

ORIGINAL ARTICLE

Development of childhood asthma prediction models using machine learning approaches

Dilini M. Kothalawala^{1,2}  | Clare S. Murray³  | Angela Simpson³  |
Adnan Custovic⁴  | William J. Tapper¹  | S. Hasan Arshad^{2,5,6}  |
John W. Holloway^{1,2}  | Faisal I. Rezwan^{1,7}  | on behalf of
STELAR/UNICORN investigators

¹Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

²NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Southampton, UK

³Division of Infection, Immunity, and Respiratory Medicine, School of Biological Sciences, University of Manchester, Manchester University Hospital NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

⁴National Heart and Lung Institute, Imperial College of Science, Technology, and Medicine, London, UK

⁵The David Hide Asthma and Allergy Research Centre, St. Mary's Hospital, Isle of Wight, UK

⁶Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

⁷Department of Computer Science, Aberystwyth University, Aberystwyth, UK

Correspondence

John W. Holloway, Southampton General Hospital, Duthie Building, MP808, Southampton, SO16 6YD, UK.
Email: j.w.holloway@soton.ac.uk

Funding information

University of Southampton Presidential Research Studentship; Medical Research Council; NIHR Southampton Biomedical Research Centre; Manchester Biomedical Research Centre

Abstract

Background: Respiratory symptoms are common in early life and often transient. It is difficult to identify in which children these will persist and result in asthma. Machine learning (ML) approaches have the potential for better predictive performance and generalisability over existing childhood asthma prediction models. This study applied ML approaches to predict school-age asthma (age 10) in early life (Childhood Asthma Prediction in Early life, CAPE model) and at preschool age (Childhood Asthma Prediction at Preschool age, CAPP model).

Methods: Clinical and environmental exposure data was collected from children enrolled in the Isle of Wight Birth Cohort ($N = 1368$, ~15% asthma prevalence). Recursive Feature Elimination (RFE) identified an optimal subset of features predictive of school-age asthma for each model. Seven state-of-the-art ML classification algorithms were used to develop prognostic models. Training was performed by applying fivefold cross-validation, imputation, and resampling. Predictive performance was evaluated on the test set. Models were further externally validated in the Manchester Asthma and Allergy Study (MAAS) cohort.

Results: RFE identified eight and twelve predictors for the CAPE and CAPP models, respectively. Support Vector Machine (SVM) algorithms provided the best performance for both the CAPE (area under the receiver operating characteristic curve, $AUC = 0.71$) and CAPP ($AUC = 0.82$) models. Both models demonstrated good generalisability in MAAS (CAPE 8-year = 0.71, 11-year = 0.71, CAPP 8-year = 0.83, 11-year = 0.79) and excellent sensitivity to predict a subgroup of persistent wheezers.

Conclusion: Using ML approaches improved upon the predictive performance of existing regression-based models, with good generalisability and ability to rule in asthma and predict persistent wheeze.

John W. Holloway and Faisal I. Rezwan contributed equally.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Clinical and Translational Allergy published by John Wiley & Sons Ltd on behalf of European Academy of Allergy and Clinical Immunology.

KEYWORDS

asthma, childhood, machine learning, prediction

MOTS-CLÉS

Asthma, Kindheit, maschinelles Lernen, Prognose

1 | INTRODUCTION

Childhood asthma is highly heterogeneous, with numerous factors contributing towards its development, persistence and severity.^{1–3} Despite approximately 80% of asthmatic children developing symptoms (such as wheeze) before the age of six, these clinical symptoms are neither universally present in early life among all future asthmatics nor specific to asthma.⁴ With the added difficulty of making an objective asthma diagnosis before the age of five, both under-treatment and over-treatment of wheezing disorders are common in early life.^{5,6}

The ability to predict the development of school-age asthma can help to identify high-risk preschool children and distinguish them from children whose symptoms are likely to be transient.⁷ Furthermore, early prediction of asthma susceptibility will be critical for the successful implementation of potential primary prevention strategies to reduce the risk of developing asthma.

A recent systematic review identified twenty-one logistic regression-based models for predicting childhood asthma.⁸ However, none of these models have been implemented into standard clinical practice, possibly due to relatively weak predictive power, poor generalisability and need for specialised clinical testing. The review further proposed that regression-based methods for predicting childhood asthma may have been exhausted, with the identified models offering similar predictive power to each other and being unable to be significantly improved upon.⁸

Machine learning approaches have increasingly been applied to a wide range of healthcare problems due to their ability to integrate large quantities of heterogeneous data, handle complex interactions between variables and identify patterns within data.⁹ Particularly for disease prediction, where interactions between biological variables are complex, machine learning approaches have the potential to identify novel predictors which may have been previously overlooked by regression-based approaches.^{9–11} Furthermore, application of methods to reduce model overfitting may address the poor generalisability of existing prediction models in independent populations. Machine learning approaches have shown promise in predicting a variety of clinical asthma outcomes, phenotypes and decisions,^{12–16} including the diagnostic or prognostic prediction of school-age asthma development.^{17–25} While these studies tend to offer improved predictive performance, none of these studies support their findings with external validations of their models or explain how their “black-box” models (where relevant) arrive at their predictions. Without these two components, machine learning models will fail to obtain the trust of physicians and continue to be limited in their clinical utility, regardless of the superior prediction accuracy they may offer.^{26,27}

This study aimed to utilise machine learning approaches to improve upon the performance of traditional regression methods and develop explainable and independently validated prediction models for childhood asthma. Two prognostic prediction models, the Childhood Asthma Prediction in Early-life (CAPE) and Childhood Asthma Prediction at Preschool-age (CAPP) models, were developed to predict school-age asthma at 10 years, within a general population-based cohort, using information available from the first two years and first four years of life, respectively.

2 | METHODS

2.1 | Developmental study population

Data was obtained from 1456 individuals from the Isle of Wight Birth Cohort (IOWBC). Study recruitment and participant details have been previously described²⁸ (see supporting information S1). Ethical approval was obtained from the Isle of Wight Local Research Ethics Committee at recruitment and 1-, 2- and 4-year assessments (No. 05/89) and 10-year assessments including genetic studies (No. 18/98). Prior to participation in the study at each follow-up, written informed consent was obtained from parents of children, and assent from children (where applicable). This study received approval from the University of Southampton Faculty of Medicine ethics committee (ERGO number 46033.R1).

