

# Application of Machine Learning Algorithms for Asthma Management with mHealth: A Clinical Review

Kevin CH Tsang<sup>1</sup>, Hilary Pinnock<sup>1</sup>, Andrew M Wilson<sup>2</sup>, Syed Ahmar Shah<sup>1</sup>

<sup>1</sup>Asthma UK Centre for Applied Research, Usher Institute, University of Edinburgh, Edinburgh, UK; <sup>2</sup>Asthma UK Centre for Applied Research, and Norwich Medical School, University of East Anglia, Norwich, UK

Correspondence: Syed Ahmar Shah, Email [ahmar.shah@ed.ac.uk](mailto:ahmar.shah@ed.ac.uk)

**Background:** Asthma is a variable long-term condition. Currently, there is no cure for asthma and the focus is, therefore, on long-term management. Mobile health (mHealth) is promising for chronic disease management but to be able to realize its potential, it needs to go beyond simply monitoring. mHealth therefore needs to leverage machine learning to provide tailored feedback with personalized algorithms. There is a need to understand the extent of machine learning that has been leveraged in the context of mHealth for asthma management. This review aims to fill this gap.

**Methods:** We searched PubMed for peer-reviewed studies that applied machine learning to data derived from mHealth for asthma management in the last five years. We selected studies that included some human data other than routinely collected in primary care and used at least one machine learning algorithm.

**Results:** Out of 90 studies, we identified 22 relevant studies that were then further reviewed. Broadly, existing research efforts can be categorized into three types: 1) technology development, 2) attack prediction, 3) patient clustering. Using data from a variety of devices (smartphones, smartwatches, peak flow meters, electronic noses, smart inhalers, and pulse oximeters), most applications used supervised learning algorithms (logistic regression, decision trees, and related algorithms) while a few used unsupervised learning algorithms. The vast majority used traditional machine learning techniques, but a few studies investigated the use of deep learning algorithms.

**Discussion:** In the past five years, many studies have successfully applied machine learning to asthma mHealth data. However, most have been developed on small datasets with internal validation at best. Small sample sizes and lack of external validation limit the generalizability of these studies. Future research should collect data that are more representative of the wider asthma population and focus on validating the derived algorithms and technologies in a real-world setting.

**Keywords:** artificial intelligence, chronic disease, smart devices, self-management, remote monitoring; asthma

## Introduction

Asthma is a variable long-term condition, affecting 339 million people worldwide,<sup>1</sup> often with diurnal, seasonal and life-time differences in symptoms and disease burden. Although, for many, asthma symptoms are controlled most of the time, some have on-going poor control and all are at risk of attacks which, at best, are inconvenient and at worst can result in hospitalization or even death.<sup>2</sup> Currently, there is no cure for asthma, therefore the focus of management is on improving symptom control and reducing the risk of attacks. Asthma is an umbrella term encompassing a range of phenotypes so personalization of management strategies is essential.

Monitoring is one of the pillars of management, allowing patients to correctly assess their health and take appropriate action. Mobile health or mHealth is commonly defined as the practice of using mobile technologies in medical care. This can range from using text reminders for medical appointments to healthcare telephone helplines to using home monitoring systems and wearable devices.<sup>3</sup> mHealth encompasses many streams of data, most of which are produced faster than a single human can comprehend; machine learning is ideal for processing this amount of data to produce actionable information and personalized feedback.

Machine learning involves using computers and algorithms to process large amounts of data (many observations and many variables) and identify patterns without explicit human programming.<sup>4</sup> It has provided insights into a very wide range of applications, including genomics,<sup>5–7</sup> images,<sup>8–10</sup> sound recordings,<sup>11,12</sup> vital signs,<sup>13</sup> and electronic health records data collected in primary,<sup>14,15</sup> secondary,<sup>16</sup> and tertiary care.<sup>17</sup> Machine learning is an umbrella term, consisting of tools and techniques that use data to learn how to perform a given task, but the algorithms generally fall into two classes, supervised and unsupervised learning. Supervised learning finds a mathematical function to link the data with known labels and is suitable for tasks that have a well-defined goal. Unsupervised learning, on the other hand, describe patterns and structures in the data without following the lead of labels or categories defined by a human. More details about machine learning algorithms are provided in the [Supplementary Material – Machine Learning](#).

Currently, most mHealth interventions that have been implemented in healthcare have focused on reminders and communications.<sup>3</sup> Areas of asthma management that machine learning and mHealth can support include monitoring,<sup>18</sup> personalizing care,<sup>19</sup> providing education,<sup>20</sup> understanding patterns in the population to better target care,<sup>21</sup> and predicting asthma attacks using a multitude of data sources.<sup>22</sup> Broadly, existing research efforts can be categorized into three types: 1) technology development, 2) attack prediction, 3) patient clustering.

This clinical review will provide a critical overview of the current research that has leveraged machine learning in the context of mHealth for remote asthma management, its shortcomings, challenges, the extent of readiness for deployment, and future research recommendations.

## Methods

We carried out a clinical review and searched PubMed for applications of machine learning to mHealth for asthma management, based on the following inclusion criteria: 1) full text available; 2) available in English; 3) published in last 5 years; 4) including at least one machine learning algorithm; 5) including data collected from humans; 6) including data other than electronic health records; 7) peer reviewed. We excluded systematic reviews, commentaries, and preprints. The terms used to search title and abstract are listed in [Table 1](#). Terms in the same column were joined by the OR operator and the search terms in different columns were joined by the AND operator. Publications in the past five years equated to publications between 1st January 2017 and 30<sup>th</sup> July 2021.

## Results

### Search Results

With our search terms, we found 90 papers available via PubMed published in the last 5 years. After reviewing the abstracts of all the papers with the inclusion and exclusion criteria, 22 papers were identified and further reviewed in this study (see [Figure 1](#)).

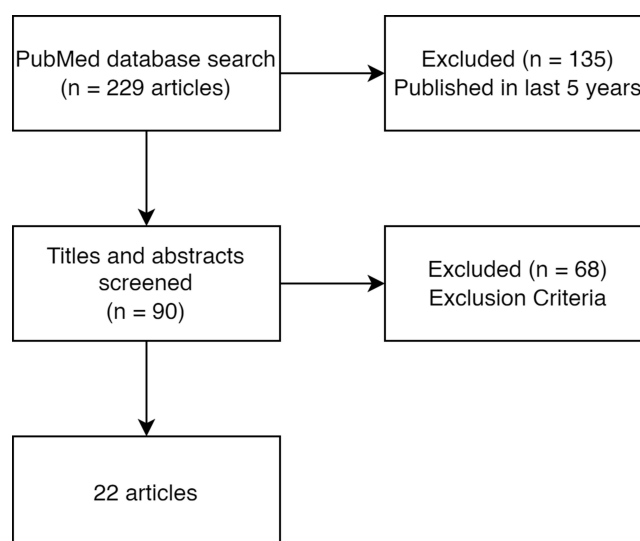
We classified the studies in three areas: technology development, attack prediction, and patient clustering. Technology development refers to contexts where machine learning is central to developing a new monitoring tool,<sup>23–33</sup> such as in cough and wheeze analysis. Attack prediction refers to studies that use machine learning to predict an asthma event (typically an attack) usually using mHealth data.<sup>34–42</sup> Patient clustering refers to studies which subtype the asthma population using unsupervised learning algorithms.<sup>43,44</sup> See [Table 2](#) for a summary of the papers.

Most applications of machine learning for asthma management in mHealth involve collecting self-reported data to form the ground truth of a patient's asthma condition, and some objective data either using smartphones or mobile

**Table 1** Search Strategy

Asthma	Machine Learning	mHealth	Validation
Asthma*	Predict*, Machine Learning, Artificial intelligence, Bayesian, Machine, Regression	mHealth, Telehealth, Telemonitor*, Monitor*, Smart*, Digital-health, eHealth, Mobile, Smartphone, Track*	AUC, Area Under the Curve, ROC, Receiver Operating Characteristic, Accuracy, Validation, Sensitivity, Specificity

**Note:** Asterisk (\*) denotes a wildcard operator; for example, "predict\*" represents "predict", "predicts", "predicting", "prediction", etc.



