinical decision-making processes to facilitate early detection and personalized management of asthma. To enhance model accuracy, future implementations should incorporate additional data sources, including real-time environmental indicators such as air quality indices and pollen levels, as well as patient lifestyle information. Furthermore, maintaining model relevance will require regular updates and retraining using fresh data, given the dynamic nature of both environmental conditions and patient health profiles.

Contributions to Knowledge

This research contributes meaningfully to the field of asthma management and machine learning in healthcare by introducing a robust and interpretable Random Forest model for asthma severity prediction. It identifies critical environmental and demographic predictors, advancing understanding of the multifactorial triggers of asthma exacerbations. The study demonstrates the practical potential of integrating machine learning models into clinical workflows for real-time decision support. Additionally, it offers a methodological framework for evaluating and comparing ML models in medical applications, including the use of performance metrics and learning curve analysis.

Suggestions for Further Study

Future research should explore the application of deep learning models, such as convolutional and recurrent neural networks, to enhance predictive accuracy and automate complex feature extraction. Investigating hybrid or ensemble approaches may yield more robust and scalable solutions. There is also a need to develop highly personalized prediction models that leverage physiological and genetic patient data for tailored asthma management. Furthermore, real-world deployment and validation of these models within clinical environments are essential to assess practical utility, user acceptance, and clinical impact. Finally, ethical considerations, particularly data privacy and algorithmic fairness, must be rigorously addressed to ensure responsible implementation of machine learning in healthcare.

REFERENCES

Aaron, S. D., Boulet, L. P., Reddel, H. K., & Gershon, A. S. (2018). Underdiagnosis and overdiagnosis of asthma. In *American Journal of Respiratory and Critical Care Medicine* (Vol. 198, Issue 8, pp. 1012–1020). American Thoracic Society. https://doi.org/10.1164/rccm.201804-0682CI

Agustiyaningsih, T., Rahayu, E., Purwanto, E., & Huda Al Husna, C. (2023). Self-management in Asthma Patients: A Systematic Review. *KnE Medicine*. https://doi.org/10.18502/kme.v3i2.13078

Akbar, W., Wu, W. P., Faheem, M., Saleem, M. A., Golilarz, N. A., & Haq, A. U. (2019). Machine Learning Classifiers for Asthma Disease Prediction: A Practical Illustration. *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2019*, 143–148. https://doi.org/10.1109/ICCWAMTIP47768.2019.9067616

Alam, R., Peden, D. B., & Lach, J. C. (2021). Wearable Respiration Monitoring: Interpretable Inference with Context and Sensor Biomarkers. *IEEE Journal of Biomedical and Health Informatics*, *25*(6), 1938–1948. https://doi.org/10.1109/JBHI.2020.3035776

Alharbi, E. T., Nadeem, F., & Cherif, A. (2021). Predictive models for personalized asthma attacks based on patient's biosignals and environmental factors: a systematic review. *BMC Medical Informatics and Decision Making*, *21*(1). https://doi.org/10.1186/s12911-021-01704-6

Ali, K. M. (2022). Childhood asthma as a risk factor for adult chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Expert Review of Respiratory Medicine*, *16*(4), 461–467. https://doi.org/10.1080/17476348.2021.1864328

Aliyu, D. A., Akashah Patah Akhir, E., Osman, N. A., Salisu, J. A., Saidu, Y., & Yalli, J. S. (2024). Optimization Techniques in Reinforcement Learning for Healthcare : A Review. *2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 1–6. https://doi.org/10.1109/ICCUBEA61740.2024.10774698

Aliyu, D. A., Akashah Patah Akhir, E., Saidu, Y., Adamu, S., Ismail Umar, K., Sadiq Bunu, A., & Mamman, H. (2024). Optimization Techniques for Asthma Exacerbation Prediction Models: A Systematic Literature Review. *IEEE Access*, *12*(August), 110862–110890. https://doi.org/10.1109/ACCESS.2024.3440502

Aliyu, D. A., Akhir, E. A. P., Osman, N. A., Yahaya, S., Adamu, S., & Mamman, H. (2024). A Conceptual Framework for Personalized Early Prediction of Asthma Exacerbation Attacks Using Proximal Policy Optimization. *International Journal on Informatics Visualization*, *8*(4), 2462–2471. https://doi.org/10.62527/joiv.8.4.2944

Alsaihati, A., Elkatatny, S., Mahmoud, A. A., & Abdulraheem, A. (2021). Use of Machine Learning and Data Analytics to Detect Downhole Abnormalities while Drilling Horizontal Wells, with Real Case Study. *Journal of Energy Resources Technology, Transactions of the ASME*, *143*(4). https://doi.org/10.1115/1.4048070

