

# Predictive Models for Personalized Asthma Attacks Based on Patient's Biosignals and Environmental Factors: A Systematic Review

Eman T. Alharbi (✉ [ealharbi0125@stu.kau.edu.sa](mailto:ealharbi0125@stu.kau.edu.sa))

King Abdulaziz University

Farrukh Nadeem

King Abdulaziz University

Asma Cherif

King Abdulaziz University

---

## Research Article

**Keywords:** Prediction, machine learning, asthma attack, Biosignals, environmental factor

**Posted Date:** August 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-770597/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Medical Informatics and Decision Making on December 1st, 2021. See the published version at <https://doi.org/10.1186/s12911-021-01704-6>.

RESEARCH

# Predictive Models for Personalized Asthma Attacks based on Patient's Biosignals and Environmental Factors: A Systematic Review

Eman T. Alharbi<sup>1\*</sup>, Farrukh Nadeem<sup>2</sup> and Asma Cherif<sup>3</sup>

\*Correspondence:

ealharbi0125@stu.kau.edu.sa

<sup>1</sup>Department of Information

Systems, King Abdulaziz

University, Jeddah, Saudi Arabia,

Full list of author information is available at the end of the article

## Abstract

**Background:** Asthma is a chronic disease that exacerbates due to various risk factors, including the patient's biosignals and environmental conditions. It is affecting on average 7% of the world population. Preventing an asthma attack is the main challenge for asthma patients, which requires keeping track of any risk factor that can cause a seizure. Many researchers developed asthma attacks prediction models that used various asthma biosignals and environmental factors. These predictive models can help asthmatic patients predict asthma attacks in advance, and thus preventive measures can be taken. This paper introduces a review of these models to evaluate the used methods, model's performance, and determine the need to improve research in this field.

**Method:** A systematic review was conducted for the research articles introducing asthma attack prediction models for children and adults. We searched the PubMed, ScienceDirect, Springer, and IEEE databases from January 2000 to December 2020. The search includes the prediction models that used biosignal, environmental, and both risk factors. The research article's quality was assessed and scored based on two checklists, the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and the Critical Appraisal Skills Programme clinical prediction rule checklist (CASP). The highest scored articles were selected to review.

**Result:** From 1068 research articles we reviewed, we found that most of the studies used asthma biosignal factors only for prediction, few of the studies used environmental factors, and limited studies used both of these factors. Fifteen different asthma attack predictive models were selected for this review. We found that most of the studies used traditional prediction methods, like Support Vector Machine and regression. We have identified the pros and cons of the reviewed asthma attack prediction models and propose solutions to advance the studies in this field.

**Conclusion:** Asthma attack predictive models become more significant when using both patient's biosignal and environmental factors. There is a lack of utilizing advanced machine learning methods, like deep learning techniques. Besides, there is a need to build smart healthcare systems that provide patients with decision-making systems to identify risk and visualize high-risk regions.

**Keywords:** Prediction; machine learning; asthma attack; Biosignals; environmental factor

## Background

Asthma is one of the most common chronic respiratory diseases that significantly impact patients and their families. It affects the lungs and causes the airways to get inflamed and narrow rising breathing difficulties [1]. The World Health Organization (WHO) has reported that it affects on average 7% of the world population, and it is estimated that the rate of asthma deaths between 2015 and 2030 for the Eastern Mediterranean Region will increase [2].

The symptoms of asthma disease are coughing, periods of wheezing, shortness of breath, and chest tightness [3]. These symptoms range from mild to severe from one person to another and may happen frequently or rarely. When symptoms get worse, it is called an asthma attack. Many risk factors lead to an asthma attack; these factors are specified in two categories, patient biosignals and environmental conditions. Many biosignal factors can indicate the possibility of getting an asthma attack, such as lung function, asthma level, smoking status, family history, and clinical data [4]. The environmental factors include weather and air pollution, which are recorded through meteorological stations using multiple sensors. The weather factors that affect asthmatic people include cold, humidity, rain, and wind. In contrast, air pollutions include the small Particulate Matter (PM<sub>2.5</sub> with an aerodynamic diameter greater than 2.5 and PM<sub>10</sub> with an aerodynamic diameter less than ten ) and ozone (O<sub>3</sub>). The particulate matter comes from cars and factories' emissions, smoke, and road dust [5, 6, 7].

Preventing asthma attacks is the main challenge for asthma patients. It requires keeping track of any risk factor that can cause an attack then predict whether special precautions should be taken to prevent the attack. Real-time monitoring and prediction of asthma attacks are the most promising fields of research. They help follow up the appropriate mitigation actions to minimize risks.

Recent studies had connected the existence of asthma disease with different features and built various models to predict its existence and exacerbations. Clinical features, patient biosignals, and environmental data are all correlated with asthma and can be used in building models for asthma attack prediction.

In this paper, we provide a systematic review of these models. We describe and discuss the existing model's properties, e.g., population characteristics, data acquisition, predictors, utilized machine learning techniques, and model performance. By understanding these properties, we will provide recommendations for future research applied in asthma attack prediction, supporting improvements and progress in this domain.

This paper is organized as follows. The related work section discusses some systematic review articles for asthma attack prediction models. In the research strategy, we define the selection criteria for the included articles. In search results, we introduce a detailed explanation for each research article from different aspects. Next, we present a discussion and future research directions. The conclusion is provided in the final section.

## Related Work

A variety of surveys have been introduced to review the existing models for asthma prediction. Most of these studies had focused on the models that predict asthma as a disease for children and adults [8, 9, 10, 11].

The models included in these reviews were mainly based on using the patient biosignals, parent's reports, clinical features, and physiological factors (e.g., the historical health record for the parents, lung functions, and the existence of atopic diseases or allergies). The reviewed models were specially designed for use in primary care diagnostic and clinician's decision-making.

It is worth noting that a few studies have introduced a systematic review of the prediction models designed to predict asthma attack and exacerbation [9, 10, 12, 13]. The selected models in existing reviews were based on analyzing biosignal/environmental risk factors. However, none of these surveys had discussed the predictive models based on the environmental risk factors in addition to the biosignal risk factors.

To overcome this lack, we present this review to study the existing models applied to predict asthma attacks using different risk factors to know the importance of monitoring asthma risk factors, predicting the attack, and visualizing the environmental risks. This review will help define whether efficient predictable models have been developed utilizing multiple predictors and identifying open research questions to determine if additional investigation is required in the field. In addition, this study provides trends in predicting asthma attacks using machine learning, limitations, and future research directions.

## Search Strategy

In this work, the PubMed, ScienceDirect, Springer, and IEEE databases were used to perform the search. The search queries used are: "asthma attack/exacerbation prediction model", "predictive model of asthma attack/exacerbation", "asthma attack prediction using biosignals data", "and asthma attack prediction based on environmental factors". Note that the search results were limited to the prediction model that predicts asthma attacks only, where the models of asthma prediction as a disease were eliminated as they are out of scope.

The title and abstract of each retrieved articles were evaluated to define potential relevancy. The full text was evaluated for each potentially relevant reference to obtain a final inclusion judgment. The studies that meet the pre-defined inclusion guidelines are included in this study. In what follows, we discuss the data extraction and selection criteria considered in our research.