**Figure 1** Article selection.

monitoring devices, or both. Frequently, a validated measures of asthma control is collected (eg, Asthma Control Questionnaire (ACQ)<sup>45</sup> or Asthma Control Test (ACT)<sup>46</sup>) in mHealth studies. Using around five questions about the symptoms experienced by patients, the questionnaires determine whether patients' asthma is controlled or uncontrolled.

Many methods and devices for monitoring different aspects of a person have been studied individually and in combination. Machine learning can be applied to breath monitoring,<sup>37,41</sup> sleep monitoring,<sup>23,34–36,38,39,42</sup> cough and wheeze,<sup>24,26,27,29–31,36</sup> lung function monitoring,<sup>23,25,33–35,38,40</sup> adherence monitoring,<sup>32,35,38,43</sup> and environment monitoring.<sup>39,40,44</sup> However, studies had different outcome measures; hence, it is difficult to conduct a direct comparison between studies.

## Technology Development

Developing monitoring tools was a goal for 11 of the included studies. These include identifying sleeping postures from wearable respiration sensor data,<sup>23</sup> activity detection using smartwatches,<sup>28</sup> home breathing monitoring,<sup>25,33</sup> and active<sup>24,27,29,31</sup> and passive cough and wheeze detection.<sup>26,30</sup> Many of the identified studies on technology development applied digital signal processing (DSP) to process the raw signals collected via sensors, a necessary step before the application of machine learning.

Two<sup>27,28</sup> out of 11 studies included data from children and five<sup>23,25,27,28,32</sup> out of 11 studies included data from adults; however, none of the 11 studies developing monitoring tools had specifically investigated data from a senior population. Some of the studies on adults were conducted purely with healthy adults who could mimic a wide range of breathing patterns.

## Sleep Posture

Among patients with asthma, posture (such as standing vs supine) can influence respiratory behavior.<sup>55</sup> However, there is conflicting evidence as to whether sleeping posture has a significant effect on respiratory behavior.<sup>55–57</sup> Identifying the posture of when the respiratory measurement was taken can be useful when studying posture-related instabilities.

Using two wearable sensors located at the abdomen and chest, four postures (standing and three sleeping) were identified with high accuracy. However, the ability to correctly identify postures from sensor data was dependent on knowing to which individual the data belonged. Using this information, the classifier jumped in performance from 21.9% accuracy to 99.5% accuracy, thus adapting this method for asthma management will require more research or include a calibration stage.<sup>23</sup>

**Table 2** Summary of Studies

Study	Category	Participants [Data Source]	Devices	Collected Data	Machine Learning Algorithms	Input Features (X)	Output (Supervised) (Y)	Output (Unsupervised)	Performance	Application to Asthma Management
Chen A, 2020 <sup>23</sup>	Technology development	11 healthy adults	2 wireless wearable sensors: the abdominal respiration (Sensor1), the chest respiration (Sensor2)	Respiratory behaviors	Random Forest	100 data points sliding window, 1200 data slices per individual	4 postures: Standing, laying on the back, laying on the left, laying on the right	-	Accuracy = 99.53% (individual classifier)	Monitor sleeping posture and respiratory behavior
Vatanparvar K, 2020 <sup>24</sup>	Technology development	131 individuals (age not specified): asthma = 69, COPD = 9, asthma and COPD = 13	Smartphone (Samsung Galaxy Note 8)	1 minute of voluntary cough	Gaussian Mixture Model, neural networks	5380 sound samples of coughs	Coughing individual	-	Sensitivity = 90.30%, specificity = 96.39%, accuracy = 93.34% (NeTrain with cough embeddings)	Passive monitoring of coughs
Prinable J, 2020 <sup>25</sup>	Technology development	9 healthy adults	Pulse oximeter, portable sleep diagnostic (Alice PDx)	Raw PPG trace, SPO2, pulse rate, and relative tidal volume (RTV)	Deep learning (LSTM)	45 recordings, 4 features each: PPG, band-passed PPG, SPO2, pulse rate	-	Inspiration time, expiration time, respiratory rate, inter-breath intervals (IBI), and the inspiration-expiration ratio (I:E)	Relative bias <4% (apart from I:E ratio)	Passive monitoring of breathing
Adhi Pramono R. X, 2019 <sup>26</sup>	Technology development	Unknown individuals from multiple repositories <sup>47</sup>	Unknown devices	Cough sounds	Logistic regression	43 recordings. Frequency bands of interest: B-HF and B-01. The spectral features: HFMaxratio, MinMaxratio, and LQMAXratio	2 classes: cough, non-cough	-	Sensitivity = 90.31%, specificity = 98.14%, F1-score = 88.70%, positive predictive value = 88.47%, Matthews Correlation Coefficient (MCC) = 87.46%	Passive monitoring of coughs

Chen H, 2019 <sup>27</sup>	Technology development	126 individuals (all ages, infants to elderly) (including asthma and COPD) [ICBHI Scientific Challenge] <sup>48</sup> and unknown individuals [R. A.L.E lung sounds] <sup>49</sup>	Digital stethoscopes	Respiratory sounds	SVM, Extreme Learning Machine (ELM), KNN	240 recordings, 2 features extracted from Enhanced Generalized S-Transform (EGST): mean and standard deviation of EGST coefficients	2 classes: wheezing, normal respiratory	-	Sensitivity = 100%, specificity = 99.27% (ELM, SVM, KNN)	Active monitoring of wheeze
Li K, 2019 <sup>28</sup>	Technology development	30 adults (age 19-48) [HARuS] <sup>50</sup> and 14 children (age 5-15) with asthma [BREATHE] <sup>51</sup>	[HARuS] waist-worn smartphone (Samsung Galaxy S II) and [BREATHE] wrist-worn smartwatch (Motorola Moto 360 Sport)	Triaxial accelerometry and gyroscopic data	XGBoost (Gradient Boosted Trees), SVM, random forest	2000 inspiratory and expiratory segments of sounds, 6 features per window of signal: arithmetic mean, SD, median absolute deviation, minimum, maximum, and entropy	6 physical activities: [HARuS] 6 activities: standing, sitting, lying, walking, walking downstairs, walking upstairs; [BREATHE] 6 activities: standing, sitting, lying, walking, walking on stairs, running	-	[HARuS] Accuracy rate = 91.06% (GGS), [BREATHE] accuracy rate = 79.4% (GGS)	Activity recognition of smart watches
Azam M.A, 2018 <sup>29</sup>	Technology development	50 individuals (age not specified) with COPD, asthma, bronchitis, and pneumonia	Smartphone (Samsung Galaxy S3)	Airflow 25 cm in front of mouth	Bag-of-Features, SVM	255 breathing cycles, 5 features extracted from instantaneous envelop (IE) and instantaneous frequency (IF)	2 classes: normal, Adventitious Signal (AS)	-	F1-score = 75%, accuracy rate = 75.21 (complete cycle)	Active monitoring of breathing sounds

(Continued)

Table 2 (Continued).