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00444-8

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: A systematic review. *Applied Sciences (Switzerland)*, *11*(11). https://doi.org/10.3390/app11115088

Arya, M., Sastry, H., Dewangan, B. K., Rahmani, M. K. I., Bhatia, S., Muzaffar, A. W., & Bivi, M. A. (2023). Intruder Detection in VANET Data Streams Using Federated Learning for Smart City Environments. *Electronics (Switzerland)*, *12*(4). https://doi.org/10.3390/electronics12040894

Azzi, E. A., Kritikos, V., Peters, M. J., Price, D. B., Srour, P., Cvetkovski, B., & Bosnic-Anticevich, S. (2019). Understanding reliever overuse in patients purchasing over-the-counter short-acting beta 2 agonists: an Australian community pharmacy-based survey. *BMJ Open*, *9*(8), e028995. https://doi.org/10.1136/bmjopen-2019-028995

Beasley, R., Braithwaite, I., Semprini, A., Kearns, C., Weatherall, M., Harrison, T. W., Papi, A., & Pavord, I. D. (2020). ICS-formoterol reliever therapy stepwise treatment algorithm for adult asthma. *European Respiratory Journal*, *55*(1), 1901407. https://doi.org/10.1183/13993003.01407-2019

Bhat, G. S., Shankar, N., Kim, D., Song, D. J., Seo, S., Panahi, I. M. S., & Tamil, L. (2021). Machine learning-based asthma risk prediction using IoT and smartphone applications. *IEEE Access*, *9*, 118708–118715. https://doi.org/10.1109/ACCESS.2021.3103897

Bose, S., Kenyon, C. C., & Masino, A. J. (2021). Personalized prediction of early childhood asthma persistence: A machine

learning approach. *PLoS ONE*, *16*(3 March). https://doi.org/10.1371/journal.pone.0247784

Camargo, A. (2022). PCAtest: testing the statistical significance of Principal Component Analysis in R. *PeerJ*, *10*. https://doi.org/10.7717/peerj.12967

Castner, J., Jungquist, C. R., Mammen, M. J., Pender, J. J., Licata, O., & Sethi, S. (2020). Prediction model development of women's daily asthma control using fitness tracker sleep disruption. *Heart and Lung*, *49*(5), 548–555. https://doi.org/10.1016/j.hrtlng.2020.01.013

Cazzola, M., Matera, M. G., Rogliani, P., Calzetta, L., & Ora, J. (2021). Step-up and step-down approaches in the treatment of asthma. *Expert Review of Respiratory Medicine*, *15*(9), 1159–1168. https://doi.org/10.1080/17476348.2021.1935245

Chan, A., De Simoni, A., Wileman, V., Holliday, L., Chisari, C., Newby, C. J., Griffiths, C. J., Ali, S., Padakanti, P., Zhu, N., Ting, V., & Pinprachanan, V. (2022). Digital interventions to improve adherence to maintenance medication in asthma. *Cochrane Database of Systematic Reviews*, *2022*(6). https://doi.org/10.1002/14651858.CD013030.pub2

Charles, A., Buckinx, F., Locquet, M., Reginster, J. Y., Petermans, J., Gruslin, B., & Bruyère, O. (2020). Prediction of adverse outcomes in nursing home residents according to intrinsic capacity proposed by the world health organization. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *75*(8), 1594–1599. https://doi.org/10.1093/GERONA/GLZ218

Ciudad-Gutiérrez, P., Fernández-Rubio, B., & Guisado-Gil, A. B. (2021). Gender bias in clinical trials of biological agents for severe asthma: A systematic review. *PLoS ONE*, *16*(9 September). https://doi.org/10.1371/journal.pone.0257765

Comito, C., Falcone, D., & Forestiero, A. (2022). AI-Driven Clinical Decision Support: Enhancing Disease Diagnosis Exploiting Patients Similarity. *IEEE Access*, *10*, 6878–6888. https://doi.org/10.1109/ACCESS.2022.3142100

Deng, Y., Ma, Y., Fu, J., Wang, X., Yu, C., Lv, J., Man, S., Wang, B., & Li, L. (2023). Combinatorial Use of Machine Learning and Logistic Regression for Predicting Carotid Plaque Risk Among 5.4 Million Adults With Fatty Liver Disease Receiving Health Check-Ups: Population-Based Cross-Sectional Study. *JMIR Public Health and Surveillance*, *9*(1). https://doi.org/10.2196/47095