## Data Extraction and Selection Criteria

The research articles were selected based on two main checklists; the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [14] and the Critical Appraisal Skills Programme clinical prediction rule checklist (CASP) [15]. These checklists were designed especially for assessing the aspect of predictive modeling studies. The abstract of the selected research articles was reviewed accordingly with the CHARMS checklist, as shown in Tab. 1. Then the body was assessed according to the CASP checklist, as illustrated in Tab.2.

## Search Results

The search yielded 1827 studies. The excluded studies were nonrelevant, reviews, Non-English references, or present inadequate details of the study for worthwhile

analysis. The remaining studies were 1068, where 897 used biosignals as risk factors in the asthma prediction, 143 used environmental risk factors, and only 28 used both categories. Fig. 1 shows the research article's searching mechanism.

We select 15 state-of-the-art research articles to review and identify various open research questions. The selection of these articles is based on giving a score using the criteria presented previously (CHARMS and CASP). Fig. 2 shows the percentage of the existing articles with their used risk factors and displays the selected references. The included articles were separated into three groups based on the used predictive features. The first research articles group covers the studies that use biosignals data only to build their prediction models [16, 17, 18, 19]. The second research articles group includes studies that use environmental data only to build their prediction models [20, 21, 22, 23, 24]. Finally, the third group includes the studies that used both biosignals and environmental data in their models [25, 26, 27, 28, 29, 30].

Each selected model was reviewed descriptively. In the upcoming subsections, we explained in detail the collected information that covered the population characteristics, data acquisition method, used risk factors, development method, and the model performance reported in the derivation cohorts such as specificity, sensitivity, precision, and accuracy ratios. At the end of this section, we placed Fig. 3 to present the review summary taxonomy that shows the derived information with different attributes.

### Population Characteristics

The designed models were developed to predict asthma attacks for children and adults. However, the validation of these models was tested on adults only, but it can be used for children. The population size varied from 1 [26] to 500,000 [20] participants. However, having a limited number of participants in a prediction study, such as [26, 28], restricts the circumstances of the study and leads to unreliable results. Some studies used a large population size [21, 22, 23] and acquired their data from national databases; therefore, some participants were excluded due to uncompleted records.

### Data Acquisition

The used dataset in the existing prediction models included data from two primary sources, biosignals and environmental. Each of these datasets was acquired by different techniques. This section introduces the used methods in the existing studies for each data source. Tab 3. introduces a comparison to summarize the population size and the data acquisition methods for all the included studies.

#### *Participant's datasets*

Most of the models were built using home-based telemonitoring technology, which was increasingly used recently to monitor and control different diseases, including asthma. This method was used in [16, 17, 25, 28, 30], where the patients record their symptoms daily and submit them through the application interface. This acquisition method relies on the patients' ability to use technology and increases the burden on them. Other studies regarded patients who had initially visit the emergency department [20, 21, 22, 23]. However, research on different groups and populations

is required to generalize the model outcomes and to confirm the study findings. This concern was solved by Tibble et al. [19] through utilizing a historical dataset of the Asthma Learning Healthcare System (ALHS), rising with the most significant sample size of all the covered studies ( $n = 500.000$ ). Besides, Jayawardene et al. [24] collected the patient's data within 49 countries by using Electronic Medical Record (EMR), including the patient's symptoms, physical examination, and lab tests.

Hosseini et al. [26] developed another method by employing a smartwatch application to record the biosignals data through a built-in wireless sensor, which gives a better way to build a personalized prediction model. One particular cause of concern was in the Do et al. [18] study as it relies solely on a randomly generated dataset, which can be highly variable and unreliable.

#### *Environmental datasets*

Most of the studies acquired the environmental data from meteorological agencies and national air pollution monitoring stations [25, 26, 27].

The other method used to acquire these data was by using built-in or networked sensors. Siddiquee et al. [29] proposed a framework based on a network of sensors that can collect different environmental data and analyze them through the Internet of Things (IoT) sensors. On the other hand, Hoq et al. [30] proposed a framework that utilized a small sensor that can be carried by the user and connected to his mobile via Bluetooth to transfer the sensed data. However, using a biosignal sensor to monitor air pollutions is not practical and difficult to deploy. Moreover, these sensors may not be accurate and do not give correct measurements as the agency stations do.

#### **Predictors**

All the predictors used in the prediction models have evidence of being linked to asthma attacks and exacerbations. These predictors are divided into two main categories, biosignal and environmental risk factors. Most researchers used one type of risk factor in the prediction process, ignoring the importance of other factors. Few researchers introduced an asthma attack prediction model using both biosignal and environmental factors. In the following, we discuss each of these categories then summarize our findings in Tab 4.

#### *Predictive Models Developed Based on The patient's Biosignal Risk Factors*

Peak Expiratory Flow Rate (PEFR) and Forced Expiratory Volume in 1 second (FEV1) are the most common biosignal factors used to measure breathing stability. PEFR variable was mainly used in seven out of the fifteen studies [16, 17, 19, 25, 26, 27, 28] to predict asthma attacks. Other biosignal factors were combined with PEFR to explore the utility of using additional attributes to increase the reliability of the prediction models.

Finkelstein and Jeong [16, 17] apply respiratory infections, the presence of cold, physical activity, sleep disturbances, and the used medications. Asthma severity, smoking Status, blood eosinophil count, obesity, comorbidity, previous medication usage, and history of asthma attacks were applied beside PEFR in the prediction

model introduced by Tibble et al. [19]. Besides all the previous factors, Lee et al. [25] applied extra attributes, including food allergy, dermatitis, rhinitis, and conjunctivitis.

On the other hand, FEV1 was used in two studies only. Hosseini et al. [26] introduced their prediction model based on FEV1 and PEFr without considering any other biosignal factors. In contrast, Do et al. [18] used FEV1 with lung functions, night awaking, used medication, activities, and symptoms.

According to these studies, using additional biosignal asthma risk factors significantly improve model performance. In Lee et al. [25], who utilized the highest number of risk factors, the model achieved a high accuracy of 0.87 and a high recall of 0.85. In comparison, the model built by Hosseini et al. [26], which utilized two factors only, achieved less accuracy of 0.80. Thus, the model performed better when using multiple attributes. The accuracy of the other models [16, 17, 27] falls between these two percentages. One exception is Kaffash-Charndabi et al. [28] predictive model, which used only the PEFr as a biosignal risk factor and got an accuracy of 0.93. However, this result comes out due to testing the model on three participants only, which makes it unreliable regarding the performance of the other models, which were trained and tested with a large dataset.

#### *Predictive Models Developed Based on The Environmental Risk Factors*

Asthma is highly associated with environmental factors. A large body of research articles were promoting the correlation between disclosure to weather and air pollution with asthma attacks [6, 7, 31, 32]. Various models were introduced to predict asthma attacks based on analyzing weather conditions and air pollution concentration. These studies' prediction target was the number of asthma patients' visits to the emergency department (ED). The most common weather and air pollution factors used through studies [20, 21, 22, 23] were temperature (T), humidity (H), nitrogen dioxide (NO<sub>2</sub>), sulfur dioxide (SO<sub>2</sub>), ozone (O<sub>3</sub>), carbon monoxide (CO), particulate matter (PM<sub>10</sub>), and (PM<sub>2.5</sub>).