2.2 | Prediction outcome

School-age asthma, evaluated at age 10, was defined as “a doctor diagnosis of asthma ever and at least one episode of wheezing or use of asthma medication in the last 12 months”. Only individuals with a reported asthma status at the 10-year follow-up were included in the analyses ($n = 1368$).

2.3 | Candidate predictors

Fifty-four candidate predictors previously reported to be associated with childhood asthma, and for which data was available in the IOWBC, were identified (Table E1). Candidate predictors included data on subject demographics, lifestyle, clinical symptoms of allergy and asthma and environmental exposures collected across three time points: at birth (prenatal and perinatal data), early life (combined

exposure at either the 1-year or the 2-year follow-ups) and at pre-school age (4-year follow-up).

2.4 | Model development

All stages of model development were performed independently for the CAPE and CAPP models (Figure 1).

2.5 | Feature selection

For each model, feature selection was performed on the complete dataset for all available candidate predictors (without any missing

values) using Recursive Feature Elimination (RFE) with a random forest algorithm, using fivefold cross-validation (see supporting information S1).

2.6 | Model construction and optimisation

To identify the best classification algorithm, seven machine learning classifiers were implemented: two support vector machines (SVM) (linear and radial basis (RBF) kernel functions), decision tree, random forest, naive Bayes, multilayer perceptron, and K-Nearest Neighbours (see supporting information S1).

Each machine learning algorithm was initially trained and evaluated on the subset of individuals who had complete data for the

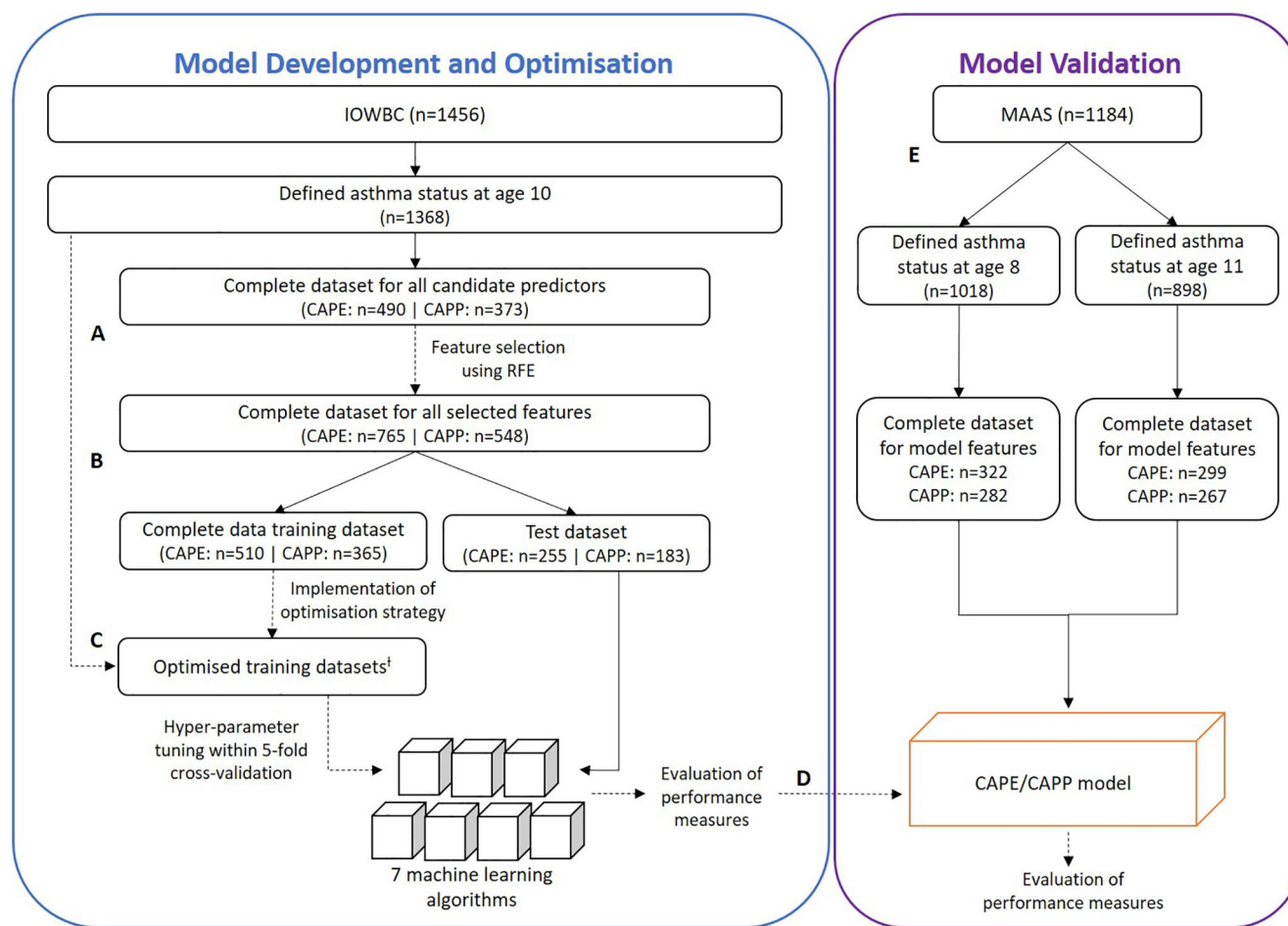


FIGURE 1 Workflow for the development and validation of asthma prediction models using machine learning approaches. Model development in the Isle of Wight Birth Cohort (IOWBC) was performed independently in for the construction of the CAPE and CAPP tools. (A) Feature selection was performed using only individuals with complete data for all candidate predictors. (B, C) Seven machine learning classifiers (two support vector machines with different kernel functions (linear and radial basis function), naive Bayes classifier, decision tree, multilayer perceptron, random forest and K-nearest neighbours) were developed. Models were developed using complete data for the subset of features identified from feature selection (B), and subsequently redeveloped using optimised training datasets (C). Training dataset optimisation consisted of the step-wise application of imputation and resampling (oversampling using ADASYN and random undersampling) to the entire IOWBC dataset not allocated to the test dataset, including those with missing predictor data (CAPE: $n = 1113$; CAPP: $n = 1185$). (D) The best CAPE and CAPP models were selected based on performance in the test set. (E) Selected models were externally validated to predict school-age asthma at ages 8 and 11 years in an independent population (Manchester Asthma and Allergy Study, MAAS). [†]The performance of the best CAPE model was developed on the complete training dataset, undersampled to balance class proportions ($n = 136$). The best CAPP model was developed on the complete training dataset, with cases oversampled by 300% and controls undersampled to balance class proportions ($n = 408$)

predictors selected through RFE. The dataset was split (ratio of 2:1, preserving class proportions) into a training and holdout test set for model development and validation, respectively (Figure 1). Within a fivefold cross-validation, the hyperparameters for each model were tuned using a grid search, optimising for its balanced accuracy (see supporting information S1, Table E2).