Study	Category	Participants [Data Source]	Devices	Collected Data	Machine Learning Algorithms	Input Features (X)	Output (Supervised) (Y)	Output (Unsupervised)	Performance	Application to Asthma Management
Adhi Pramono R. X, 2019 <sup>30</sup>	Technology development	Unknown individuals from multiple repositories <sup>47</sup>	Unknown devices	Cough sounds	Logistic regression	43 unique recordings, 4 features each: Linear Predictive Coding (LPC) coefficient, tonality index, spectral flatness, and spectral centroid	2 classes: cough, non-cough	-	Sensitivity = 86.78%, specificity = 99.42%, F1-score = 88.74%	Passive monitoring of coughs
Infante C, 2017 <sup>31</sup>	Technology development	87 individuals (age not specified): COPD = 7, asthma = 15, allergic rhinitis = 11, asthma and allergic rhinitis = 17, COPD & allergic rhinitis = 4, healthy = 33	Custom-built electronic stethoscope and Android application	Voluntary coughs recorded from the trachea (30 seconds), standard auscultation lung sound data, peak flow meter reading and clinical questionnaire	Logistic regression with L1 penalization (LASSO)	4 features: Zero Crossing Irregularity, Rate of Decay, Kurtosis, Variance	2 set of labels, diagnosis; cough type (wet or dry).	-	Sensitivity = 100%, specificity = 87%, AUC = 94% (Wet vs dry) Sensitivity = 35.7%, specificity = 100%, AUC = 67.8% (classifying unhealthy patients with cough type)	Active monitoring of coughs
Taylor T.E, 2018 <sup>32</sup>	Technology development	20 healthy adults	Inhaler Compliance Assessment (INCA) audio recording device, pneumotachograph spirometer	Audio recording of placebo Ellipta inhalation, inhalation rate	Linear regression, power law regression	15 inhalations per person, acoustic envelope of the inhaler inhalation	Flow rate: peak inspiratory flow rate (PIFR), volume or inspiratory capacity (IC), and the inhalation ramp time (Tr)	-	Accuracy = 90.89% (power law model)	Measure correct inhaler technique

Purnomo A. T, 2021 <sup>33</sup>	Technology development	Unknown individuals	FMCW radar	5 to 15 seconds of chest displacement breathing waveforms	XGBoost (Gradient Boosted Trees)	4000 breathing waveforms. [set 1] breathing wave form; [set 2] 8 features: mean, median, maximum, variance, standard deviation, absolute deviation, kurtosis, and skewness; [set 3] MFCC feature extraction	5 classes: normal breathing, deep and quick breathing, deep breathing, quick breathing, holding the breath	-	Precision > 80%, sensitivity > 70%, F1-score > 75% (for all classes, MFCC feature extraction)	Active monitoring of breathing
Zhang O, 2020 <sup>34</sup>	Attack prediction	2010 individuals (age > 16) with severe and persistent asthma [SAKURA] <sup>52</sup>	Paper diary	Daily questionnaire: PEF, morning symptoms, evening symptoms, reliever inhaler usage, asthma sleep wakening	Recursive feature elimination, PCA, random under-sampling, random over-sampling, SMOTE, logistic regression, naïve Bayes, decision tree, perceptron	728,535 daily records, 432 features, 9 basic features	2 classes: exacerbation event, no exacerbation	-	Sensitivity = 90%, specificity = 83%, AUC = 85% (logistic regression)	Attack prediction from daily diary and PEF
Tsang K.C. H, 2020 <sup>35</sup>	Attack prediction	554 adults with asthma [AMHS] <sup>53</sup>	Smartphone (BYOT)	Daily and weekly questionnaire: symptoms, healthcare usage, medication usage, triggers encountered, PEF	Decision trees, logistic regression, naïve Bayes, and SVM	2659 periods, 25 features per 14-day period before unstable event, 6 basic features	2 classes: stable, unstable period	-	Sensitivity = 86.6%, specificity = 72.5%, AUC = 87.1% (naïve Bayes)	Attack prediction from daily diary
Tinschert P, 2020 <sup>36</sup>	Attack prediction	79 adults with asthma	Smartphone (Samsung Galaxy A3) application based on MobileCoach	ACT, Pittsburgh Sleep Quality Index, Nocturnal cough frequencies (manually labelling audio recordings from the smartphone's built-in microphone)	Mixed-effects regressions, decision trees based on recursive partitioning analysis	2291 nights, 7 combinations of Pittsburgh Sleep Quality Index and Nocturnal cough frequencies	ACT score	Prediction of exacerbation risk in the next 7 days	56% < balanced accuracy < 70%	Attack prediction from sleep quality

(Continued)

Table 2 (Continued).

Study	Category	Participants [Data Source]	Devices	Collected Data	Machine Learning Algorithms	Input Features (X)	Output (Supervised) (Y)	Output (Unsupervised)	Performance	Application to Asthma Management
Tenero L, 2020 <sup>37</sup>	Attack prediction	38 children (age 6-16): persistent asthma = 28, control = 10	Electronic nose (Cyranose 320)	VOCs in exhaled breath, spirometry	PCA, penalized logistic model	1 recording per person, 32 e-Nose nanosensors	4 classes: control (CON), controlled asthma (AC), partially controlled asthma (APC) or uncontrolled asthma (ANC)	6 most important sensors, 5 principal components	Sensitivity = 79%, specificity = 84%, cross-validated AUC = 80%	Asthma control prediction from exhaled breath
Finkelstein J, 2017 <sup>38</sup>	Attack prediction	Adults with asthma	Peak flow meter connected to laptop	PEF, daily questionnaire: symptoms, medication usage, trigger exposure, sleep	Naïve Bayes, adaptive Bayesian network, SVM	7001 records, 147 features, 21 basic variables x 7 days	2 classes: high-alert, no-alert PEF zone on day 8	-	Sensitivity = 100.0%, specificity = 100.0%, accuracy = 100.0% (adaptive Bayesian network)	Attack prediction from daily diary and PEF
Castner J, 2020 <sup>39</sup>	Attack prediction	43 adults (working aged women) with poorly controlled asthma	Fitness tracker (Fitbit Charge), activity monitor (Actigraph GT3X+), spirometer (Vyntus), spirometer (MicroDiary), home monitor (Hobo Data Logger)	ACT, Mini Asthma Quality of Life Questionnaire (AQLQ), trait emotionality PANAS-X questionnaire, Consensus Sleep Diary, asthma control diary (ACD), physiologic and environmental sensors, medical record review, and spirometry.	Generalized linear mixed models	900 daily scores, [set 1] 8 features; [set 2] 10 features	[set 1] self-reported asthma-specific waking; [set 2] FEV1	-	[set 1] AUC = 77% (sleep waking) [set 2] AUC = 83% (FEV1)	Measure sleep disruption using fitness tracker



Khasha R, 2019 <sup>40</sup>	Attack prediction	96 individuals (age >5) with asthma	Weather reports, Air quality, questionnaires, medical records	140 variables about patient demographics, lung function, symptoms, environmental factors, medical history	Ensemble learning, multinomial logistic regression, SVM, random forest, extreme gradient boosting, KNN, decision tree, Gaussian naïve Bayesian, rule-based classifier created from clinical knowledge	2870 daily records, 35 selected variables	3 classes: well-controlled, not well-controlled, very poorly-controlled levels	-	Sensitivity = 88.3%, precision = 89.4%, specificity = 94.9%, neg pred value = 94.3%, accuracy = 92.7% (Ensemble Learning 2)	Asthma control prediction using health records and weather reports
Van Vliet D, 2017 <sup>41</sup>	Attack prediction	96 children (age 6-18) with asthma	NIOX analyzer (NIOX MINO), 5 liter inert bag with a resistant free valve (Tedlar bag), spirometer (ZAN 100®)	ACQ, GINA respiratory symptom score, online FeNO assessment, collection of exhaled breath (VOCs), dynamic spirometry	Random forest, PCA	574 chromatograms, 7 VOCs	Exacerbations	Most important VOCs, separation of children with and without exacerbation, possible difference in co-factors between samples of children with an exacerbation 14 days after sampling and those without an exacerbation after 2 months	Sensitivity = 88%, specificity = 75%, AUC = 90% (attack after 14 days)	Attack prediction from exhaled breath and home monitoring
Huffaker M. F, 2018 <sup>42</sup>	Attack prediction	16 children (age 5-18) with persistent asthma	BCG accelerometer-based passive bed sensor (Murata Technologies SCA11H)	Heart rate (HR), respiratory rate (RR), HR variability HRV, calculated RR variability (RRV), relative stroke volume (SV), HR percentile based on age, RR percentile based on age, movement, relative Q (= SV × HR), VO2	Random forest	891 nights, 16 features, 8 basic features	Report of asthma symptoms	-	Sensitivity = 47.2%, specificity = 96.3%, accuracy = 87.4%	Attack prediction from sleep and bed monitoring

(Continued)