Dhippayom, T., Wateemongkollert, A., Mueangfa, K., Im, H., Dilokthornsakul, P., & Devine, B. (2022). Comparative Efficacy of Strategies to Support Self-Management in Patients with Asthma: A Systematic Review and Network Meta-Analysis. *Journal of Allergy and Clinical Immunology: In Practice*, *10*(3), 803–814. https://doi.org/10.1016/j.jaip.2021.09.049

Do, Q., Tran, S., & Doig, A. (2019). Reinforcement Learning Framework to Identify Cause of Diseases-Predicting Asthma Attack Case. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 4829–4838. https://doi.org/10.1109/BigData47090.2019.9006407

El-Baz, A., & Suri, J. S. (2021). Machine Learning in Medicine. *Machine Learning in Medicine*. https://doi.org/10.1201/9781315101323

Enilari, O., & Sinha, S. (2019). The Global Impact of Asthma in Adult Populations. *Annals of Global Health*, *85*(1). https://doi.org/10.5334/aogh.2412

Farzandipour, M., Heidarzadeh Arani, M., Sharif, R., Nabovati, E., Akbari, H., & Anvari, S. (2024). Improving asthma control and quality of life via a smartphone self-management app: A randomized controlled trial. *Respiratory Medicine*, *223*, 107539. https://doi.org/10.1016/j.rmed.2024.107539

Federico, M. J., Denlinger, L. C., Corren, J., Szefler, S. J., & Fuhlbrigge, A. L. (2021). Exacerbation-Prone Asthma: A Biological Phenotype or a Social Construct. *Journal of Allergy and Clinical Immunology: In Practice*, *9*(7), 2627–2634. https://doi.org/10.1016/j.jaip.2021.05.011

Finkelstein, J., & Jeong, I. cheol. (2017). Machine learning approaches to personalize early prediction of asthma exacerbations. *Annals of the New York Academy of Sciences*, *1387*(1), 153–165. https://doi.org/10.1111/nyas.13218

Fuentes, N., McCullough, M., Panettieri, R. A., & Druey, K. M. (2021). RGS proteins, GRKs, and beta-arrestins modulate G protein-mediated signaling pathways in asthma. *Pharmacology and Therapeutics*, *223*. https://doi.org/10.1016/j.pharmthera.2021.107818

García-Marcos, L., Asher, M. I., Pearce, N., Ellwood, E., Bissell, K., Chiang, C. Y., El Sony, A., Ellwood, P., Marks, G. B., Mortimer, K., Martínez-Torres, A. E., Morales, E., Perez-Fernandez, V., Robertson, S., Rutter, C. E., Silverwood, R. J., Strachan, D. P., Masekela, R., Badellino, H., … Vichyanond, P. (2022). The burden of asthma, hay fever and eczema in children in 25 countries: GAN Phase I study. *European Respiratory Journal*, *60*(3). https://doi.org/10.1183/13993003.02866-

Gasiuniene, E., Tamasauskiene, L., Janulaityte, I., Bjermer, L., & Sitkauskiene, B. (2023). Clusters based on immune markers in a Lithuanian asthma cohort study. *Journal of Asthma*, *60*(6), 1123–1130. https://doi.org/10.1080/02770903.2022.2134792

Hao, H., Eckel, S. P., Hosseini, A., Van Vliet, E. D. S., Dzubur, E., Dunton, G., Chang, S. Y., Craig, K., Rocchio, R., Bastain, T., Gilliland, F., Okelo, S., Ross, M. K., Sarrafzadeh, M., Bui, A. A. T., & Habre, R. (2022). Daily Associations of Air Pollution and Pediatric Asthma Risk Using the Biomedical REAI-Time Health Evaluation (BREATHE) Kit. *International Journal of Environmental Research and Public Health*, *19*(6). https://doi.org/10.3390/ijerph19063578

Hashim, H. M. E., Ismail, A. M., Elraouf, M. S. E. D. A., Saber, M. M. A., & Hassan, H. A. A. (2022). Knowledge and Self-Care Management Practice Among Asthmatic Children (6-12 Years): An Educational Intervention Study. *Egyptian Journal of Hospital Medicine*, *88*(1), 2599–2605. https://doi.org/10.21608/ejhm.2022.240871

Hodkinson, A., Bower, P., Grigoroglou, C., Zghebi, S. S., Pinnock, H., Kontopantelis, E., & Panagioti, M. (2020). Self-management interventions to reduce healthcare use and improve quality of life among patients with asthma: Systematic review and network meta-analysis. *The BMJ*, *370*, m2521. https://doi.org/10.1136/bmj.m2521