The training dataset was then optimised to further improve the performance of the classification algorithms. Multiple imputation using Multivariate Imputation by Chain Equations (MICE),²⁹ over-sampling using an adaptive synthetic sampling approach (ADASYN),³⁰ and random under-sampling were implemented in a stepwise approach to address the degree of missing data and class imbalance in the training set (see supporting information S1). The seven algorithms were then redeveloped, with hyperparameters retuned, on each optimised training set to identify the best asthma prediction model(s) and tested on the same holdout test set (Figure 1).

The best CAPE and CAPP models were selected based on their discriminative performance on the test set using the area under the receiver operating characteristics curve (AUC). Sensitivity, specificity, positive and negative predictive values (PPV and NPV), positive and negative likelihood ratios (LR+ and LR–), balanced accuracy, F1-score and Brier score were reported at the optimal threshold that maximized the Youden's Index, with 2000 bootstrap samples used to calculate 95% confidence intervals for the performance measures.

2.7 | External validation

The best performing models were validated in the Manchester Asthma and Allergy Study (MAAS) cohort³¹ to predict school-age asthma at ages eight and eleven (Figure 1, see supporting information S1). Data extracted from MAAS was closely matched to maximise the similarity of predictor and outcome definitions used in the development cohort (Table E3).

2.8 | Sensitivity analyses

Sensitivity analyses were conducted to comprehensively evaluate the developed models, including evaluations of (i) their generalisability in high risk subgroups; (ii) their robustness to predict an alternative definition of school-age asthma; (iii) the resolution of the predictions to distinguish between individuals presenting with distinct wheeze phenotypes throughout childhood and adolescence; and (iv) their performance compared to similar regression-based models (see supporting information S1).

2.9 | Explaining the 'black-box' models

SHapley Additive exPlanations (SHAP)³² were used to evaluate feature importance and provide global explanations for how predictions were made by the CAPE and CAPP models (see

supporting information S1). Examples of how SHAP can be used locally to explain individual predictions were also provided.

3 | RESULTS

In the IOWBC, 1368 enrolled participants had a defined asthma outcome at age 10, of whom 201 (14.69%) were asthmatic. Baseline characteristics between individuals with complete data were largely comparable with the full IOWBC dataset (Table E4).

3.1 | Childhood Asthma Prediction in Early-life (CAPE) model

Complete data on all 39 predictors collected by age two was available for 490 individuals. RFE identified a subset of eight predictors for inclusion in the CAPE model, with an average balanced accuracy of 64.49%. Figure 2A details the feature importance, direction, and magnitude of asthma risk for each selected predictor based on SHAP. Complete data for these eight predictors was available for 765 individuals; 510 (68 asthmatics) and 255 (34 asthmatics) individuals were allocated to the initial training and test sets, respectively. An SVM classifier (RBF kernel) was the best performing classification algorithm for the CAPE model (AUC = 0.71, Brier score = 0.21) (Table 1A).

3.2 | External validation of the CAPE model

To predict the development of asthma at the 8-year and 11-year time-points in MAAS, complete data on the eight CAPE predictors was available for 322 and 299 individuals, respectively. Table E5 compares the distribution of predictors in the IOWBC and MAAS. The CAPE model demonstrated moderate generalisability, maintaining an AUC = 0.71 at both 8 and 11 years (Table 1A; Figure 3), despite slight reductions in PPV. In the high-risk subgroups, despite a 3%–4% increase in PPV, overall predictive performance decreased (Table 1A).

3.3 | Childhood Asthma Prediction at Preschool-age (CAPP) model

For the CAPP model, 373 individuals had complete data for all 54 candidate predictors available by age four. RFE identified an optimal subset of 12 predictors for inclusion in the model, with an average balanced accuracy of 74.93% (Figure 2B). Complete data for these 12 predictors was available for 548 individuals, of whom 365 (51 asthmatics) and 183 (25 asthmatics) individuals were assigned to the initial training and test sets, respectively. The best performing classification algorithm for the CAPP model was an SVM (linear kernel) classifier (AUC = 0.82, Brier score = 0.18) (Table 1B).