**Table 2** (Continued).

Study	Category	Participants [Data Source]	Devices	Collected Data	Machine Learning Algorithms	Input Features (X)	Output (Supervised) (Y)	Output (Unsupervised)	Performance	Application to Asthma Management
Tibble H, 2020 <sup>43</sup>	Patient clustering	211 children (age 6-15) with asthma attack <sup>54</sup>	Electronic inhaler monitoring devices	Medication dose taken	PCA, K-means, decision trees	35,161 person-days of data, 5 features: the percentage of doses taken, the percentage of days on which zero doses were taken, the percentage of days on which both doses were taken, the number of treatment intermissions per 100 study days, and the duration of treatment intermissions per 100 study days	-	3 clusters: poor adherence, moderate adherence, good adherence	-	Characterize asthma patients by adherence
Tignor N, 2017 <sup>44</sup>	Patient clustering	334 adults with asthma [AMHS] <sup>53</sup>	Smartphone (BYOT)	Daily questionnaire: symptom diary, medication, triggers encountered	Probability based imputation with consensus clustering (PIC) method (utilize k-means)	1 recording per person. [Cluster formation] daily symptoms; [Characterize formation] 10 features: 4 clinical features, 3 demographic features, 3 trigger features	-	3 clusters: high day symptom rate, medium day symptom rate, low day symptom rate	-	Subtyping asthma patients for personalized alerts based on triggers

**Abbreviations:** AUC, Area under the ROC curve; BYOT, Bring your own technology; COPD, Chronic Obstructive Pulmonary Disease; FeNO, Fractional exhaled nitric oxide; GINA, Global Initiative for Asthma; LSTM, Long short-term memory; PCA, Principal Component Analysis; PPG, Photoplethysmogram; SVM, Support Vector Machine; VOCs, Volatile organic compounds.

## Activity Detection

Smartwatches are increasingly prevalent amongst the public, healthy individuals, and elite athletes to measure their health. This has promoted technology development, so that the sensors are more reliable, affordable, and comparable between brands.<sup>58</sup> Motion data (triaxial accelerometry and gyroscopic data) commonly collected in smartwatches was used in activity detection, which could improve the capabilities of passive monitoring potentially replacing the need to ask questions about activity. Using DSP to process the raw signals and supervised learning (gradient boosted tree classification) on two datasets, various activities like standing, sitting, and walking were identified from signals from the wrist worn device with promising accuracy.<sup>28</sup>

In a comparison between the performance of algorithms trained on two datasets, one in adults and one in children, found the activity detection performed better in adults, but this was confounded by the adults performing tightly proscribed movements and the children recording more natural movements.<sup>28</sup>

## Breathing Monitoring

Breathing monitoring and detecting difficulties in breathing could help potentially identify asthma attacks early. Tools that have been proposed for home monitoring include portable sleep diagnostic devices to monitor breathing,<sup>25</sup> and radar to measure chest movement.<sup>33</sup> Using deep learning and features from a pulse oximeter, there were accurate predictions of the respiratory waveforms.<sup>25</sup> Likewise, applying supervised learning (XGBoost) on features extracted from chest movement recorded by the radar gave promising accuracy of identifying different breathing patterns.<sup>33</sup>

## Cough Monitoring

Like sleep monitoring, wheeze and cough are widely captured as a measure of asthma control and included in validated asthma questionnaires. However, there are also studies combining mHealth and machine learning to develop new tools for monitoring wheeze and cough, both actively<sup>24,27,29,31</sup> and passively.<sup>26,30</sup> Recording and analyzing voluntary coughs and respiratory sounds from people with different respiratory diseases could provide a tool to assist diagnosis. Although separating wet (cough with phlegm) and dry coughs was successful, there were varying levels of performance when making a diagnosis using recordings alone.<sup>24,29,31</sup> Using voluntary cough recordings, one study accurately predicted individuals who were either healthy, had asthma, had chronic obstructive pulmonary disease (COPD), or had comorbid asthma and COPD with an accuracy of 93.3%.<sup>24</sup> In contrast, another study using cough type to distinguish healthy people from those with respiratory disease had a much lower performance, with AUC of 67.8%.<sup>31</sup> Developing new DSP methods (an essential step to be able to extract relevant information from raw sound signals) have shown promise in wheeze and cough detection from digital stethoscope recordings.<sup>26,27,30</sup>

## Inhaler Technique Monitoring

Measuring adherence to medication is widely studied in asthma research. In addition to measuring when patients took medication, measuring how the inhalers were used and checking for correct technique is another application of mHealth and machine learning. Regression models of DSP processed adult audio recordings from the INhaler Compliance Assessment (INCA) device were found to accurately estimate the inhaler inhalation flow profile with 91% accuracy.<sup>32</sup> This objective measure of inhaler technique could help patients improve how they take their medication.

## Attack Prediction

Machine learning was applied to several different mHealth data sources to predict asthma attacks and change in symptoms. The data included volatile organic compounds,<sup>37,41</sup> sleep quality,<sup>36,39,42</sup> peak flow,<sup>34,35,38,40</sup> preventer medication adherence,<sup>35,38</sup> and environmental triggers.<sup>39,40</sup> Two<sup>34,40</sup> of the nine studies included data collected from children or teenagers, and adults, but the population was considered as a whole in both cases. Three studies<sup>37,41,42</sup> focused on children with asthma, four studies<sup>35,36,38,39</sup> focused on adults with asthma, and none of the studies focused on seniors. The performance of the algorithms was unlikely to have been affected by the age group of the study population.

## Breath Analysis

Volatile organic compounds (VOCs), stemming from indoor pollutants, that are present in the breath of patients could be used to understand the development of asthma attacks, but evidence is inconsistent.<sup>59</sup> Gas chromatography–mass spectrometry (GC-MS) is the gold standard in VOC analysis, but electronic nose (e-Nose) could be a portable alternative. The e-Nose can detect and recognize individual chemical compounds in mixtures of chemical vapors.

The VOCs in exhaled breath of children were analyzed using both supervised and unsupervised learning.<sup>37,41</sup> Supervised learning methods (penalized logistic models and random forest) were used to identify the most important VOCs for attack prediction. Classifiers were trained to identify which VOCs would predict an upcoming asthma attack or worsening control. The study reported good performance, with sensitivity and specificity between 70% and 90%, and an AUC upwards of 80%. Furthermore, unsupervised learning (principal component analysis (PCA)) was used to pre-process the data to form combinations of VOCs for attack prediction and for visualizing high-dimensional data in a two-dimensional graph.<sup>37,41</sup>

## Sleep Monitoring

Aligned with the clinical recognition of exaggerated diurnal variation causing sleep disturbance as a sign of poorly controlled asthma,<sup>60,61</sup> disturbance to sleep was widely used as a potential predictor of worsening asthma. Many studies captured night symptoms and sleep quality using questionnaires,<sup>34,35,38</sup> but some collected objective sleep data using devices.<sup>36,39,42</sup> Out of 25 features used to predict asthma attacks with daily (symptom diary like-) questionnaires about asthma, night symptoms-related features were two of the four most predictive features.<sup>35</sup> Also, night-time waking was selected as one of three basic variables used for prediction.<sup>34</sup> When the objective data were combined with machine learning algorithms (random forest, generalized linear mixed models, regression), it enabled smartphone recordings to analyze nocturnal coughs,<sup>36</sup> related fitness tracker activity data with sleep waking,<sup>39</sup> and bed sensors to predict asthma control.<sup>42</sup> The usefulness of using sensors to predict self-reported asthma control is unclear, using nocturnal cough and sleep quality alone achieving balanced accuracy of no more than 70% in predicting attacks,<sup>36</sup> but using fitness tracker data to predict sleep waking had an AUC of 77%,<sup>39</sup> and an accuracy of 87.4% in predicting reports of asthma symptoms.<sup>42</sup>

## Lung Function Monitoring

Falling peak expiratory flow (PEF) is a major indicator of asthma attacks. Peak flow meters are sometimes used by patients at home to take objective measurements and used to inform whether action needs to be taken. Spirometers are another device that measures lung function, but in more detail than peak flow meters.<sup>62</sup> Action plans use thresholds of 80% of their best PEF to determine that action needs to be taken, and urgent action is required if a person's PEF falls below 60%.<sup>60</sup> A drop in PEF and/or a change in symptom score are widely used in asthma action plans to determine self-management in response to deterioration.<sup>63</sup> Smart peak flow meters enable patients to measure and track their PEF, and are often linked with a mobile app to function.