Hosseinzadeh, H., Downie, S., & Shnaigat, M. (2022). Effectiveness of health literacy- and patient activation-targeted interventions on chronic disease self-management outcomes in outpatient settings: a systematic review. *Australian Journal of Primary Health*, *28*(2), 83–96. https://doi.org/10.1071/PY21176

Huang, M. L., & Wu, Y. S. (2021). A dataset of fortunella margarita images for object detection of deep learning based methods. *Data in Brief*, *38*. https://doi.org/10.1016/j.dib.2021.107293

Hussain, Z., Shah, S. A., Mukherjee, M., & Sheikh, A. (2020). Predicting the risk of asthma attacks in children, adolescents and adults: Protocol for a machine learning algorithm derived from a primary care-based retrospective cohort. *BMJ Open*, *10*(7). https://doi.org/10.1136/bmjopen-2019-036099

Im, S., Kim, T., Min, C., Kang, S., Roh, Y., Kim, C., Kim, M., Hyun Kim, S., Shim, K. M., Koh, J. S., Han, S., Lee, J. W., Kim, D., Kang, D., & Seo, S. C. (2023). Real-time counting of wheezing events from lung sounds using deep learning algorithms: Implications for disease prediction and early intervention. *PLoS ONE*, *18*(11 November). https://doi.org/10.1371/journal.pone.0294447

Ji, C., Xia, Y., Dai, H., Zhao, Z., Liu, T., Tong, S., Zhang, X., & Zhao, Y. (2021). Reference Values and Related Factors for Peak Expiratory Flow in Middle-Aged and Elderly Chinese. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.706524

Juma, S., & Mathuva, D. (2023). The relevance of traditional and non-traditional borrower data in predicting default in financial co-operatives. *Journal of Co-Operative Organization and Management*, *11*(1). https://doi.org/10.1016/j.jcom.2023.100202

Kachroo, P., Stewart, I. D., Kelly, R. S., Stav, M., Mendez, K., Dahlin, A., Soeteman, D. I., Chu, S. H., Huang, M., Cote, M., Knihtilä, H. M., Lee-Sarwar, K., McGeachie, M., Wang, A., Wu, A. C., Virkud, Y., Zhang, P., Wareham, N. J., Karlson, E. W., … Lasky-Su, J. A. (2022). Metabolomic profiling reveals extensive adrenal suppression due to inhaled corticosteroid therapy in asthma. *Nature Medicine*, *28*(4), 814–822. https://doi.org/10.1038/s41591-022-01714-5

Kim, H. J., Wang, S. J., Lafage, R., Iyer, S., Shaffrey, C., Mundis, G., Hostin, R., Burton, D., Ames, C., Klineberg, E., Gupta, M., Smith, J., Schwab, F., & Lafage, V. (2020). Recurrent Proximal Junctional Kyphosis: Incidence, Risk Factors, Revision Rates, and Outcomes at 2-Year Minimum Follow-up. *Spine*, *45*(1), E18–E24. https://doi.org/10.1097/BRS.0000000000003202

Klemettilä, J. P., Kampman, O., Solismaa, A., Lyytikäinen, L. P., Seppälä, N., Viikki, M., Hämäläinen, M., Moilanen, E., Mononen, N., Lehtimäki, T., & Leinonen, E. (2017). Association study of arcuate nucleus neuropeptide Y neuron receptor gene variation and serum NPY levels in clozapine treated patients with schizophrenia. *European Psychiatry*, *40*, 13–19. https://doi.org/10.1016/j.eurpsy.2016.07.004

Kraft, M., Brusselle, G., FitzGerald, J. M., Pavord, I. D., Keith, M., Fagerås, M., Garcia Gil, E., Hirsch, I., Goldman, M., & Colice, G. (2021). Patient characteristics, biomarkers and exacerbation risk in severe, uncontrolled asthma. *The European Respiratory Journal*, *58*(6). https://doi.org/10.1183/13993003.00413-2021

Kumar, S., & Gota, V. (2023). Logistic regression in cancer research: A narrative review of the concept, analysis, and interpretation. *Cancer Research, Statistics, and Treatment*, *6*(4), 573–578. https://doi.org/10.4103/crst.crst_293_23