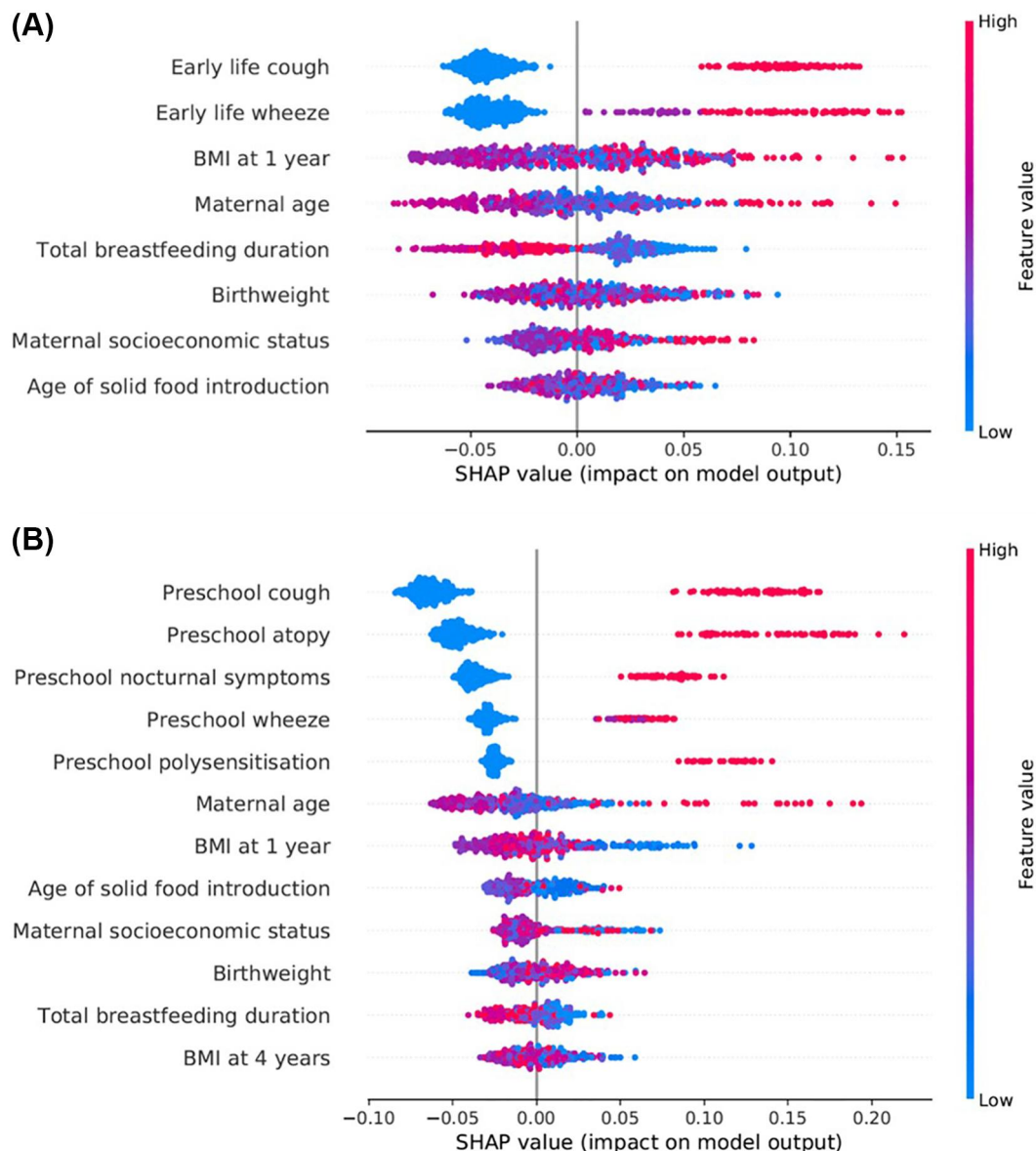


FIGURE 2 SHAP feature explanations for the CAPE and CAPP models. The SHAP summary plots describe the contribution of the features selected during RFE for inclusion in the CAPE (A) and CAPP models (B). Predictors are listed in descending order of their SHAP value. The higher the SHAP value, the larger its contribution on model predictions. Each dot in each predictor row corresponds to a separate individual. The placing of the dot along the x-axis represents the contribution of the predictor in the individual's asthma prediction. The colour of the dot refers to the feature value, with higher values in red and lower values in blue. For example, early life cough offers the highest contribution to the random forest model, with higher values (presence of early life cough) having a positive contribution towards a prediction of asthma. The absence of early life cough (blue dots) reduces the impact/contribution of the model delivering a prediction of asthma

3.4 | External validation of the CAPP model

For validation of the CAPP model in MAAS at the 8-year and 11-year time-points, complete data for the 12 CAPP predictors was available for 282 and 267 individuals, respectively. The model demonstrated good generalisability to predict asthma at both 8 and 11 years ($AUC = 0.83$ and 0.79 , respectively) in the unselected MAAS subgroup (Table 1B; Figure 3). PPV also remained comparable in MAAS ($PPV = 0.45$ and 0.41 , respectively), with further improvements reported in the high-risk subgroup validations at both time-points (Table 1B).

3.5 | Sensitivity analysis

The CAPE and CAPP models were robust in correctly predicting non-asthmatics using the alternative asthma definition (similar NPV). However, neither model was robust in predicting asthmatics, with an increase in false positive predictions reducing the PPV by approximately 50% for both models, likely due to disagreement between the original and modified asthma definitions (Table E6; Figure E1).

Furthermore, both models showed excellent power to predict a persistent wheeze phenotype, with 100% and 90% of individuals with persistent wheeze offered a positive prediction by the CAPE and