Measuring PEF to monitor lung function is commonplace in asthma studies. This could be either reporting the results from a traditional peak flow meter,<sup>34,35,40</sup> or using a smart peak flow meters that sends the data through a computer or smartphone.<sup>38</sup> PEF measurements are used as both predictors of asthma attacks as well as defining severity and informing management. Using daily diaries and PEF measurements to predict worsening condition with supervised learning (adaptive Bayesian network) achieved a performance of 100.0% accuracy, sensitivity, and specificity.<sup>38</sup>

## Adherence Monitoring

Adherence to regular preventative medication is sometimes captured by questionnaire and used as a predictor for asthma attacks.<sup>35,38</sup> Although clinically important, the two studies did not identify the adherence to controller medication as an important predictive feature in their methods. In contrast, and consistent with clinical recommendations, features based on the use of short-acting reliever medication were two of the four most predictive features.<sup>35</sup>

## Environment Monitoring

Some common asthma triggers in the environment, such as pollen, meteorological change, and air pollution (eg, particulate matter, carbon monoxide (CO), nitrogen dioxide (NO<sub>2</sub>)), could be monitored to reduce risk of exposure to known triggers. Also, recording asthma triggers encountered, such as viral infections, passive smoke, and pets, could give a better understanding of a person's asthma and their symptoms.<sup>64–66</sup> Connecting data from pollution monitoring stations and meteorology stations with patient health records provides a wealth of information for analysis.

Furthermore, combining physicians' knowledge using a rule-based classifier (analogous to a decision tree created based on knowledge) with conventional supervised learning techniques (multinomial logistic regression, SVM, random forest, extreme gradient boosting, KNN, decision tree, Gaussian naïve Bayesian) created an accurate (sensitivity of 88.3% and precision of 89.4%) ensemble learning algorithm for predicting levels of asthma control.<sup>40</sup> Based on the joined dataset, the most important features for prediction were lung function and symptoms: PEF in the morning and before bedtime, ACT score, and shortness of breath in the last 24 hours. Although environmental features were not ranked highly, daily NO<sub>2</sub> concentration and daily temperatures were useful.<sup>40</sup> Further, home environment measuring device has also been shown to be useful in predicting self-reported asthma-specific waking.<sup>39</sup>

## Patient Clustering

Two studies<sup>43,44</sup> used unsupervised learning to form data-driven clusters using data collected via mHealth. One study was investigated clusters in children with asthma,<sup>43</sup> the other had focused on data collected by adults with asthma.<sup>44</sup>

## Adherence Monitoring

In addition to capturing adherence to regular controller medication via questionnaires, there has also been in-depth studies of medication adherence. Smart inhalers are devices that objectively measure how inhaler medication is taken, as an alternative to self-report. Monitoring can be applied to the long-acting controller inhaler or the short-acting reliever inhaler, or both. By analyzing electronic inhaler monitoring data of controller medication with unsupervised learning algorithms (PCA and k-mean), asthma patients were characterized by multi-dimensional inhaler adherence measures, which formed three groups, poor (on average 16% of their prescribed doses), moderate (averaged 60% of dose), and good (averaged 91% of dose) adherence.<sup>43</sup> Furthermore, comparison with clusters formed by another data-driven method (decision trees) yielded similar results.<sup>43</sup>

## Environment Monitoring

Like many daily questionnaires, recording encounters with asthma triggers can be difficult and lead to missing data. To tackle this, probability-based imputation with consensus clustering was developed as a method of imputing the missing data and clustering patients, which can be used to subtype asthma patients for personalized alerts based on their triggers.<sup>44</sup> Using the imputation method, three patient clusters were formed using the daily asthma symptom data. The characteristics of each cluster was investigated on four clinical, three demographic, and three trigger features. Cluster 1, with the highest average day symptom level, had patients who frequently reported pollen and heat as their triggers. On the other hand, cluster 3, with the lowest average day symptoms, was characterized more by patients citing air quality as their trigger.<sup>44</sup> Prospectively, weather forecasts could be useful in predicting the risk of a future asthma attack for patients who are sensitive to environmental triggers such as sudden temperature changes or high pollen levels.

## Discussion

This review has described a range of machine learning applications being used to support asthma management, in the areas of developing novel technology,<sup>23–33</sup> predicting acute attacks at an individual level,<sup>34–42</sup> and informing understanding of asthma phenotypes by clustering patients within populations.<sup>43,44</sup> There were examples of successful application of machine learning to achieve a novel task (such as attack prediction from sleep quality, control prediction from exhaled breath, characterize asthma patients by medication adherence)<sup>36,37,42,43</sup> or to improve existing methodology by using fewer resources for similar or better performance (such as smartphone-based passive monitoring of coughs).<sup>24,26,27,30,31,40,41</sup>

Most of the machine learning algorithms applied were easily interpretable,<sup>26–32,34–39</sup> a desirable characteristic to help easily understand the decision process in a clinical context. However, a few studies applied more complex but less interpretable machine learning algorithms.<sup>24,25,40</sup>

## Developing Novel Technology: Proof-of-Concept with Clinical Potential

Using machine learning, new home monitoring tools were under development, including for activity detection, breath monitoring, cough monitoring, and inhaler technique monitoring.<sup>23–33</sup> Most studies were in the proof-of-concept stage and although they were developed on selected small populations, many had achieved promising performance.<sup>23–25</sup> An initial challenge, before considering the clinical potential of novel technology, is to process the incoming data so that background noise is removed and clear signals emerge.<sup>29</sup> This was the focus of several of the papers that described development of new methods to filter the signal data.<sup>26,27,29</sup> Before using the novel technology to monitor asthma at home, validation studies should be conducted in a real-world environment.

## Prediction of Attacks: Supporting Individual Self-Management

Asthma is a variable condition,<sup>67</sup> and central to supported self-management is the ability to recognize early evidence of deterioration and to take appropriate timely action to prevent a serious attack.<sup>68,69</sup> A key aim of many of the machine learning papers was to use a wide variety of data sources to identify an individual's risk of uncontrolled asthma and to improve prediction of asthma attacks.<sup>34–42</sup> All the predictors explored (asthma symptoms, PEF, VOCs, fractional exhaled nitric oxide (FeNO), heart rate, respiratory rate, sleep quality, medication adherence, and environment) showed promise, though it was widely discussed that combining multiple varied data sources could help improve asthma attack prediction.<sup>28,34,35,38,40</sup> Importantly, the prediction algorithms were developed retrospectively and require external validation in different datasets before they can be used in clinical practice. Besides the need for external validation, future studies should also consider evaluating the algorithms by comparison to existing effective “action plans” in clinical practice.

## Clustering Patients: Informing Phenotypes and Targeting Care

Contemporary understanding of asthma as an umbrella term describing a heterogeneous group of conditions<sup>70</sup> has increased interest in identifying phenotypes of asthma amenable to specific treatments or carrying specific risks of poor symptom control and/or acute attacks. Using unsupervised learning algorithms, progress has been made on forming patient clusters representing natural patterns spotted in the data.<sup>43</sup> Understanding phenotypes not only has value in terms of individual risk and targeting care to “treatable traits” but can inform health service delivery as appropriate care can be targeted on high-risk populations.<sup>71</sup> However, many of the studies used relatively small datasets – and often of populations selected for frequent symptoms or willingness to monitor – with limited generalizability to the whole asthma population.<sup>23–25,31,36,37,39,41–43</sup> Future research should consider larger sample sizes that can better represent the general asthma population.