Ram et al. [20] used Twitter data and Google search interests besides these environmental factors for asthma attack prediction. Using such datasets provides a better opportunity for early prediction of asthma risk and can prevent the attack. However, the model performance gets lower when using these data, with an accuracy of 0.70 and 0.89 without these data. Barometric pressure (BP) was one of the weather factors tested by Mireku et al. [21] to determine its effect on increasing asthma symptoms besides all the previous variables. The model shows that BP has no impact on asthma patients and did not change the prediction result.

Delamater et al. [22] had tested each of the previously mentioned factors and found that the increase in CO, NO<sub>2</sub>, and PM<sub>2.5</sub> were significantly impacted the rate of asthma hospitalization with a positive coefficient correlation of 0.79, 0.79, 0.93, respectively. In contrast, T and H produced mixed results, and O<sub>3</sub> was non-significant with a relationship of 0.03. Vegetation density (VD) was applied as a new predictor in the model introduced by Ayres-Sampaio et al. [23]. The outcomes of the study show that VD has consistent relationships with asthma with 0.6 Pearson correlation coefficients. The result illustrates that people with asthma are at high risk of getting asthma attacks in the urbanized area and sparsely vegetated areas, leading to hospital admissions.

The result of these studies leads other researchers to the most harmful air pollution factors to indicate them in future studies. However, the studies lack in using individual data and variations in the demographic structure of the population, which limits the ability to establish a causal link between the change in environmental factors and asthma outcomes.

#### *Predictive Models Developed Based on The Biosignal and Environmental Risk Factors*

Predicting the asthma attack becomes more accurate and gives better performance when using both of the risk factors, the biosignal and environmental. Lee et al. [25] introduced a study that shows how to contribute to the model performance by applying multiple predictors from different sources. They used 37 attributes; 19 for the patient biosignals, 14 for air pollution, and 4 for the weather. The performance of the model shows the ability to predict asthma attacks with 0.87 accuracy.

Other studies [26, 27] considered limited biosignal attributes and focused on applying more environmental data, including air quality index (AQI), weather data, and traffic. Their model's performance was under 0.80 accuracy, which could be improved by applying additional biosignals data. Kaffash-Charandabi et al. [28] developed another study based on using PEFr and the patient's medical history. The limitation of this study was the limited number of its participants, as they were only three participants who record their data in different situations.

Two studies built a prediction framework with a new interpretation of the situation of concern in this review. Siddiquee et al. [29] and Hoq et al. [30] proposed a smart framework to predict asthma attacks by deploying wireless sensors to acquire biosignals and environmental data. However, the proposed structures were limit these data to a few risk factors such as food allergy, humidity, CO, and NO<sub>2</sub>, which is not clear how the model would perform with these factors in the real world.

#### **Development Method and Model Performance**

There were four unique model development methodologies specified in the reviewed research articles: Probabilistic Approach, Classification and Regression Tree Analysis (CART), Regression Analysis, and Real-Time Analysis. Tab.5 presents details for the used methods with the performance of each model.

#### *Probabilistic Approach*

The probabilistic approach is based on the probability theory. It consolidates probability distributions and random variables into a model of phenomenon or events and gives a probability distribution as a solution [33]. A probabilistic approach was used in four of the analyzed research articles to build the prediction model [16, 18, 24, 25].

Finkelstein and Jeong [16] and Lee et al. [25] built a model to assign a high or low risk of asthma exacerbation using machine learning. They applied approximately 70% of the data for model development and 30% for verification. The models were evaluated using different performance matrices. These matrices were introduced for the model as a unit but not to particular values for development and verification. Finkelstein and Jeong [16] used eight-day lag inputs from a telemonitoring system as predictors. A naïve Bayesian classifier (NB), adaptive Bayesian network (ABN),

and a support vector machine (SVM) were used. This study found that the ABN classifier has higher sensitivity and specificity (1.00, 1.00) than SVM (sensitivity 0.84 and specificity 0.80), and NB classifier (sensitivity 0.80, specificity 0.77). Do et al. [18] built a model to classify the asthma attack into three categories. They trained their model with Bag of Words and tested it with a random text using Tensorflow Text Classification (TC) result in an accuracy of 0.98.

Jayawardene et al. [24] used a generalized estimating Equation (GEE) to predict asthma attacks. They used a cut-off value estimated by a 3-year mean value of seizures per day in the period between 2008 -2010. The highest asthma attacks are found in the fall and summer, and the lowest are found in spring and winter.

Lee et al. [25] used a five-day lag to train the classifier with the dataset acquired from EMR and environmental sensors. Pattern-Based Decision Tree (PBDT) and Pattern-Based Class-Association Rule (PBCAR) were applied and compared. The experimental results show that PBDT outperformed PBCAR with a slight difference, which delivers an accuracy of 0.87 and 0.86, respectively.

#### *Classification and Regression Tree Analysis (CART)*

CART analysis is unlike traditional data analysis methods. It is a tree-building technique that is effective for creating clinical decision rules [34]. In addition, CART can reveal complicated interactions between predictors that may be difficult to discover using traditional multivariate techniques.

Finkelstein and Jeong [17] developed an asthma exacerbation prediction model one day in advance based on the previous 7-day lag by using the CART algorithm. Telemonitoring data was used in the development and verification of the model. Three evaluation models were conducted, and each model predicts whether the asthma is "Normal" or "Exacerbation" for the next day. The resulting algorithm had a specificity of 0.97, a sensitivity of 0.64, and an accuracy of 0.80.

#### *Regression Analysis*

Regression is a statistical measure used to determine the relationship between independent and dependent parameters [35].

Different categories of regressing were used in three of the analyzed research articles [21, 22, 23]. These studies introduced different models to determine the probability of increasing asthma ED visits according to weather or air pollution concentration changes. Mireku et al. [21] used Bayesian regression and found that with two to three days lag, the increase in humidity and temperature cause a rise with 1.8 additional asthma ED visits.

Delamater et al. [22] built multiple models to determine the relationship between environmental variables and asthma ED visits. Bayesian regression was used with temporal random effects, which improved model fit and produced better accurate and precise predictions.

Ayres-Sampaio et al. [23] applied Land-Use regression (LUR) results with a consistent relationship between T, NO<sub>2</sub>, and VD with asthma in all seasons. T and NO<sub>2</sub> were positively associated with the admissions where Pearson correlation coefficients ranged from 0.351 to 0.600, 0.405 to 0.513, respectively. VD was negatively related to the ED admission where the coefficients ranged from 0.376 to 0.498. Also, H had a high relationship where the ratio ranged from 0.54 to 0.81.

### *Real-Time Computing*

Real-time computing is based on recognizing and performing different functions within a predictable time and storing a large amount of data acquired from the controlled system [36].

There were four works [20, 26, 27, 28] that used real-time computing to predict asthma attacks per individual. Different wireless sensors were employed to acquire data from users and the environment. Hosseini et al. [26] introduced a model to classify the risk of getting an asthma attack into three different classes (high, medium, and low risk) using a random forest classifier with 10-fold cross-validation. The biosignals data were recorded through a built-in smartwatch wireless sensor, while the environmental data were acquired from different meteorological wireless sensors. Data were analyzed in real-time through the cloud platform. The prediction model could classify the risk level with an accuracy of 0.80.