TABLE 1 Performance of the CAPE and CAPP models

A. Performance of the CAPE model											
	Dataset	Sample size (no. asthmatic)	Balanced accuracy	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-	F1 score
Development—IOWBC: 10 years	Training ^a	136 (68 asthmatic)	0.65 (0.57, 0.74)	0.76 (0.68, 0.84)	0.56 (0.44, 0.68)	0.75 (0.63, 0.85)	0.69 (0.59, 0.79)	0.63 (0.56, 0.71)	2.24 (1.44, 3.80)	0.59 (0.42, 0.78)	0.62 (0.52, 0.71)
	Testing	255 (34 asthmatic)	0.71 (0.62, 0.78)	0.71 (0.61, 0.80)	0.74 (0.56, 0.88)	0.68 (0.62, 0.74)	0.26 (0.21, 0.32)	0.94 (0.91, 0.97)	2.29 (1.69, 3.01)	0.39 (0.18, 0.63)	0.38 (0.31, 0.46)
External Validation —MAAS: 8 years	Unselected	322 (38 asthmatic)	0.67 (0.60, 0.74)	0.71 (0.63, 0.79)	0.84 (0.71, 0.95)	0.51 (0.45, 0.56)	0.19 (0.16, 0.21)	0.96 (0.93, 0.99)	1.71 (1.40, 2.03)	0.31 (0.10, 0.57)	0.30 (0.26, 0.35)
	High-risk ^c	208 (31 asthmatic)	0.66 (0.59, 0.73)	0.71 (0.61, 0.80)	0.87 (0.74, 0.97)	0.46 (0.39, 0.53)	0.22 (0.19, 0.25)	0.95 (0.91, 0.99)	1.61 (1.31, 1.95)	0.28 (0.06, 0.59)	0.35 (0.30, 0.40)
	High-risk ^d	81 (16 asthmatic)	0.57 (0.45, 0.67)	0.64 (0.47, 0.80)	0.81 (0.63, 1.00)	0.32 (0.22, 0.43)	0.23 (0.18, 0.28)	0.88 (0.75, 1.00)	1.20 (0.86, 1.56)	0.58 (0.00, 1.35)	0.36 (0.27, 0.43)
External Validation —MAAS: 11 years	Unselected	299 (32 asthmatic)	0.68 (0.60, 0.74)	0.71 (0.62, 0.79)	0.84 (0.72, 0.97)	0.51 (0.45, 0.57)	0.17 (0.14, 0.20)	0.96 (0.94, 0.99)	1.72 (1.39, 2.05)	0.31 (0.07, 0.58)	0.28 (0.24, 0.33)
	High-risk ^c	192 (25 asthmatic)	0.67 (0.59, 0.74)	0.71 (0.62, 0.80)	0.88 (0.76, 1.00)	0.47 (0.40, 0.54)	0.20 (0.17, 0.23)	0.96 (0.92, 1.00)	1.65 (1.34, 2.03)	0.26 (0.00, 0.57)	0.32 (0.27, 0.37)
	High-risk ^d	72 (12 asthmatic)	0.58 (0.44, 0.69)	0.60 (0.43, 0.76)	0.83 (0.58, 1.00)	0.33 (0.22, 0.45)	0.20 (0.15, 0.25)	0.91 (0.78, 1.00)	1.25 (0.85, 1.66)	0.50 (0.00, 1.39)	0.32 (0.23, 0.40)
B. Performance of the CAPP model											
	Population	Sample size (no. asthmatic)	Balanced accuracy	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-	F1 score
Development—IOWBC: 10 years	Training ^b	408 (204 asthmatic)	0.78 (0.75, 0.82)	0.85 (0.81, 0.89)	0.80 (0.74, 0.82)	0.77 (0.71, 0.85)	0.78 (0.73, 0.82)	0.79 (0.75, 0.84)	3.47 (2.73, 4.61)	0.26 (0.19, 0.34)	0.79 (0.75, 0.83)
	Testing	183 (25 asthmatic)	0.80 (0.70, 0.89)	0.82 (0.71, 0.91)	0.72 (0.52, 0.88)	0.88 (0.83, 0.92)	0.47 (0.38, 0.62)	0.95 (0.92, 0.98)	5.99 (3.79, 10.11)	0.32 (0.13, 0.54)	0.56 (0.45, 0.70)
External Validation—MAAS: 8 years	Unselected	282 (33 asthmatic)	0.73 (0.64, 0.81)	0.83 (0.75, 0.90)	0.55 (0.36, 0.70)	0.91 (0.88, 0.95)	0.45 (0.33, 0.59)	0.94 (0.92, 0.96)	6.17 (3.64, 10.69)	0.50 (0.33, 0.69)	0.49 (0.36, 0.62)
	High-risk ^c	178 (26 asthmatic)	0.70 (0.60, 0.80)	0.80 (0.70, 0.88)	0.50 (0.31, 0.69)	0.90 (0.85, 0.95)	0.46 (0.32, 0.63)	0.91 (0.89, 0.94)	5.07 (2.77, 9.95)	0.55 (0.34, 0.77)	0.48 (0.32, 0.63)
	High-risk ^d	70 (13 asthmatic)	0.73 (0.59, 0.87)	0.80 (0.62, 0.94)	0.54 (0.23, 0.77)	0.93 (0.86, 0.98)	0.64 (0.40, 0.90)	0.90 (0.84, 0.95)	7.67 (2.92, 39.46)	0.50 (0.24, 0.81)	0.58 (0.32, 0.78)
External Validation—MAAS: 11 years	Unselected	267 (29 asthmatic)	0.73 (0.63, 0.82)	0.79 (0.68, 0.88)	0.55 (0.38, 0.72)	0.90 (0.87, 0.94)	0.41 (0.29, 0.55)	0.94 (0.92, 0.96)	5.71 (3.44, 9.85)	0.50 (0.30, 0.71)	0.47 (0.33, 0.62)
	High-risk ^c	169 (22 asthmatic)	0.72 (0.61, 0.82)	0.76 (0.61, 0.88)	0.55 (0.36, 0.73)	0.89 (0.84, 0.94)	0.43 (0.29, 0.59)	0.93 (0.90, 0.96)	5.01 (2.75, 9.65)	0.51 (0.30, 0.74)	0.48 (0.32, 0.63)
	High-risk ^d	64 (10 asthmatic)	0.75 (0.59, 0.90)	0.73 (0.47, 0.94)	0.60 (0.30, 0.90)	0.91 (0.83, 0.98)	0.55 (0.31, 0.86)	0.92 (0.87, 0.98)	6.48 (2.40, 32.40)	0.44 (0.11, 0.80)	0.57 (0.30, 0.78)

Note: Performance measures are reported, with 95% confidence intervals from 2000 bootstrap samples in brackets. Performance measures in the IOWBC test set and MAAS are evaluated at thresholds of 0.42 (CAPE model) and 0.73 (CAPP model).

^aThe CAPE model was developed using an SVM classification algorithm using a radial basis function kernel ($C = 45.1$, $\gamma = 0.0054$). The model was trained on the training dataset consisting of individuals with complete data, with controls under-sampled to obtain a 1:1 class ratio.

^bThe CAPP model was developed using an SVM classification algorithm using a linear kernel ($C = 0.33$), and was trained on the training dataset consisting of individuals with complete data, with cases oversampled by 300% and controls under-sampled to obtain a 1:1 class ratio.

^cHigh-risk, defined as a child having at least one parent with allergic disease (asthma, eczema or allergic rhinitis).

^dHigh-risk, defined as a child with both parents with allergic disease (asthma, eczema or allergic rhinitis).