## Machine Learning Applied to Asthma Management: Challenges

### Tailored Data Collection

The performance of machine learning algorithms largely depends on the input data; hence, the sample size and data pre-processing methods must be considered in conjunction with the performance metrics. Most data used to train the machine learning algorithms in this review had small sample sizes, and sometimes used narrow inclusion criteria to collect the data.<sup>23–25,31,36,37,39,41–43</sup> For example, a common exclusion criterion for asthma studies is “other respiratory disease”,<sup>23,37,41,43,44</sup> which makes for a homogeneous dataset (which may be easier to analyze) but it reduces the likelihood of the results being generalizable. It also overlooks the possibility that the conditions excluded may be part of the phenotype. Even within asthma, different individuals have different medication regimes, which complicates the analysis,<sup>43</sup> but selection according to a specific regime (say prescribed combination controller medication) will only give information on a selected population. Importantly, in longitudinal studies where participant retention is a factor, different

individuals may provide different amounts of data for analysis, which will skew analysis towards patients who are more engaged with the study, more adherent to data collection, possibly influenced by the characteristics of their asthma.<sup>42,44</sup>

### Secondary Analysis of Existing Datasets

To tackle the problem of small sample sizes, some studies have conducted secondary analysis on data that were collected for a different purpose.<sup>27,34</sup> Eight studies (36%) were based on data that were publicly available or available on request.<sup>26–28,30,34,35,43,44</sup> This makes for efficient use of data, but the aims (and thus eligibility) of the original dataset may not match the aims of the new analysis thereby making the interpretation of the results more challenging.

### Missing Data

How the analysis handled missing data will be important to understand the differences between studies.<sup>35,40,42,44</sup> If the amount of missing data is small, removing the cases with missing data is an option. Alternatively, imputing the missing values is a method that avoids losing data, but is a major challenge when there is a low response rate or the data are not missing at random<sup>44,72,73</sup> (eg, people with frequent attacks may monitor more regularly than those who rarely have symptoms). Other methods to handle missing data include interpolation into regular spacing or creating summary windows,<sup>35</sup> which can then be analyzed using regular methods. However, each method of handling missing data carries their assumptions (for example, assuming people with missing inhaler data and people who reporting using and not using their have the same inhaler usage rate).

### Low Event Rate in the Dataset

For many people with less severe asthma, attacks are infrequent leading to large “class imbalance”. In some populations, the imbalance can be upwards of 90%.<sup>26,34–36,38,40</sup> Data analysis sampling techniques, such as Synthetic Minority Oversampling TEchnique (SMOTE),<sup>74</sup> have been applied to balance out the classes by essentially multiplying the minority class, which allows machine learning techniques to function properly. For example, oversampling techniques can be used to artificially enlarge the number of asthma attacks such that the data now has 50% attacks and 50% controlled asthma.

### Inconsistent Output Definitions During Modelling

Different studies of asthma attack predictions had different definitions of an asthma attack and outcome measures. This included using patient symptoms,<sup>36,37,39–42</sup> self-reported asthma attack treatment,<sup>34,35</sup> and spirometry measurements.<sup>38,39</sup> Although sometimes similar, the different definitions cannot be used in direct comparison.<sup>73</sup> Furthermore, some outcomes were easier to model based on the input data, thus leading to over-optimistic performance results. For example, Finkelstein and Jeong used 21 daily measures, including symptoms and PEF, to predict asthma attacks.<sup>38</sup> However, the asthma attacks were defined as the PEF zone on day 8, which is directly related to one of the input features, namely PEF on day 7. Consequently, it is not sufficient to assess any study based solely on the performance metrics without the broader context.

### External Validation

For external validation, the “new” dataset must be the similar in at least the key parameters as the training dataset to meaningfully compare the machine learning algorithms. Ideally, and especially for health data, the methods should be robust and comparable even if there are slight differences in the data. It is highly challenging to externally validate machine learning models partly due to major differences in inclusion criteria and outcome definitions, and most often due to lack of access to comparable data.<sup>26,30,41</sup> Slight differences in wording of questions or device choice can create datasets that are similar yet not directly comparable, hence not applicable for external validation (for example, acute attacks might be measured as “needing an oral steroid course” or “unscheduled care” and might be assessed over a year or a few months). In the context of mHealth, this requires similar devices to be used, but rapidly advancing technology may make this a challenge. However, this may change in the future as devices become validated and widely used (like how validated questionnaires and guidelines have allowed studies to be comparable).

None of the machine learning algorithms in the 22 studies had been externally validated and were only internally validated.



## Data Quality

Conducting data collection in controlled environments enables cleaner data to be collected and analyzed.<sup>27,29</sup> However, real-world settings will most likely lead to reduced data quality. Consequently, it is important that a given model's performance is evaluated for use by actual patients in their day-to-day lives.<sup>32,33</sup>

## Future Direction

Machine learning algorithms are dependent on the data that is inputted. Since most existing studies are based on relatively small sample sizes and often selected populations, the next natural step is to validate the results in larger – and more representative – populations.<sup>25,39,43</sup> Future research should consider adding other data sources to existing models, collecting multi-dimensional data using several devices and data sources simultaneously to provide a more complete picture about a person and their environment, whilst also assessing the utility of individual devices.<sup>25,28,34,35,38,40</sup> Studies like MyAirCoach<sup>22</sup> and Biomedical REAL-Time Health Evaluation (BREATHE)<sup>51</sup> that combine several sources of data longitudinally are important for future development of mHealth technologies for asthma.

The data used to train the machine learning models included data collected from children, teenagers, and adults, patients with asthma, COPD, and other respiratory diseases, some exclusively and others in combination. Although any variation of the performance in the algorithms trained on data from either age group was unlikely to be directly related to the age, it remains to be seen if the model developed for one population can perform comparably with a new or more general population.

Expanding the functionality of technologies developed, improving performance, and validating results against other devices is another area for future research.<sup>23,24,27,31,33,37,41</sup> For example, wheeze detection could be extended to other breath sounds,<sup>27</sup> expanding its application to other respiratory diseases. Cough detection could be applied to more difficult data, such as a mix of multiple individuals and background noise,<sup>24</sup> much like the “cocktail party problem” in machine learning. Developments in image recognition and video analysis using machine learning is promising<sup>8–10</sup> and could be applied to enhance inhaler technique monitoring.

The data generated by mHealth devices for home monitoring are increasingly reliable and validated against existing gold-standard equipment.<sup>58,75,76</sup> However, the validity of the information created by machine learning analysis has not yet reached the standards required by health services. Many more large-scale studies, akin to clinical trials, will be required to test the outputs of real-time analysis using mHealth and machine learning algorithms deployed in the real world.<sup>23,28–30,34,42</sup> Although training machine learning models often require a large amount of computing power, the resulting models may be easy to use and can be deployed and run on a mobile phone.

An ideal asthma management system combining machine learning and mHealth would intelligently utilize both active and passive monitoring and be validated with clinical trials. Passive monitoring requires minimal input from the patient, such as wearing a smartwatch or switching on a sleep monitoring device, capturing data without interfering with the patient's daily life. In contrast, active monitoring requires more input from the patient but could provide more detailed information about a person's condition, such as measuring peak flow or answering questions about asthma control. Using machine learning to infer when active monitoring is required based on passive monitoring data would minimize the need for intrusive data collection, while not reducing the attention given to patients.<sup>36,40</sup> Most importantly, systems must be evaluated clinically to ensure clinical (and cost) effectiveness and safety.

## Strengths and Limitations

A reproducible search strategy was implemented using the free search engine PubMed database to search for the latest developments in applications of machine learning algorithms, where the focus was placed only on the past five years. The interdisciplinary team who interpreted the papers consisted of practicing clinicians (covering both primary and secondary care) and applied machine learning experts. However, this is not a systematic review, and it was challenging to directly compare studies and algorithms due to diverse contexts.



## Conclusion

Recent developments in applying machine learning to asthma management have tested a wide range of functionalities using mHealth devices. The algorithms have demonstrated promising results, but they have only been assessed with internal validation at best. Further, the algorithms were mostly developed on small datasets and a select population. Consequently, the likely performance of these algorithms in the general population in a real-world environment is unknown. Future research should include external validation with large sample size and a focus on combining multiple, diverse sources of data.

## Abbreviations

ACT, Asthma Control Test; ACQ, Asthma Control Questionnaire; AUC, area under the ROC curve; BYOT, bring your own technology; COPD, chronic obstructive pulmonary disease; DSP, digital signal processing; FeNO, fractional exhaled nitric oxide; FN, false negative; FP, false positive; GINA, Global Initiative for Asthma; kNN, k-nearest neighbors; LSTM, long short-term memory; mHealth, mobile health; PCA, principal component analysis; PEF, peak expiratory flow; PPG, photoplethysmogram; RCT, randomized control trial; ROC, receiver operating characteristic; SVM, support vector machine; TN, true negative; TP, true positive; VOC, volatile organic compound.