Kaffash-Charandabi et al. [28] introduced another real-time prediction model based on multiple reasoning mechanisms for a more reliable model. Rule-based inference with SVM was used to predict asthma attacks for each individual and performed an average accuracy of 0.93. The model was trained and tested on three participants only, and it is not clear how the model would perform when tested on more populations.

Ram et al. [20] developed a real-time prediction model for asthma ED visits based on Twitter, Google search data, and environmental variables. They examined different techniques to validate the model include Decision Tree (DT), Artificial Neural Network (ANN), and a combination of both (DT+ANN). The result of using ANN alone was the same as using both techniques and its surpassed DT technique, resulting in an accuracy of 0.70 and 0.67, respectively, which considered low and could be enhanced by using different machine learning techniques or more specified dataset.

Another comparison was performed by Khasha et al. [27] between six different methods. Also, DT had better performance and surpassed LR, RF, SVM, and gradient boosting (GB). The study used different measurements like sensitivity, specificity, F-measure, G-means, accuracy, and precision. A weakness of this study was that it fails to represent the output of these measurements with percentages or numbers, and it used only figures to describe the outcomes. Two more articles [29, 30] also proposed a prediction model based on real-time analysis. The proposed frameworks are built based on activating multiple wireless sensors to acquire real-time data from the user and his surrounded environment. However, using a biosignal sensor to monitor air pollutions is not practical and challenging to employ.

## **Discussion and Future Research Direction**

There are numerous limitations to the existing predictive models for asthma attacks (see Tab. 6). To overcome this, we define some recommendations and propose different solutions to advance the studies in this field.

### **Use of multiple risk factors**

Most existing asthma attack prediction models use limited risk factors. Large numbers of research articles utilize only biosignals data and ignore the environmental

factors. Also, using the telemonitoring system in collecting biosignals data is not an effective method since data is recorded manually by the patients themselves.

In practice, Electronic Medical Record (EMR) represents a better data source as it is much more consistent and robust than telemonitoring data. Also, EMR often includes additional variables observed by the physician. Combining EMR and the wireless sensor is the ultimate method that provides the best historical data and precise biosignals measurements for more accurate prediction results. It is essential to devise new prediction systems that intelligently take the full potential of EMR to collect and update patient data.

### Introducing Personalized Prediction and Prevention Models

Most existing asthma attack prediction models were established for clinical use to facilitate the disease diagnosis and decision-making for the physician. Unfortunately, this has led to design models that are unsuitable for personal usage. Indeed, asthma patients need to follow up on all the asthma symptoms and risk factors to avoid sudden attacks. So far, most current models make predictions to identify the ED visits or find the relationship between the attack and different risk factors, ignoring the need for personalized asthma attack prediction and the required precautions that should be followed.

To improve asthma patients' lives, a predictive model for asthma attacks should notify the user of any undesirable situation as soon as a risk factor threatens the patient life without any delay that could worsen the health condition. Ideally, the model should alarm the user with a real-time update for air pollution concentration and climate changes as additional services.

### Improving Prediction Accuracy

To provide a valuable asthma attack predictive model, it is crucial to provide high sensitivity and accuracy. Indeed, high sensitivity guarantees the model's ability to classify most patients who will get an asthma attack according to the surrounding risk factors. In contrast, high accuracy ensures that a patient with high predicted risk is undoubtedly expected to have an asthma attack.

As reviewed in this paper, most of the predictive models did not exceed an accuracy of 0.87, with two exceptions that did not use a suitable dataset. Besides, sensitivity was not reported for most of the models. It remains an open problem to enhance the accuracy of asthma attack prediction models without affecting the computation time. To the best of our knowledge, currently, no such model can achieve efficiency high enough for individual usage. Using large datasets and exhaustive variable sets with superior machine learning methods can improve the model's accuracy.

With rare exceptions [19, 24] existing asthma attack prediction models were introduced utilizing small datasets, including (typically much) less than 350 patients. Generally, the accuracy of the prediction model increases as the training dataset becomes more significant, mainly if the model utilizes multiple predictors. There are a variety of asthma attack risk factors identified through different studies and researches [3, 5, 7, 37]. However, with one exception [25], most existing asthma attack prediction models used less than ten variables. By utilizing a comprehensive set of predictors linked with a large number of populations, the model's accuracy will further likely improve.

## Conclusion

Establishing a self-management tool for asthma control and attack prediction is still an attractive area for researchers. By reviewing the studies and research in asthma monitoring and prediction models, it can be concluded that most of these studies focused on using specific factors and neglected other essential factors. Asthma prediction requires the use of various risk factors, including biosignals and environmental conditions. Numerous studies introduce a linkage between particular variables and asthma exacerbation, and such articles can be used in the predictor selection process. Moreover, most of the current prediction models do not deliver services to patients. Subsequently, there is a limitation in developing a smart monitoring system of environmental variables that can provide a personalized map with the concentration of surrounding air pollution. Constructing a smart healthcare system to predict asthma attacks for asthma patients and visualize environmental risks is critically important. There is a need to consider all the categories of risk factors with various variables as much as possible to deliver an accurate model. It is required to provide patients with a decision-making system that alarms them by identifying the risk factor and the high-risk regions. Besides, a dynamic visualizing map for air quality is required to enable citizens and policy-makers to take immediate action and avoid polluted sites.

### Declarations

#### Ethics approval and consent to participate

Not applicable

#### Consent for publication

Not applicable

#### Availability of data and material

Not applicable

#### Competing interests

The authors declare that they have no conflicts of interest to report regarding the present study.

#### Funding

The authors received no specific funding for this study.

#### Authors' contributions

Eman Alharbi\* Searched for articles on the databases and retrieved many articles. She decided on the inclusion criteria and reviewed the retrieved articles based on the target checklists. She wrote the manuscript and draw the figures and construct the tables.

Farrukh Nadeem and Asma Cherif reassessed the selected articles, reviewed the manuscript, and edited the figures and tables.

#### Acknowledgements

Not applicable

#### Abbreviations

World Health Organization (WHO)

Emergency Department (ED)

Electronic Medical Record (EMR)

Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS)

Critical Appraisal Skills Programme clinical prediction rule checklist (CASP)

Peak Expiratory Flow Rate (PEFR)

Forced Expiratory Volume in 1 second (FEV1)

Air Quality Index (AQI)

Classification and Regression Tree Analysis (CART)

Gradient Boosting (GB)

Support Vector Machine (SVM)

Logistic Regression (LR)

Artificial Neural Network (ANN)

Naïve Bayesian (NB)

Adaptive Bayesian Network (ABN)  
 Pattern-Based Decision Tree (PBDT)  
 Pattern-Based Class-Association Rule (PBCAR)  
 Receiver operator characteristic (ROC)  
 Area Under Curve (AUC)

#### Author details

<sup>1</sup>Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia., <sup>2</sup>Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia., <sup>3</sup>Department of Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia.,