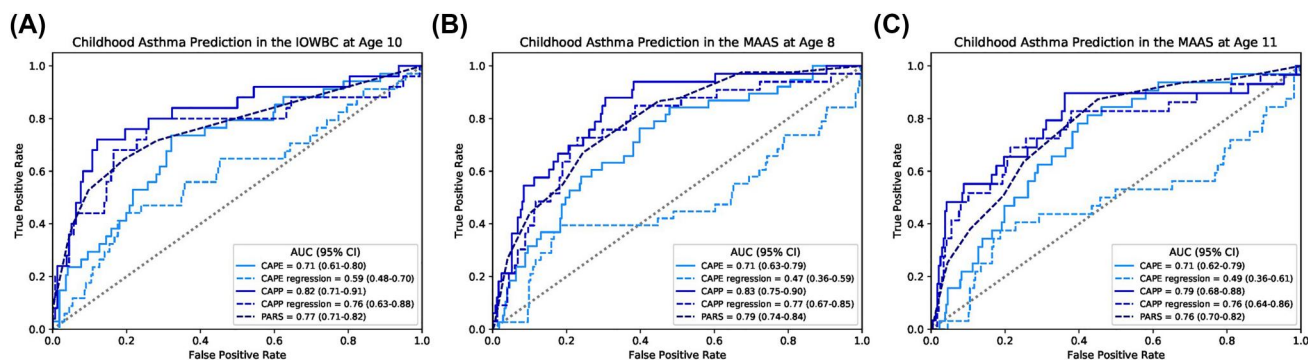


FIGURE 3 Discriminative performance of the CAPE and CAPP machine learning models. The ROC curves compare the discriminative performance of the CAPE and CAPP machine learning models, their equivalent logistic regression models and the PARS model in the IOWBC at age 10 (A) and upon validation in MAAS at age 8 years (B) and 11 years (C)

CAPP models in the IOWBC, respectively (90% and 57% in MAAS, respectively) (Figure 4).

3.6 | Comparison with regression methods

Both the CAPE and CAPP models outperformed their equivalent logistic regression models (Table E7; Figure 3). There was a substantial decline in predictive performance of the CAPE-logistic regression model (AUC = 0.59 vs. 0.71), with predictions being no better than chance in MAAS at 8 and 11 years (AUC = 0.47 and 0.49, respectively). Predictive power of the CAPP-logistic regression model was also lower compared to the CAPP-machine learning model (AUC = 0.76 vs. 0.82, PPV = 0.33 vs. 0.47).

Whilst the benchmark regression-based model for the CAPE model (Persistent Asthma Predictive Score)³³ was unable to be replicated due to lack of data on key predictors in the IOWBC, the model comparable with the CAPP model, PARS (Paediatric Asthma Risk Score),³⁴ was replicated in the IOWBC and MAAS (AUC in IOWBC = 0.77, MAAS 8-year = 0.79, MAAS 11-year = 0.76). Among individuals with predictions available for both the CAPP and PARS models, positive net reclassification indices show that the proportion of reclassifications made by the CAPP model offered equal, if not greater, accuracy to predict future asthmatics than PARS in both the IOWBC (Table 2) and MAAS (Table E8).

3.7 | Explaining the “black-box” models

Based on SHAP, only a subset of predictors included in each model were shown to have a major contribution on the predictions—early life cough and wheeze for the CAPE model and preschool cough, atopy and polysensitisation for the CAPP model (Figure E2). The contributions of these predictors were consistent with explanations of individual predictions (Figure E3). Redevelopment of the models including only these highly contributing predictors showed similar performance for the CAPP model but a 10% fall in AUC for the CAPE model (Figure E4).

4 | DISCUSSION

4.1 | Summary of findings

Two models, predicting school-age asthma at age 10 within a general population, were developed using machine learning classification methods. The CAPE model uses a RBF SVM classifier and eight predictors to predict school-age asthma in early life. The CAPP model uses a linear SVM classifier and twelve predictors available by age four. Both machine learning models offered superior predictive power and generalisability upon external validation compared to equivalent models developed using logistic regression methods as well as existing regression-based models. Whilst the primary prediction outcome was school-age asthma, both models demonstrated excellent sensitivity in predicting individuals likely to experience persistent wheeze throughout childhood.

4.2 | Comparisons with existing models

To date, twenty-one regression-based prediction models have been developed for childhood asthma (reviewed in Kothalawala et al.⁸), of which only six have been externally validated (Table E9). A recent systematic review further identified 10 studies that developed prediction models for childhood asthma using machine learning approaches, but only eight specifically predicted school-age asthma (5–14 years).²⁶ Another study directly compared the performance of a current regression-based asthma prediction model, PARS, with a conditional inference tree-based decision rule model using the same predictors.²⁵ However, none of these studies externally validated the machine learning models they proposed.

Similar to the CAPE and CAPP models, most published asthma prediction models are very good at ruling out asthma rather than ruling in asthma, resulting partly from low power due to low asthma prevalence.⁸ Even if existing models offer good PPV, this often degrades upon validation.⁸ Indeed, despite having similar asthma prevalence to existing studies in the original training set, the machine learning-based CAPP model offered a 30% improvement in