## Acknowledgement

This work is funded by Asthma+Lung UK as part of the Asthma UK Centre for Applied Research [AUK-AC-2018-01]

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Global Asthma Network. The global asthma report 2018. Global Asthma Network; 2018.
2. Reddel HK, Taylor DR, Bateman ED, et al. An Official American Thoracic Society/European Respiratory society statement: asthma control and exacerbations. *Am J Respir Crit Care Med*. 2009;180(1):59–99. doi:10.1164/rccm.200801-060ST
3. World Health Organization. mHealth: new horizons for health through mobile technology. Who Press; 2011. Available from: <http://www.who.int/about/>. Accessed September 3, 2021.
4. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev*. 1959;3(3):210–229. doi:10.1147/rd.33.0210
5. Zhou Y, Zhao L, Zhou N, et al. Predictive big data analytics using the UK biobank data. *Sci Rep*. 2019;9(1):6012. doi:10.1038/s41598-019-41634-y
6. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020;577(7792):706–710. doi:10.1038/s41586-019-1923-7
7. Caravagna G, Giarratano Y, Ramazzotti D, et al. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nat Methods*. 2018;15(9):707–714. doi:10.1038/s41592-018-0108-x
8. Gornale SS, Patravali PU, Manza RR. Detection of osteoarthritis using knee x-ray image analyses: a machine vision based approach. *Int J Comput Appl*. 2016;145(1):20–26. doi:10.5120/ijca2016910544
9. Falcini F, Lami G, Costanza AM. Deep learning in automotive software. *IEEE Softw*. 2017;34(3):56–63. doi:10.1109/MS.2017.79
10. Giarratano Y, Bianchi E, Gray C, et al. Automated segmentation of optical coherence tomography angiography images: benchmark data and clinically relevant metrics. *Transl Vis Sci Technol*. 2020;9(13):5. doi:10.1167/tvst.9.13.5
11. Palaniappan R, Sundaraj K, Ahamed NU. Machine learning in lung sound analysis: a systematic review. *Biocybern Biomed Eng*. 2013;33(3):129–135. doi:10.1016/j.bbe.2013.07.001
12. Li R, Jiang J-Y, Wu X, Hsieh -C-C, Stolcke A. Speaker identification for household scenarios with self-attention and adversarial training. In: Interspeech 2020, ISCA: 2272–2276.
13. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. *J Med Internet Res*. 2017;19(3):e69. doi:10.2196/jmir.7207
14. Hill NR, Ayoubkhani D, McEwan P, et al. Predicting atrial fibrillation in primary care using machine learning. *PLoS One*. 2019;14(11):e0224582. doi:10.1371/JOURNAL.PONE.0224582
15. Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. *PLoS One*. 2012;7(1):e30412. doi:10.1371/journal.pone.0030412
16. Shah SA. Vital sign monitoring and data fusion for paediatric triage. [PhD Thesis]; 2012. Available from: <https://ora.ox.ac.uk/objects/uuu:d:80ae66e3-849b-4df1-b064-f9eb7530200d>. Accessed October 25, 2021.
17. Shah SA, Brown P, Gimeno H, Lin J-P, McClelland VM. Application of machine learning using decision trees for prognosis of deep brain stimulation of globus pallidus internus for children with dystonia. *Front Neurol*. 2020;11:825. doi:10.3389/fneur.2020.00825
18. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med*. 2020;26:1037–1040. doi:10.1038/s41591-020-0916-2

19. Berry SE, Valdes AM, Drew DA, et al. Human postprandial responses to food and potential for precision nutrition. *Nat Med.* **2020**;26(6):964–973. doi:10.1038/s41591-020-0934-0
20. North M, Bourne S, Green B, et al. A randomised controlled feasibility trial of E-health application supported care vs usual care after exacerbation of COPD: the RESCUE trial. *Npj Digit Med.* **2020**. doi:10.1038/s41746-020-00347-7
21. Horne E, Tibble H, Sheikh A, Tsanas A. Challenges of clustering multimodal clinical data: review of applications in asthma subtyping. *JMIR Med Informatics.* **2020**;8(5):e16452. doi:10.2196/16452
22. Honkoop PJ, Simpson A, Bonini M, et al. MyAirCoach: the use of home-monitoring and mHealth systems to predict deterioration in asthma control and the occurrence of asthma exacerbations; study protocol of an observational study. *BMJ Open.* **2017**;7(1):e013935. doi:10.1136/bmjopen-2016-013935
23. Chen A, Zhang J, Zhao L, et al. Machine-learning enabled wireless wearable sensors to study individuality of respiratory behaviors. *Biosens Bioelectron.* **2020**;173:112799. doi:10.1016/j.bios.2020.112799
24. Vatanparvar K, Nemati E, Nathan V, Rahman MM, Kuang J. CoughMatch – subject verification using cough for personal passive health monitoring. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE: **2020**: 5689–5695.
25. Prinable J, Jones P, Boland D, Thamrin C, McEwan A. Derivation of breathing metrics from a photoplethysmogram at rest: machine learning methodology. *JMIR mHealth uHealth.* **2020**;8(7):e13737. doi:10.2196/13737
26. Adhi Pramono RX, Anas Imtiaz S, Rodriguez-Villegas E. Automatic cough detection in acoustic signal using spectral features. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: **2019**: 7153–7156.
27. Chen H, Yuan X, Li J, Pei Z, Zheng X. Automatic multi-level in-exhale segmentation and enhanced generalized S-transform for wheezing detection. *Comput Methods Programs Biomed.* **2019**;178:163–173. doi:10.1016/j.cmpb.2019.06.024
28. Li K, Habre R, Deng H, et al. Applying multivariate segmentation methods to human activity recognition from wearable sensors' data. *JMIR mHealth uHealth.* **2019**;7(2):e11201. doi:10.2196/11201
29. Azam MA, Shahzadi A, Khalid A, Anwar SM, Naeem U. Smartphone based human breath analysis from respiratory sounds. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: **2018**: 445–448.
30. Adhi Pramono RX, Anas Imtiaz S, Rodriguez-Villegas E. Automatic identification of cough events from acoustic signals. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: **2019**: 217–220.
31. Infante C, Chamberlain DB, Kodgule R, Fletcher RR. Classification of voluntary coughs applied to the screening of respiratory disease. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: **2017**: 1413–1416.
32. Taylor TE, Lacalle Muls H, Costello RW, Reilly RB. Estimation of inhalation flow profile using audio-based methods to assess inhaler medication adherence. Kou YR, ed. *PLoS One.* **2018**;13(1):e0191330. doi:10.1371/journal.pone.0191330
33. Purnomo AT, Lin D-B, Adiprabowo T, Hendria WF. Non-contact monitoring and classification of breathing pattern for the supervision of people infected by COVID-19. *Sensors.* **2021**;21(9):3172. doi:10.3390/s21093172
34. Zhang O, Minku LL, Gonen S. Detecting asthma exacerbations using daily home monitoring and machine learning. *J Asthma.* **2021**;58(11):1518–1527. doi:10.1080/02770903.2020.1802746
35. Tsang KCH, Pinnock H, Wilson AM, Ahmar Shah S. Application of machine learning to support self-management of asthma with mHealth. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE: **2020**: 5673–5677. doi:10.1109/EMBC44109.2020.9175679.
36. Tinschert P, Rassouli F, Barata F, et al. Nocturnal cough and sleep quality to assess asthma control and predict attacks. *J Asthma Allergy.* **2020**;13:669–678. doi:10.2147/JAA.S278155
37. Tenero L, Sandri M, Piazza M, Paiola G, Zaffanello M, Piacentini G. Electronic nose in discrimination of children with uncontrolled asthma. *J Breath Res.* **2020**;14(4):046003. doi:10.1088/1752-7163/ab9ab0
38. Finkelstein J, Jeong I. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann N Y Acad Sci.* **2017**;1387(1):153–165. doi:10.1111/nyas.13218
39. Castner J, Jungquist CR, Mammen MJ, Pender JJ, Licata O, Sethi S. Prediction model development of women's daily asthma control using fitness tracker sleep disruption. *Hear Lung.* **2020**;49(5):548–555. doi:10.1016/j.hrtlng.2020.01.013
40. Khasha R, Sepehri MM, Mahdavi SA. An ensemble learning method for asthma control level detection with leveraging medical knowledge-based classifier and supervised learning. *J Med Syst.* **2019**;43(6):158. doi:10.1007/s10916-019-1259-8
41. van Vliet D, Smolinska A, Jöbss Q, et al. Can exhaled volatile organic compounds predict asthma exacerbations in children? *J Breath Res.* **2017**;11(1):016016. doi:10.1088/1752-7163/aa5a8b
42. Huffaker MF, Carchia M, Harris BU, et al. Passive nocturnal physiologic monitoring enables early detection of exacerbations in children with asthma. a proof-of-concept study. *Am J Respir Crit Care Med.* **2018**;198(3):320–328. doi:10.1164/rccm.201712-2606OC
43. Tibble H, Chan A, Mitchell EA, et al. A data-driven typology of asthma medication adherence using cluster analysis. *Sci Rep.* **2020**;10(1):14999. doi:10.1038/s41598-020-72060-0
44. Tignor N, Wang P, Genes N, et al. Methods for clustering time series data acquired from mobile health apps. In: Biocomputing 2017, WORLD SCIENTIFIC: **2017**: 300–311.
45. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR. Development and validation of a questionnaire to measure asthma control. *Eur Respir J.* **1999**;14(4):902–907. doi:10.1034/j.1399-3003.1999.14d29.x
46. Nathan RA, Sorkness CA, Kosinski M, et al. Development of the asthma control test: a survey for assessing asthma control. *J Allergy Clin Immunol.* **2004**;113(1):59–65. doi:10.1016/j.jaci.2003.09.008
47. Adhi Pramono RX, Imtiaz SA, Rodriguez-Villegas E, Cough-Based A. Algorithm for automatic diagnosis of pertussis. *PLoS One.* **2016**;11(9):e0162128. doi:10.1371/journal.pone.0162128
48. Rocha BM, Filos D, Mendes L, et al. A respiratory sound database for the development of automated classification. In: *IFMBE Proceedings*. Vol 66, Singapore: Springer: **2018**: 33–37.
49. Ward JJ. *Rale lung sounds 3.1 professional edition.* *Respir Care.* **2005**;50(10):1385–1388.
50. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL. A public domain dataset for human activity recognition using smartphones. In: 2013 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, (ESANN): **2013**.
51. Bui AAT, Hosseini A, Rocchio R, et al. Biomedical REAL-Time Health Evaluation (BREATHE): toward an mHealth informatics platform. *JAMIA Open.* **2020**;3(2):190–200. doi:10.1093/jamiaopen/ooaa011