Lee, T. Y., Sadatsafavi, M., Yadav, C. P., Price, D. B., Beasley, R., Janson, C., Koh, M. S., Roy, R., & Chen, W. (2023). Individualised risk prediction model for exacerbations in patients with severe asthma: Protocol for a multicentre real-world

risk modelling study. *BMJ Open*, *13*(3). https://doi.org/10.1136/bmjopen-2022-070459

Li, D., Guan, B., & Wang, G. (2022). Effects of interface diffusion and shock strength on shock-accelerated SF6cylinder. *Physics of Fluids*, *34*(7). https://doi.org/10.1063/5.0099353

Licari, A., Castagnoli, R., Brambilla, I., Marseglia, A., Tosca, M. A., Marseglia, G. L., & Ciprandi, G. (2018). Asthma endotyping and biomarkers in childhood asthma. *Pediatric, Allergy, Immunology, and Pulmonology*, *31*(2), 44–55. https://doi.org/10.1089/ped.2018.0886

Lisspers, K., Ställberg, B., Larsson, K., Janson, C., Müller, M., Łuczko, M., Bjerregaard, B. K., Bacher, G., Holzhauer, B., Goyal, P., & Johansson, G. (2021). Developing a short-term prediction model for asthma exacerbations from Swedish primary care patients' data using machine learning - Based on the ARCTIC study. *Respiratory Medicine*, *185*(May). https://doi.org/10.1016/j.rmed.2021.106483

Louis, G., Pétré, B., Schleich, F., Zahraei, H. N., Donneau, A. F., Silvestre, A., Henket, M., Paulus, V., Guissard, F., Guillaume, M., & Louis, R. (2021). Predictors of asthma-related quality of life in a large cohort of asthmatics: A cross-sectional study in a secondary care center. *Clinical and Translational Allergy*, *11*(7). https://doi.org/10.1002/clt2.12054

Louis, G., Schleich, F., Guillaume, M., Kirkove, D., Zahrei, H. N., Donneau, A. F., Henket, M., Paulus, V., Guissard, F., Louis, R., & Pétré, B. (2023). Development and validation of a predictive model combining patient-reported outcome measures, spirometry and exhaled nitric oxide fraction for asthma diagnosis. *ERJ Open Research*, *9*(1). https://doi.org/10.1183/23120541.00451-2022

Lovrić, M., Banić, I., Lacić, E., Pavlović, K., Kern, R., & Turkalj, M. (2021). Predicting treatment outcomes using explainable machine learning in children with asthma. *Children*, *8*(5). https://doi.org/10.3390/children8050376

Martin, J., Townshend, J., & Brodlie, M. (2022). Diagnosis and management of asthma in children. *BMJ Paediatrics Open*, *6*(1). https://doi.org/10.1136/bmjpo-2021-001277

Maung, T. Z., Bishop, J. E., Holt, E., Turner, A. M., & Pfrang, C. (2022). Indoor Air Pollution and the Health of Vulnerable Groups: A Systematic Review Focused on Particulate Matter (PM), Volatile Organic Compounds (VOCs) and Their Effects on Children and People with Pre-Existing Lung Disease. *International Journal of Environmental Research and Public Health*, *19*(14). https://doi.org/10.3390/ijerph19148752

Meghji, J., Mortimer, K., Agusti, A., Allwood, B. W., Asher, I., Bateman, E. D., Bissell, K., Bolton, C. E., Bush, A., Celli, B., Chiang, C. Y., Cruz, A. A., Dinh-Xuan, A. T., El Sony, A., Fong, K. M., Fujiwara, P. I., Gaga, M., Garcia-Marcos, L., Halpin, D. M. G., … Marks, G. B. (2021). Improving lung health in low-income and middle-income countries: from challenges to solutions. *The Lancet*, *397*(10277), 928–940. https://doi.org/10.1016/S0140-6736(21)00458-X

Menendez, M. E., Janssen, S. J., & Ring, D. (2016). Electronic health record-based triggers to detect adverse events after outpatient orthopaedic surgery. *BMJ Quality and Safety*, *25*(1), 25–30. https://doi.org/10.1136/bmjqs-2015-004332

Mortimer, K., Reddel, H. K., Pitrez, P. M., & Bateman, E. D. (2022). Asthma management in low and middle income countries: case for change. *European Respiratory Journal*, *60*(3), 1–17. https://doi.org/10.1183/13993003.03179-2021

Mosbech, C. H., Godtfredsen, N. S., Ulrik, C. S., & Westergaard, C. G. (2023). Biomarker-guided withdrawal of inhaled corticosteroids in asthma patients with a non-T2 inflammatory phenotype – a randomized controlled trial study protocol. *BMC Pulmonary Medicine*, *23*(1), 372. https://doi.org/10.1186/s12890-023-02679-y