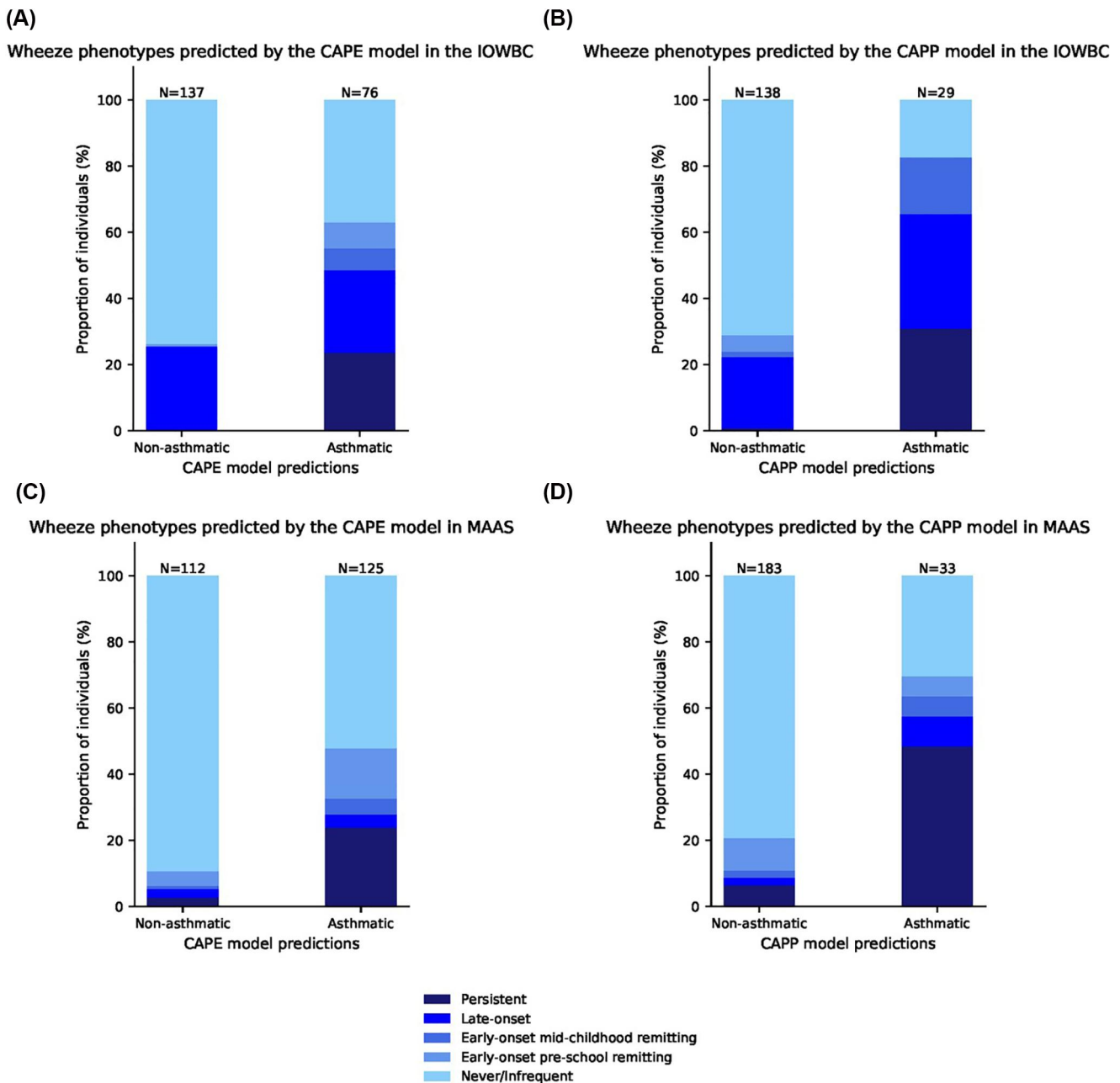


FIGURE 4 CAPE and CAPP model predictions and corresponding wheeze trajectories. The proportion of individuals corresponding to their most probable wheeze phenotype is presented for those predicted to be asthmatic or non-asthmatic by the CAPE and CAPP models in the IOWBC (A, B) and MAAS (C, D)

sensitivity compared to the only model suggested in asthma guidelines described to date (sensitivity: CAPP = 0.72 vs. loose Asthma Predictive Index = 0.42)³⁵ and further 10% improvement in PPV compared to its benchmark model, PARS. This is consistent with Owora et al.'s novel tree-based model offering better predictive performance compared to an equivalent regression-based PARS model (AUC = 0.85 vs. 0.71).²⁵ Many of the other machine learning models also demonstrated greater performance to predict asthma than existing regression-based models.²⁶ However, with low sample sizes and indications of overfitting in many of these studies, the lack of external validation renders it impossible to evaluate any superior

performance offered by these models, especially since they were all developed in high-risk populations. Importantly, our CAPP machine learning model was more generalisable and retained its positive predictive power upon replication compared to its equivalent logistic regression model. Furthermore, reclassification tables comparing the CAPP and PARS models were suggestive of the CAPP model predicting future asthmatics more accurately than PARS, with a greater proportion of correct reclassifications than incorrect reclassification made by the CAPP model in both the IOWBC and MAAS. However, this needs to be confirmed within a larger cohort. The moderate but limited predictive power of the CAPE model compared to the CAPP

TABLE 2 Reclassification table showing changes in prediction categorisation between the PARS and CAPP model

Predicted risk (PARS model)	Predicted risk (CAPP model)			Reclassified by CAPP (%)			
	No asthma	Asthma	Total	Increased risk	Decreased risk	Correctly reclassified	NRI
No asthma at age 10 (n = 149)							
No asthma	130	9^b	139				
Asthma	1^a	9	10	9 (6%)	1 (<1%)	1 (<1%)	−0.05
Total	131	18	149				
Asthma at age 10 (n = 25)							
No asthma	7	8^a	15				
Asthma	0^b	10	10	8 (32%)	0 (0%)	8 (32%)	0.32
Total	7	18	25				
			Total	17	1	9	

Note: Reclassification table comparing the change in individual asthma predictions with the CAPP model instead of the PARS model (reference model). For the PARS model, categorisations of predictions as asthmatic and non-asthmatic was based on the optimal threshold (cutoff = 7) as defined in their original publication. Results are presented separately for individuals who were asthmatic and non-asthmatic at age 10. Values in bold identify the number of individuals with disagreement in their asthma predictions made by the CAPP and PARS models. NRI = net reclassification index is given separately for true asthmatics and non-asthmatics.

^aReclassified into a more appropriate risk group by the CAPP model with respect to classifications made by the PARS model.

^bReclassified into a less appropriate risk group by the CAPP model with respect to classifications made by the PARS model.

model was unsurprising given the known difficulty of predicting the future development of childhood asthma using data from the first few years of life.³⁶

4.3 | Predictor selection and availability

Both the CAPE and CAPP models include data collected across multiple time-points (Figures E5 and E6). Given the variable nature of asthma development and risk throughout early childhood, the consideration of predictors across multiple time-points allowed for the identification of novel combinations of predictors that together improved the ability of the models to predict asthma. Whilst data collected across multiple time-points may hinder the utility of the models, the selected predictors are all typically reported during routine health visits or tracked in child health records. Only the predictors of atopy and polysensitisation, which require a skin prick test (SPT), may restrict the applicability of the CAPP model in primary care. However, as these predictors are well-established in the literature, were shown to make large contributions to the predictions (Figure E2), and resulted in a 10% reduction in AUC when excluded from the model (Table E10), the predictive benefit offered by the inclusion of sensitisation was deemed to outweigh the potential reduction in applicability.