#### References

- Melinda, D.: Asthma. <https://www.webmd.com/asthma/what-is-asthma>
- Asthma: Health Topics. <http://www.emro.who.int/health-topics/asthma/index.html>
- Asthma. <https://www.nlm.nih.gov/health-topics/asthma>
- Klaassen, E.M.M., van de Kant, K.D.G., Jöbbs, Q., van Schayck, O.C.P., Smolinska, A., Dallinga, J.W., van Schooten, F.J., den Hartog, G.J.M., de Jongste, J.C., Rijkers, G.T., Dompeling, E.: Exhaled biomarkers and gene expression at preschool age improve asthma prediction at 6 years of age. **191**(2), 201–207. doi:10.1164/rccm.201408-1537OC. Place: United States
- Asthma Triggers and Management. <https://www.aaaai.org/conditions-and-treatments/library/asthma-library/asthma-triggers-and-management>
- Asthma - Causes. <https://www.nhs.uk/conditions/asthma/causes/>
- Weather as an asthma trigger. <https://www.asthma.org.uk/advice/triggers/weather/>
- Daines, L., McLean, S., Buelo, A., Lewis, S., Sheikh, A., Pinnock, H.: Systematic review of clinical prediction models to support the diagnosis of asthma in primary care **29**(1), 19. doi:10.1038/s41533-019-0132-z
- Loymans, R.J.B., Debray, T.P.A., Honkoop, P.J., Termeer, E.H., Snoeck-Stroband, J.B., Schermer, T.R.J., Assendelft, W.J.J., Timp, M., Chung, K.F., Sousa, A.R., Sont, J.K., Sterk, P.J., Reddel, H.K., Ter Riet, G.: Exacerbations in adults with asthma: A systematic review and external validation of prediction models. **6**(6), 1942–195215. doi:10.1016/j.jaip.2018.02.004. Place: United States
- Luo, G., Nkoy, F.L., Stone, B.L., Schmick, D., Johnson, M.D.: A systematic review of predictive models for asthma development in children **15**, 99–99. doi:10.1186/s12911-015-0224-9. Publisher: BioMed Central. 26615519
- Smit, H.A., Pinart, M., Antó, J.M., Keil, T., Bousquet, J., Carlsen, K.H., Moons, K.G.M., Hooft, L., Carlsen, K.C.L.: Childhood asthma prediction models: a systematic review. **3**(12), 973–984. doi:10.1016/S2213-2600(15)00428-2. Place: England
- Bridge, J., Blakey, J.D., Bonnett, L.J.: A systematic review of methodology used in the development of prediction models for future asthma exacerbation **20**(1), 22. doi:10.1186/s12874-020-0913-7
- Fleming, L.: Asthma exacerbation prediction: recent insights. **18**(2), 117–123. doi:10.1097/ACI.0000000000000428. Place: United States
- Moons, K.G.M., de Groot, J.A.H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D.G., Reitsma, J.B., Collins, G.S.: Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. **11**(10), 1001744. doi:10.1371/journal.pmed.1001744
- CASP - Critical Appraisal Skills Programme. <https://casp-uk.net/casp-tools-checklists/>
- Finkelstein, J., Jeong, I.C.: Machine learning approaches to personalize early prediction of asthma exacerbations. **1387**(1), 153–165. doi:10.1111/nyas.13218
- Finkelstein, J., Jeong, I.C.: Using cart for advanced prediction of asthma attacks based on telemonitoring data. In: 2016 IEEE 7th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), pp. 1–5 (2016). doi:10.1109/UEMCON.2016.7777890
- Do, Q.T., Doig, A.K., Son, T.C., Chaudri, J.M.: Personalized prediction of asthma severity and asthma attack for a personalized treatment regimen. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1–5 (2018). doi:10.1109/EMBC.2018.8513281
- Tibble, H., Tsanas, A., Horne, E., Horne, R., Mizani, M., Simpson, C.R., Sheikh, A.: Predicting asthma attacks in primary care: protocol for developing a machine learning-based prediction model **9**(7). doi:10.1136/bmjopen-2018-028375. Publisher: British Medical Journal Publishing Group .eprint: <https://bmjopen.bmj.com/content/9/7/e028375.full.pdf>
- Ram, S., Zhang, W., Williams, M., Pengetnze, Y.: Predicting asthma-related emergency department visits using big data. **19**(4), 1216–1223. doi:10.1109/JBHI.2015.2404829. Place: United States
- Mireku, N., Wang, Y., Ager, J., Reddy, R.C., Baptist, A.P.: Changes in weather and the effects on pediatric asthma exacerbations. **103**(3), 220–224. doi:10.1016/S1081-1206(10)60185-8. Place: United States
- Delamater, P.L., Finley, A.O., Banerjee, S.: An analysis of asthma hospitalizations, air pollution, and weather conditions in los angeles county, california **425**, 110–118. doi:10.1016/j.scitotenv.2012.02.015. Edition: 2012/04/02. 22475217
- Ayres-Sampaio, D., Teodoro, A.C., Sillero, N., Santos, C., Fonseca, J., Freitas, A.: An investigation of the environmental determinants of asthma hospitalizations: An applied spatial approach **47**, 10–19. doi:10.1016/j.apgeog.2013.11.011
- Jayawardene, W.P., Youssefagha, A.H., Lohrmann, D.K., El Afandi, G.S.: Prediction of asthma exacerbations among children through integrating air pollution, upper atmosphere, and school health surveillances. **34**(1), 1–8. doi:10.2500/aap.2013.34.3629. Place: United States
- Lee, C.-H., Chen, J.C.-Y., Tseng, V.S.: A novel data mining mechanism considering bio-signal and environmental data with applications on asthma monitoring **101**(1), 44–61. doi:10.1016/j.cmpb.2010.04.016

26. Hosseini, A., Buonocore, C.M., Hashemzadeh, S., Hojaiji, H., Kalantarian, H., Sideris, C., Bui, A.A.T., King, C.E., Sarrafzadeh, M.: HIPAA compliant wireless sensing smartwatch application for the self-management of pediatric asthma **2016**, 49–54. doi:10.1109/BSN.2016.7516231. Edition: 2016/07/21. 29354688
27. Khasha, R., Sepehri, M.M., Mahdavi, S.A., Khatibi, T.: Mobile GIS-based monitoring asthma attacks based on environmental factors **179**, 417–428. doi:10.1016/j.jclepro.2018.01.046
28. Kaffash-Charandabi, N., Alesheikh, A.A., Sharif, M.: A ubiquitous asthma monitoring framework based on ambient air pollutants and individuals' contexts **26**(8), 7525–7539. doi:10.1007/s11356-019-04185-3
29. Siddiquee, J., Roy, A., Datta, A., Sarkar, P., Saha, S., Biswas, S.S.: Smart asthma attack prediction system using internet of things. In: 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 1–4 (2016). doi:10.1109/IEMCON.2016.7746252
30. Hoq, M.N., Alam, R., Amin, A.: Prediction of possible asthma attack from air pollutants: Towards a high density air pollution map for smart cities to improve living. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–5 (2019). doi:10.1109/ECACE.2019.8679335
31. How Weather Affects Asthma. <https://www.everydayhealth.com/asthma/how-weather-affects-asthma.aspx>
32. Monitoring Ambient Air Quality for Health Impact Assessment
33. Alon, N., Spencer, J.H.: The Probabilistic Method, 4th edition edn. Wiley
34. Lewis, R.J.: An introduction to classification and regression tree (cart) analysis. (2000)
35. V, N.: Introduction to Linear Regression Analysis, 5th ed.[Douglas.C.Montgomery,Elizabeth.A.Peck,and G.]pdf. [https://www.academia.edu/32079757/Introduction\\_to\\_Linear\\_Regression\\_Analysis\\_5th\\_ed\\_Douglas\\_C\\_Montgomery\\_Elizabeth\\_A\\_Peck\\_and\\_G\\_Pdf\\_Accessible](https://www.academia.edu/32079757/Introduction_to_Linear_Regression_Analysis_5th_ed_Douglas_C_Montgomery_Elizabeth_A_Peck_and_G_Pdf_Accessible) 07 – 26
36. Furht, B., Grostick, D., Gluch, D., Rabbat, G., Parker, J., McRoberts, M.: Real-Time UNIX® Systems: Design and Application Guide, 1991st edition edn. Springer
37. Wandalsen, G.F., Solé, D., Bacharier, L.B.: Identification of infants and preschool children at risk for asthma: predictive scores and biomarkers. **16**(2), 120–126. doi:10.1097/ACI.0000000000000240. Place: United States