Mudiyanselage, S. B., Stevens, J., Toscano, J., Kotowicz, M. A., Steinfort, C. L., Hayles, R., & Watts, J. J. (2023). Cost-effectiveness of personalised telehealth intervention for chronic disease management: A pilot randomised controlled trial. *PLoS ONE*, *18*(6 June). https://doi.org/10.1371/journal.pone.0286533

Nanthakumaran, P., & Liyanage, L. (2021). Tree-Based Models Using Random Grid Search Optimization for Disease Classification Based on Environmental Factors: A Case Study on Asthma Hospitalizations. *2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning, PRML 2021*, 136–142. https://doi.org/10.1109/PRML52754.2021.9520720

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access*, *9*, 78658–78700. https://doi.org/10.1109/ACCESS.2021.3083060

Nickovic, S., Petković, S., Ilić, L., Pejanović, G., Mijić, Z., Huete, A., & Marks, G. (2023). Prediction of airborne pollen and sub-pollen particles for thunderstorm asthma outbreaks assessment. *Science of the Total Environment*, *864*(December 2022), 1–3. https://doi.org/10.1016/j.scitotenv.2022.160879

Perry, T. T., Marshall, A., Berlinski, A., Rettiganti, M., Brown, R. H., Randle, S. M., Luo, C., & Bian, J. (2017). Smartphone-based vs paper-based asthma action plans for adolescents. *Annals of Allergy, Asthma & Immunology*, *118*(3), 298–303. https://doi.org/10.1016/j.anai.2016.11.028

Pooja, M. R., Alsubaie, N., Alqahtani, M. S., Abbas, M., & Souene, B. O. (2023). *An Optimized Machine Learning Model for the Early Prognosis of Chronic Respiratory Diseases with specic focus On Asthma*. https://doi.org/10.21203/rs.3.rs-3075448/v1

Qi, H., Yang, X., Xu, C., Shi, H., Liu, S., Zhang, X., Wang, T., & Huang, M. (2023). Identification of Prognostic Genes and a Subtype Sensitive to Cetuximab in Colorectal Cancer. *Journal of Biological Regulators and Homeostatic Agents*, *37*(6), 2903–2911. https://doi.org/10.23812/j.biol.regul.homeost.agents.20233706.288

Ramphul, M. (2023). Increased inhaled corticosteroids for treating acute asthma exacerbations. In *Clinical and Experimental Allergy* (Vol. 53, Issue 4, pp. 388–391). John Wiley and Sons Inc. https://doi.org/10.1111/cea.14306

Reel, P. S., Reel, S., van Kralingen, J. C., Langton, K., Lang, K., Erlic, Z., Larsen, C. K., Amar, L., Pamporaki, C., Mulatero, P., Blanchard, A., Kabat, M., Robertson, S., MacKenzie, S. M., Taylor, A. E., Peitzsch, M., Ceccato, F., Scaroni, C., Reincke, M., … Jefferson, E. (2022). Machine learning for classification of hypertension subtypes using multi-omics: A multi-centre, retrospective, data-driven study. *EBioMedicine*, *84*. https://doi.org/10.1016/j.ebiom.2022.104276

Ren, Z., Wan, J., & Deng, P. (2022). Machine-Learning-Driven Digital Twin for Lifecycle Management of Complex Equipment. *IEEE Transactions on Emerging Topics in Computing*, *10*(1), 9–22. https://doi.org/10.1109/TETC.2022.3143346

Rosenfeld, N., & Last, M. (2022). Using ECG signals for hypotensive episode prediction in trauma patients. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE*, *223*. https://doi.org/10.1016/j.cmpb.2022.106955 WE - Science Citation Index Expanded (SCI-EXPANDED)

Saleem, S. S., Khan, A., Aman, R., Saleem, S. S., Bibi, A., Ahmad, N., Mushtaq, S., Mehsud, S., Green, I. R., Rashid, S. A., Khan, A., & Shah, K. U. (2022). Impact of pharmacist-led educational intervention on knowledge of self-management among asthmatic patients: A prospective cohort study. *BMJ Open*, *12*(6). https://doi.org/10.1136/bmjopen-2021-058861

Şanlıtürk, D., & Ayaz-Alkaya, S. (2021). The Effect of a Theory of Planned Behavior Education Program on Asthma Control and Medication Adherence: A Randomized Controlled Trial. *Journal of Allergy and Clinical Immunology: In Practice*, *9*(9), 3371–3379. https://doi.org/10.1016/j.jaip.2021.03.060