52. Atienza T, Aquino T, Fernández M, et al. Budesonide/formoterol maintenance and reliever therapy via turbuhaler versus fixed-dose budesonide/formoterol plus terbutaline in patients with asthma: phase III study results. *Respirology*. 2013;18(2):354–363. doi:10.1111/RESP.12009
53. Chan Y-FY, Wang P, Rogers L, et al. The asthma mobile health study, a large-scale clinical observational study using ResearchKit. *Nat Biotechnol*. 2017;35(4):354–362. doi:10.1038/nbt.3826
54. Chan AHY, Stewart AW, Harrison J, Camargo CA, Black PN, Mitchell EA. The effect of an electronic monitoring device with audiovisual reminder function on adherence to inhaled corticosteroids and school attendance in children with asthma: a randomised controlled trial. *Lancet Respir Med*. 2015;3(3):210–219.
55. Katz S, Arish N, Rokach A, Zaltzman Y, Marcus E-L. the effect of body position on pulmonary function: a systematic review. *BMC Pulm Med*. 2018;18(1):159. doi:10.1186/s12890-018-0723-4
56. Kera T, Maruyama H. The effect of posture on respiratory activity of the abdominal muscles. *J Physiol Anthropol Appl Human Sci*. 2005;24(4):259–265. doi:10.2114/jpa.24.259
57. Penzel T, Möller M, Becker HF, Knaack L, Peter JH. Effect of sleep position and sleep stage on the collapsibility of the upper airways in patients with sleep apnea. *Sleep*. 2001;24(1):90–95. doi:10.1093/sleep/24.1.90
58. Price K, Bird SR, Lythgo N, Raj IS, Wong JYL, Lynch C. Validation of the fitbit one, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *J Med Eng Technol*. 2017;41(3):208–215. doi:10.1080/03091902.2016.1253795
59. Nurmatov UB, Tagiyeva N, Semple S, Devereux G, Sheikh A. Volatile organic compounds and risk of asthma and allergy: a systematic review. *Eur Respir Rev*. 2015;24(135):92–101. doi:10.1183/09059180.00000714
60. Scottish Intercollegiate Guidelines Network/ British Thoracic Society. SIGN 158 British guideline on the management of asthma. BTS/SIGN; 2019. Available from: <https://www.brit-thoracic.org.uk/document-library/guidelines/asthma/btssign-guideline-for-The-management-of-asthma-2019/>. Accessed June 17, 2022.
61. Clark TJ, Hetzel MR. Diurnal Variation of Asthma. *Br J Dis Chest*. 1977;71(2):87–92.
62. Moore VC. Spirometry: step by Step. *Breathe*. 2012;8(3):232–240. doi:10.1183/20734735.0021711
63. Honkoop PJ, Taylor DR, Smith AD, Snoeck-Stroband JB, Sont JK. Early detection of asthma exacerbations by using action points in self-management plans. *Eur Respir J*. 2013;41(1):53–59. doi:10.1183/09031936.00205911
64. Gautier C, Charpin D. Environmental triggers and avoidance in the management of asthma. *J Asthma Allergy*. 2017;10:47–56. doi:10.2147/JAA.S121276
65. Fang W, Zhang Y, Li S, et al. Effects of air pollutant exposure on exacerbation severity in asthma patients with or without reversible airflow obstruction. *J Asthma Allergy*. 2021;14:1117–1127. doi:10.2147/JAA.S328652
66. Baldacci S, Maio S, Cerrai S, et al. Allergy and asthma: effects of the exposure to particulate matter and biological allergens. *Respir Med*. 2015;109(9):1089–1104. doi:10.1016/J.RMED.2015.05.017
67. Global Initiative for Asthma (GINA). Global strategy for asthma management and prevention; 2021. <https://ginasthma.org/gina-reports/>. Accessed June 17, 2021.
68. Pinnock H, Parke HL, Panagioti M, et al. Systematic meta-review of supported self-management for asthma: a healthcare perspective. *BMC Med*. 2017;15(1):64. doi:10.1186/s12916-017-0823-7
69. Pearce G, Parke HL, Pinnock H, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. *J Health Serv Res Policy*. 2016;21(2):73–82. doi:10.1177/1355819615602725
70. Pavord ID, Beasley R, Agusti A, et al. After asthma: redefining airways diseases. *Lancet*. 2018;391(10118):350–400. doi:10.1016/S0140-6736(17
71. Morjaria JB, Polosa R. Recommendation for optimal management of severe refractory asthma. *J Asthma Allergy*. 2010;3:43–56. doi:10.2147/jaa.s6710
72. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338(7713):157–160. doi:10.1136/BMJ.B2393
73. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R A review of challenges and opportunities in machine learning for health. In: AMIA Joint Summits on Translational Science. Vol 2020, American Medical Informatics Association: 2020: 191–200.
74. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform*. 2013;14(1):106. doi:10.1186/1471-2105-14-106
75. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. *JMIR mHealth uHealth*. 2019;7(3):e10828. doi:10.2196/10828
76. VanZeller C, Williams A, Pollock I. Comparison of bench test results measuring the accuracy of peak flow meters. *BMC Pulm Med*. 2019;19(1):74. doi:10.1186/s12890-019-0837-3