**Figures**

**Figure 1** Research Article's Searching Mechanism.

**Figure 2** Existing Asthma Attack Prediction Models with the Used Factors Percentage.

**Figure 3** Taxonomy of the Review Output

**Tables**

**Table 1** CHARMS Checklist for Assessing the Abstracts of Selected Research Articles [15]

	Abstract selection criteria	Criteria description
1	“Type of prediction model”	A predictive model to predict the expected event of Asthma attack
2	“Intended scope”	To monitor asthma patients, make decisions on the need for therapy, and help prevent asthma attacks
3	“Type of prediction modeling studies”	Model development researches in which a previously generated model was improved
4	“Target population”	All age groups
5	“Outcome to be predicted”	Asthma attack or exacerbation, asthma risk level, the cause of the asthma attack
6	“Timespan”	Predictors measured between 1 -2 weeks / j 1 month; outcomes measured directly after data acquisition
7	“Intended moment of using the model”	Preventing asthma attacks and recognizing its exacerbation

**Table 2** CASP Checklist for Assessing the Body of the Selected Research Articles [16]

CASP questions	
Q1	Is the predictive model clearly defined?
Q2	Did the population from which the predictive model was derived include an appropriate spectrum of patients?
Q3	Was the predictive model validated in a different group of patients?
Q4	Were the predictor variables and the outcome evaluated in a blinded fashion?
Q5	Were the predictor variables and the outcome evaluated in the whole sample selected initially?
Q6	Are the statistical methods used to construct and validate the predictive model clearly described?

**Table 3** Comparison of The Reviewed Articles Regarding The Population Size and Data Acquisition Methods

Ref	Population size	Data Acquisition method
[16]	7001 records, 350 patients	Telemonitoring data
[17]	3470 records, 174 patients	Telemonitoring data
[18]	Random generated dataset with size 130,000 text	Randomly generated text
[19]	Asthma Learning Healthcare System (ALHS) dataset includes 500,000 individuals	Historical data
[20]	50 ED visits, 3768 tweets	Twitter data, Google search interests, and ED data from hospital
[21]	25,401	ED visits Historical data+ Health records
[22]	100,000	ED visits
[23]	18,409	ED visits
[24]	168,825	EMR
[25]	33 participants	Historical data+ MS
[26]	1 participant	Physiological and environmental wireless sensors
[27]	32 participants	Historical data+ MS
[28]	3 participants in 80 different situations	Historical data+ MS
[29]	-	IoT Sensors
[30]	-	Personal sensor

**Table 4** Comparison of The Reviewed Articles Regarding the Used Risk Factors

Ref.	Risk Factors		
	Personal	Weather	Air Pollution
[16, 17]	PEFR, respiratory symptoms, presence of cold, sleep disturbances, medication, and physical activity	-	-
[18]	FEV1, Lung functions, night awaking, used medication, activities, and symptoms	-	-
[19]	PEFR, Asthma severity, history of asthma attacks, medication, obesity, blood eosinophil count, smoking status, and comorbidity	-	-
[20]	Number of ED visits	-	PM2.5, O3, CO, and NO2.
[21]	Number of ED visits	Humidity, temperature, and barometric pressure	-
[22]	Number of ED visits	Humidity and temperature	CO, O3, PM10, PM2.5, NO2
[23]	Number of ED visits	Temperature and humidity	Vegetation density and NO2
[24]	Demographic data, number of asthmatic children in the selected schools, and the used medications.	Temperature	PM2.5, PM10, SO2, NO2, CO, and O3
[25]	PEFR, Nose symptom, eye symptom, skin symptom, night symptom, day symptom, fever, dermatitis, rhinitis, asthma medicine instructions, conjunctivitis.	Temperature, absolute maximum and minimum temperature, and humidity.	CO, SO2, O3, NO2, PH PM10, HydraCarbon,HydraCarbon2.
[26]	PEFR and FEV 1.	Temperature, humidity, pressure, ozone, wind speed, visibility, and precipitation intensity.	PM2.5, PM10, CO, NO2, and SO2.
[27]	PEFR and patients' location.	Humidity, absolute maximum and minimum temperatures, average temperature, pressure, rainfall, and wind.	O3, SO2, NO2, CO, PM10
[28]	PEFR, location, and medical history.	Temperature, barometric pressure, and humidity.	CO, O3, NO2, SO2, PM10
[29]	Food allergen	Humidity	Pollen
[30]	Medication plan	-	CO, NO, dust and smoke

**Table 5** Comparison of The Reviewed Articles Regarding the Used Development Method and Model Performance (the used symbols SN: sensitivity, SP: specificity, AC: accuracy)

Ref.	Development Method/ Technique	Machine Learning	Model Performance
[16]	NB, ABN, and SVM		SN: 80%, 100%, 84%; SP:77%, 100%, 80%; AC: 77%, 100%, 80%
[17]	CART		SN: 64%, SP: 97%, AC: 80%
[18]	TC		AC: 97%
[19]	NB, SVM, and random forests.		-
[20]	DT, ANN, and DT+ANN		AC: 65.18%, 66.25%, 66.25% respectively.
[21]	Bayesian regression		Increase in T -¿1.8 ED visit Increase in H -¿1 ED visit
[22]	LR		Coefficient correlation of 0.79 (CO), 0.79 (NO <sub>2</sub> ), 0.93 (PM <sub>2.5</sub> ), and 0.03 (O <sub>3</sub> )
[23]	LR		Pearson correlation coefficient:0.60
[24]	GEE		Probability 50% of an asthma attack with low T and high AP
[25]	PBCAR and PBDT		AC: 86.89% and 87.52%, Recall: 84.12%, and 85.59% respectively.
[26]	Random forest		AC: 80%.
[27]	LR, DT, ANN, SVM, gradient boosting, and random forests		AC, SN, SP, G-mean, precision, and F- measure
[28]	SVM		AC: 93.55%.