Serebrisky, D., & Wiznia, A. (2019). Pediatric asthma: A global epidemic. *Annals of Global Health*, *85*(1), 1–6. https://doi.org/10.5334/aogh.2416

Shaveta. (2023). A review on machine learning. *International Journal of Science and Research Archive*, *9*(1), 281–285. https://doi.org/10.30574/ijsra.2023.9.1.0410

Sheares, B. J., Mellins, R. B., Dimango, E., Serebrisky, D., Zhang, Y., Bye, M. R., Dovey, M. E., Nachman, S., Hutchinson, V., & Evans, D. (2015). Do Patients of Subspecialist Physicians Benefit from Written Asthma Action Plans? *American Journal of Respiratory and Critical Care Medicine*, *191*(12), 1374–1383. https://doi.org/10.1164/rccm.201407-1338OC

Shiri, T., Birungi, J., Garrib, A. V., Kivuyo, S. L., Namakoola, I., Mghamba, J., Musinguzi, J., Kimaro, G., Mutungi, G., Nyirenda, M. J., Okebe, J., Ramaiya, K., Bachmann, M., Sewankambo, N. K., Mfinanga, S., Jaffar, S., & Niessen, L. W. (2021). Patient and health provider costs of integrated HIV, diabetes and hypertension ambulatory health services in low-income settings — an empirical socio-economic cohort study in Tanzania and Uganda. *BMC Medicine*, *19*(1). https://doi.org/10.1186/s12916-021-02094-2

Skolik, A., Jerbi, S., & Dunjko, V. (2022). Quantum agents in the Gym: a variational quantum algorithm for deep Q-learning. *Quantum*, *6*. https://doi.org/10.22331/Q-2022-05-24-720

Soremekun, S., Heaney, L. G., Skinner, D., Bulathsinhala, L., Carter, V., Chaudhry, I., Hosseini, N., Eleangovan, N., Murray, R., Tran, T. N., Emmanuel, B., Garcia Gil, E., Menzies-Gow, A., Peters, M., Lugogo, N., Jones, R., & Price, D. B. (2023). Asthma exacerbations are associated with a decline in lung function: A longitudinal population-based study. *Thorax*, *78*(7), 643–652. https://doi.org/10.1136/thorax-2021-217032

Sowah, R. A., Kuditchar, B., Mills, G. A., Acakpovi, A., Twum, R. A., Buah, G., & Agboyi, R. (2021). HCBST: An Efficient Hybrid Sampling Technique for Class Imbalance Problems. *ACM Transactions on Knowledge Discovery from Data*, *16*(3). https://doi.org/10.1145/3488280

Sun, J., Guo, H., Wang, W., Wang, X., Ding, J., He, K., & Guan, X. (2022). Identifying novel subgroups in heart failure patients with unsupervised machine learning: A scoping review. *Frontiers in Cardiovascular Medicine*, *9*. https://doi.org/10.3389/fcvm.2022.895836

Supriya, M., & Chattu, V. K. (2021). A review of artificial intelligence, big data, and blockchain technology applications in medicine and global health. *Big Data and Cognitive Computing*, *5*(3). https://doi.org/10.3390/bdcc5030041

Tan, Y., Zhang, Z., Li, M., Pan, F., Duan, H., Huang, Z., Deng, H., Yu, Z., Yang, C., Shen, G., Qi, P., Yue, C., Liu, Y., Hong, L., Yu, H., Fan, G., & Tang, Y. (2024). MedChatZH: A tuning LLM for traditional Chinese medicine consultations. *Computers in Biology and Medicine*, *172*. https://doi.org/10.1016/j.compbiomed.2024.108290

Thakur, A., Sharma, P., & Clifton, D. A. (2022). Dynamic Neural Graphs Based Federated Reptile for Semi-Supervised Multi-Tasking in Healthcare Applications. *IEEE Journal of Biomedical and Health Informatics*, *26*(4), 1761–1772. https://doi.org/10.1109/JBHI.2021.3134835

Tibble, H., Tsanas, A., Horne, E., Horne, R., Mizani, M., Simpson, C. R., & Sheikh, A. (2019). Predicting asthma attacks in primary care: Protocol for developing a machine learning-based prediction model. *BMJ Open*, *9*(7). https://doi.org/10.1136/bmjopen-2018-028375

Tiotiu, A. (2021). Applying personalized medicine to adult severe asthma. *Allergy and Asthma Proceedings*, *42*(1), E8–E16. https://doi.org/10.2500/aap.2021.42.200100