**Table 6** Comparison of the The Reviewed Articles Regarding their Pros and Cons

Ref.	Pros	Cons
[16, 17]	<ul style="list-style-type: none"> <li>• Number of participants is adequate.</li> <li>• The performance of the model is provided with different measurements.</li> </ul>	<ul style="list-style-type: none"> <li>• Compared different machine learning techniques</li> <li>• The used data are not enough to predict asthma attacks.</li> <li>• Population is not generalized.</li> <li>• Environmental risk factors were not considered.</li> </ul>
[18]	<ul style="list-style-type: none"> <li>• Number of datasets is adequate.</li> <li>• Model accuracy is high.</li> </ul>	<ul style="list-style-type: none"> <li>• Some personal risk factors were ignored.</li> <li>• Environmental risk factors were not considered.</li> </ul>
[19]	<ul style="list-style-type: none"> <li>• Used exhaustive dataset.</li> <li>• Compared different machine learning techniques.</li> </ul>	<ul style="list-style-type: none"> <li>• The study proposed only a protocol without the results of the prediction process.</li> </ul>
[20]	<ul style="list-style-type: none"> <li>• Applied real-time dataset.</li> <li>• Compared different machine learning techniques</li> </ul>	<ul style="list-style-type: none"> <li>• The tweets were taken from the English language only.</li> <li>• The ED visits data were taken from one hospital only.</li> <li>• The accuracy of the model is low.</li> </ul>
[21]	<ul style="list-style-type: none"> <li>• Used sufficient dataset.</li> </ul>	<ul style="list-style-type: none"> <li>• Air pollution was not considered.</li> </ul>
[22]	<ul style="list-style-type: none"> <li>• Dataset has many records.</li> </ul>	<ul style="list-style-type: none"> <li>• Data was lack of patient demographic and residential location information.</li> </ul>
[23]	<ul style="list-style-type: none"> <li>• Used sufficient dataset.</li> </ul>	<ul style="list-style-type: none"> <li>• Air pollution factors were limited.</li> </ul>
[24]	<ul style="list-style-type: none"> <li>• Dataset has many records.</li> </ul>	<ul style="list-style-type: none"> <li>• Weather factors were limited</li> </ul>
[25]	<ul style="list-style-type: none"> <li>• Used exhaustive personal and environmental predictors.</li> <li>• Compared two different machine learning techniques.</li> <li>• Increased the model performance by applying a feature selection algorithm.</li> </ul>	<ul style="list-style-type: none"> <li>• Bio-signals were daily recorded by users.</li> <li>• The interpretation of the study is difficult for users.</li> <li>• No services after prediction.</li> </ul>
[26]	<ul style="list-style-type: none"> <li>• Used exhaustive environmental factors.</li> </ul>	<ul style="list-style-type: none"> <li>• The prediction result is only three states, without any suggestions for treatment or precautions.</li> </ul>
[27]	<ul style="list-style-type: none"> <li>• Used exhaustive environmental factors.</li> <li>• Compared different machine learning techniques.</li> <li>• The performance of the model is provided with different measurements.</li> </ul>	<ul style="list-style-type: none"> <li>• Patient medical history and bio-signals were not used.</li> <li>• Validation results were not expressed by numbers.</li> </ul>
[28]	<ul style="list-style-type: none"> <li>• Model accuracy is high.</li> <li>• Map of polluted and safe areas was introduced.</li> </ul>	<ul style="list-style-type: none"> <li>• Study was performed on three participants only.</li> </ul>
[29]	<ul style="list-style-type: none"> <li>• A complete framework was introduced to predict attacks, alarm used, and view polluted sites.</li> </ul>	<ul style="list-style-type: none"> <li>• The model is based on environmental data only.</li> <li>• It is a proposed framework only.</li> </ul>
[30]	<ul style="list-style-type: none"> <li>• The proposed framework considering real-time prediction and warning users with risks.</li> </ul>	<ul style="list-style-type: none"> <li>• Weather factors were not considered.</li> <li>• Personal sensor does not give an accurate reading.</li> <li>• Difficult to employ.</li> </ul>

# Figures

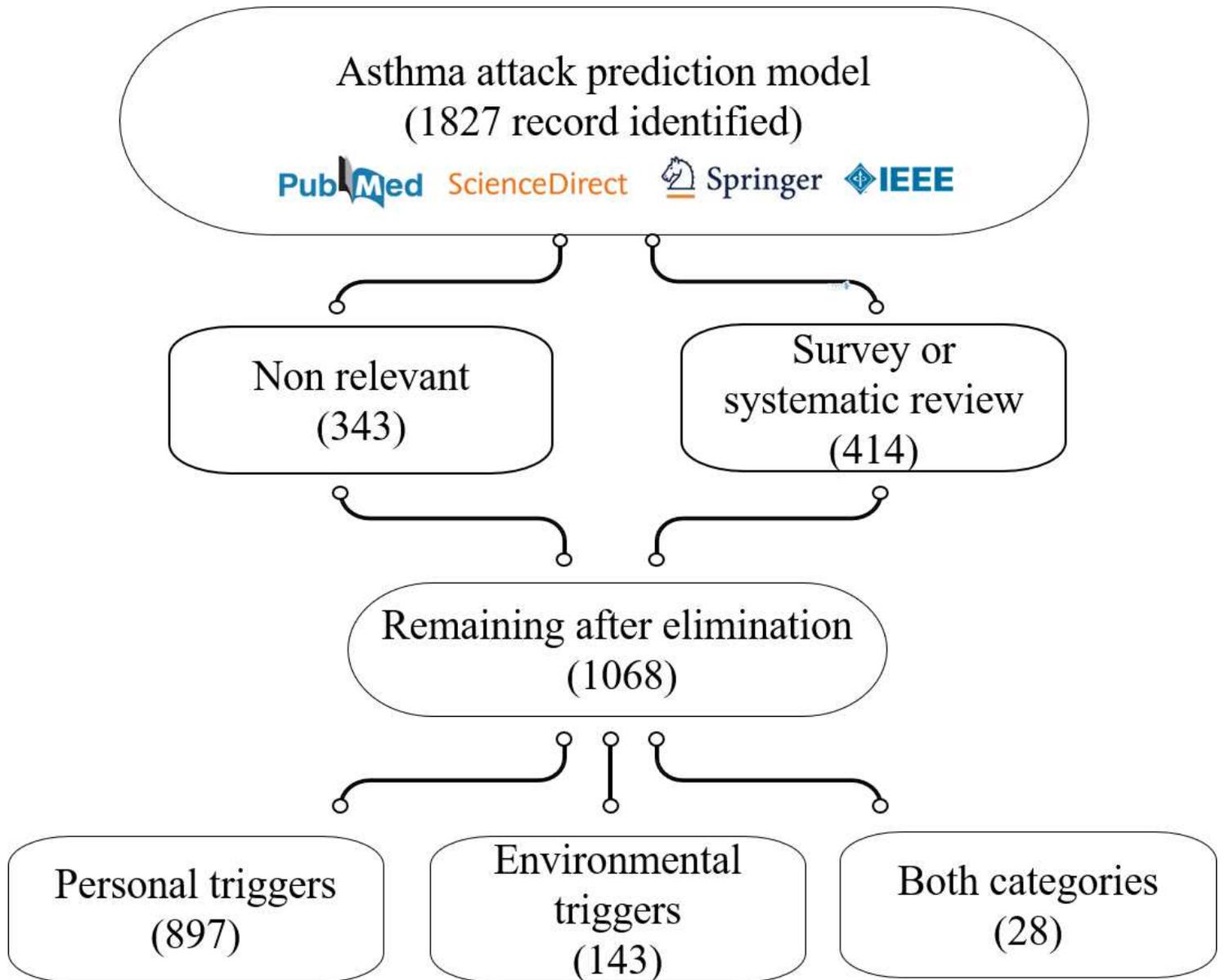
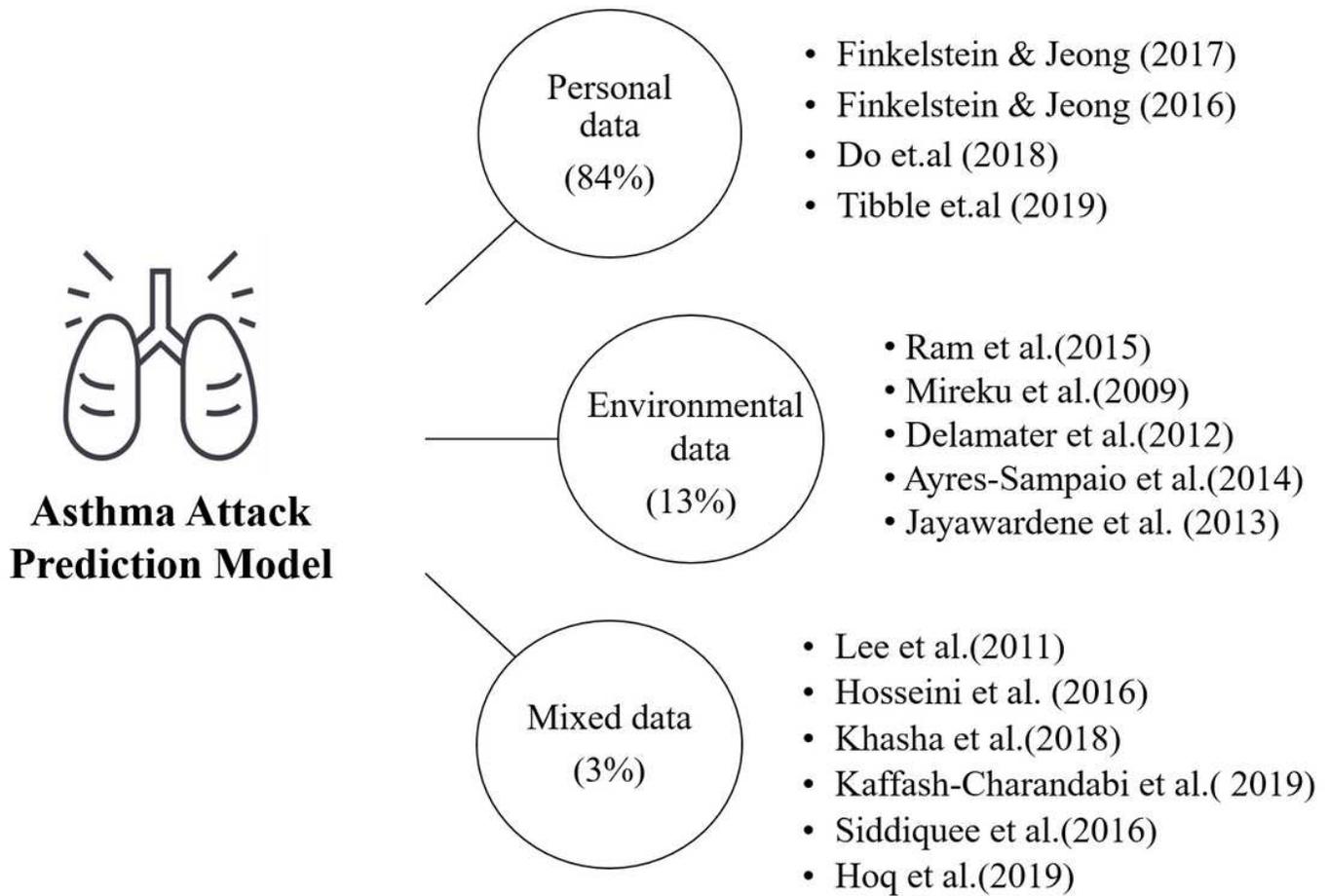


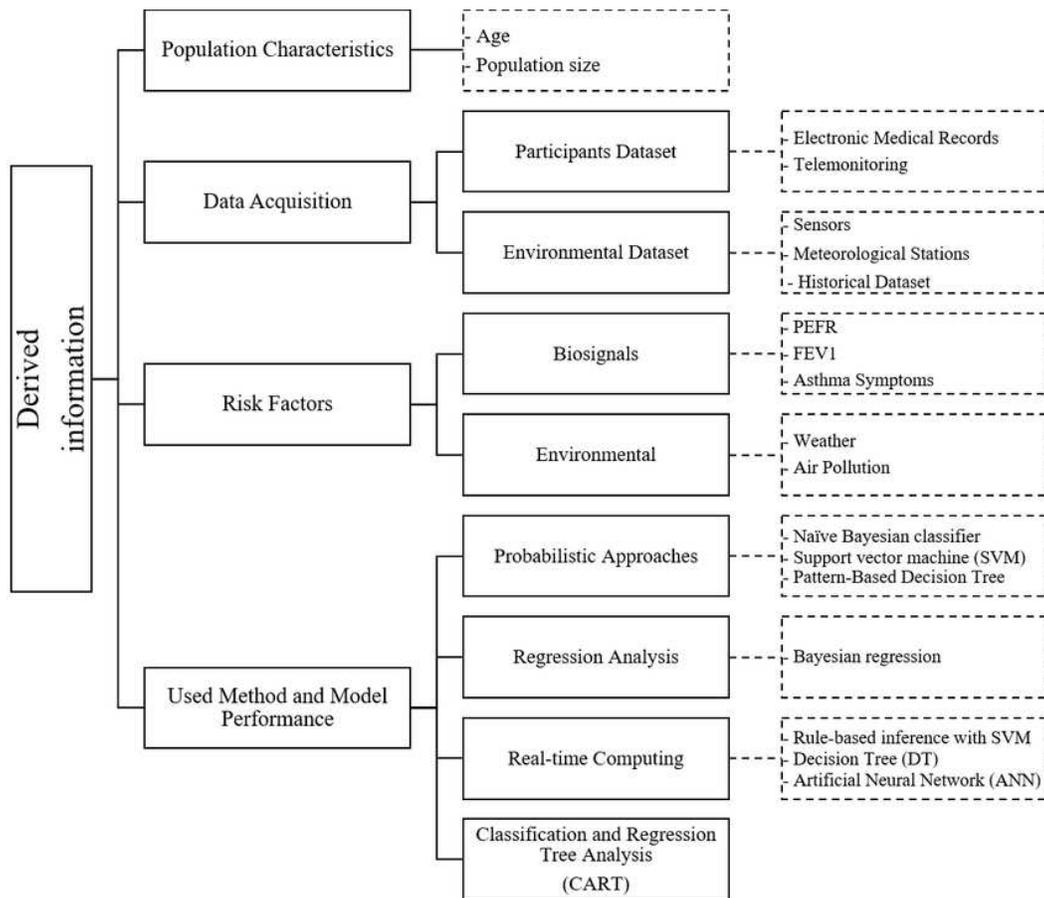
Figure 1

Research Article's Searching Mechanism.



**Figure 2**

Existing Asthma Attack Prediction Models with the Used Factors Percentage.



**Figure 3**

Taxonomy of the Review Output

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PRISMA2020abstractchecklist.docx](#)
- [PRISMA2020checklist.docx](#)