Tong, Y., Messinger, A. I., Wilcox, A. B., Mooney, S. D., Davidson, G. H., Suri, P., & Luo, G. (2021). Forecasting future asthma hospital encounters of patients with asthma in an academic health care system: Predictive model development and secondary analysis study. *Journal of Medical Internet Research*, *23*(4). https://doi.org/10.2196/22796

Tsang, K. C. H., Pinnock, H., Wilson, A. M., & Ahmar Shah, S. (2020). Application of Machine Learning to Support Self-Management of Asthma with mHealth. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020-July*, 5673–5677. https://doi.org/10.1109/EMBC44109.2020.9175679

Tyrovolas, S., Stergachis, A., Krish, V. S., Chang, A. R. A. Y., Skirbekk, V., Dieleman, J. L., Chatterji, S., Abd-Allah, F., Abdollahi, M., Abedi, A., Abolhassani, H., Abosetugn, A. E., Abreu, L. G., Abrigo, M. R. M., Abu Haimed, A. K., Adabi, M., Adebayo, O. M., Adedeji, I. A., Adekanmbi, V., … Collaborators, G. 2019 A. (2022). Global, regional, and national burden of diseases and injuries for adults 70 years and older: Systematic analysis for the Global Burden of Disease 2019 Study. *Bmj*, *376*. https://doi.org/10.1136/bmj-2021-068208

Viquerat, J., Duvigneau, R., Meliga, P., Kuhnle, A., & Hachem, E. (2023). Policy-based optimization: single-step policy gradient method seen as an evolution strategy. *Neural Computing and Applications*, *35*(1), 449–467. https://doi.org/10.1007/s00521-022-07779-0

Walsh, J., Moore, Z., Hale, E. Mac, Greene, G., Doyle, F., Costello, R. W., & Murray, B. (2022). The Impact of Biofeedback on Self-Efficacy in Adults with Asthma: A Cross-Sectional Descriptive Survey. *Patient Preference and Adherence*, *16*, 1469–1475. https://doi.org/10.2147/PPA.S358232

Wen, T. Y., & Mohd Aris, S. A. (2022). Hybrid Approach of EEG Stress Level Classification Using K-Means Clustering and Support Vector Machine. *IEEE Access*, *10*, 18370–18379. https://doi.org/10.1109/ACCESS.2022.3148380

Yang, C. Y., Shiranthika, C., Wang, C. Y., Chen, K. W., & Sumathipala, S. (2023). Reinforcement learning strategies in cancer chemotherapy treatments: A review. *Computer Methods and Programs in Biomedicine*, *229*. https://doi.org/10.1016/j.cmpb.2022.107280

Yi, Z., Liu, X. C., Wei, R., & Grubesic, T. H. (2021). Snowplow Truck Performance Assessment and Feature Importance Analysis Using Machine-Learning Techniques. *Journal of Transportation Engineering, Part A: Systems*, *147*(2). https://doi.org/10.1061/jtepbs.0000486

Yip, M., Salcudean, S., Goldberg, K., Althoefer, K., Menciassi, A., Opfermann, J. D., Krieger, A., Swaminathan, K., Walsh, C. J., Huang, H. H., & Lee, I. C. (2023). Artificial intelligence meets medical robotics. *Science (New York, N.Y.)*, *381*(6654), 141–146. https://doi.org/10.1126/science.adj3312

Yoon, C. H., Torrance, R., & Scheinerman, N. (2022). Machine learning in medicine: should the pursuit of enhanced interpretability be abandoned? *Journal of Medical Ethics*, *48*(9), 581–585. https://doi.org/10.1136/medethics-2020-107102

Yu, B., Jiao, J., Li, C., Sun, S., Wang, J., Zhang, C., & Liu, T. (2023). Effect of spinal fusion on degenerative scoliosis in elderly patients and analysis of influencing factors. *American Journal of Translational Research*, *15*(2), 1168–1176. http://www.ncbi.nlm.nih.gov/pubmed/36915737%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC10006747

Zein, J. G., Wu, C. P., Attaway, A. H., Zhang, P., & Nazha, A. (2021). Novel Machine Learning Can Predict Acute Asthma Exacerbation. *Chest*, *159*(5), 1747–1757. https://doi.org/10.1016/j.chest.2020.12.051

Zhang, O., Minku, L. L., & Gonem, S. (2021). Detecting asthma exacerbations using daily home monitoring and machine learning. *Journal of Asthma*, *58*(11), 1518–1527. https://doi.org/10.1080/02770903.2020.1802746