Of the predictors selected for inclusion in the two models, some were well-established risk factors with a clear inferred direction of asthma risk (Figure 2). Others were predictors which have not previously been used in childhood asthma prediction models (maternal age at the time of the child's birth, age of solid food introduction and total breastfeeding duration) and offer a less clear direction of

asthma risk. The selection of these novel predictors, over others that have a more established biological relevance in the literature (such as parental asthma, eczema or allergic rhinitis), may be cautiously accepted by the clinical community. However, RFE identifies the subset of features that collectively offer the best predictive accuracy rather than devising a comprehensive list of childhood asthma risk factors, which may be biologically sound but lacking in predictive power.³⁷ In fact, the predictors of wheeze and cough were among those repeatedly included in the majority of machine learning models identified to date.²⁶ The predictors of atopy, polysensitisation and wheeze were also included in Owora et al.'s machine learning model, however the predictors were taken from the PARS model rather than being identified from an independent feature selection.²⁵ It is also important to acknowledge the possibility that the selection of these novel predictors may stem from an inherent bias of the random forest algorithm to assign greater importance to features which are continuous or which have a large number of categories.³⁸ However, as the CAPE and CAPP models developed using these selected predictors demonstrated improved performance against existing prediction models, any bias stemming from the feature selection process did not appear to limit the inclusion of features that were truly predictive of school-age asthma.

4.4 | Prediction generalisability, robustness and resolution

In the unselected MAAS cohort, the CAPE and CAPP models showed moderate to good generalisability to predict asthma across school ages, despite the marginal decline in the PPV of the CAPE model.

Validation in high-risk MAAS subgroups showed the PPV of both models to increase with the number of allergic parents, suggesting that confidence in ruling in asthma improves in high-risk groups; but replication in a larger study population is required.

The lack of a clear definition for asthma is an unavoidable limitation in epidemiological studies.³⁹ The asthma definition used in this study aimed to account for children with a clinical indication of asthma (physician diagnosed) who were actively symptomatic, but also those potentially asymptomatic at the time of assessment due to the use of symptom relieving medications. Whilst both models were robust in predicting non-asthmatics using an alternative asthma definition of wheeze and bronchial hyper-responsiveness (BHR), they had reduced power to predict true asthmatics (~50% decline in PPV). The latter may be explained by objective tests, such as spirometry and BHR, being influenced by treatment; potential asthmatics on controller medications, whom the models are trained to identify as asthmatic, may be considered as non-asthmatic with the alternative definition, resulting in greater false positive predictions.

As the aim of this study was to compare whether machine learning approaches could improve upon existing regression-based models that predict childhood asthma, the primary prediction outcome for this study was restricted to school-age asthma rather than predicting asthma phenotypes. However, acknowledging the importance of exploring specific sub-phenotypes of asthma, the resolution of the machine learning models to predict an individual's future wheeze trajectory was explored. Notably, both the CAPE and CAPP models showed excellent sensitivity to predict individuals with a persistent wheeze phenotype; these individuals would likely benefit from early primary or secondary asthma prevention/management.

To promote the clinical use of complex machine learning methods, studies must address the major hurdle of model interpretability. This study demonstrates how tools such as SHAP values³² can be used to unravel explanations of complex black-box machine learning algorithms that have shown to improve the accuracy of childhood asthma predictions.

4.5 | Strengths and limitations

This study had a number of strengths. First, each model was developed to make timely predictions to identify future asthmatics within a general population, rather than among those already considered at high-risk (mainly those experiencing wheeze or with a familial history of asthma/allergy). Second, by utilising machine learning methods, novel predictor subsets for school-age asthma were identified and the developed models offered improved predictive performance over current regression-based methods. Third, to our knowledge, this is the first study to externally validate asthma prediction models developed using machine learning approaches. The models demonstrated good generalisability to predict school-age asthma across multiple time-points, without degrading the predictive power to rule in asthma (particularly with the CAPP model). Fourth, the two models displayed excellent sensitivity to predict a subgroup of individuals

with persistent wheeze. Finally, this study was able to use SHAP to address one of the key issues preventing the uptake of machine learning methods in clinical practice—the inability to interpret the models and explain the individual predictions made.

However, this study was limited by both model development and validation being conducted in predominantly Caucasian populations. Machine learning also requires large datasets—the introduction of more data would undoubtedly improve the performance of the machine learning models and offer more precise performance estimates with smaller confidence intervals. To retain a sample size appropriate for machine learning, feature selection was conducted before performing a train-test split. This decision could have resulted in information leakage, potentially biasing the performance seen in the IOWBC test sets. To mitigate any bias, external replication was used to evaluate the models; as performance in MAAS was similar to the IOWBC, data leakage was not deemed a significant problem. Finally, whilst genomic data was available in the IOWBC, only clinical and environmental predictors were considered in order to maximize the clinical applicability of the models. It is possible that the consideration of genomic predictors might significantly improve childhood asthma predictions further^{22,40}; however, the aim of this study was to explore whether machine learning methods could surpass the predictive ceiling that existing logistic regression methods appeared to be limited to. Hence, to provide a fair comparison with existing regression-based models, such asthma biomarkers were not incorporated into this study.

5 | CONCLUSION AND FUTURE WORK

Using machine learning, the CAPE and CAPP models were able to surpass the predictive performance of similar models developed using traditional logistic regression-based methods. Both models were generalisable in an independent population, with the CAPP model also demonstrating superior predictive power to rule in true asthmatics compared to its benchmark model (and was retained upon validation). Future application of these models could include the development of a personalised tool/app capable of providing explanations of which predictors contributed to an individual's predicted probability of developing asthma. Both models also demonstrated excellent sensitivity to predict a subgroup of persistent wheezers. Therefore, rather than developing an all-encompassing asthma prediction tool, further research into predicting specific 'asthmas' using machine learning approaches may offer greater predictive insight and clinical utility. Finally, continued exploration of machine learning approaches and the identification and integration of novel biomarkers is warranted to further improve the power to predict future childhood asthma.

ACKNOWLEDGMENTS

The authors would like to acknowledge the help of all the staff at the David Hide Asthma and Allergy Research Centre in undertaking the assessments of the Isle of Wight Birth Cohort. The authors would

also like to thank the IOWBC and MAAS study participants and their parents for their continued support and enthusiasm. Recruitment and initial assessment for the first 4 years of age for the IOWBC was supported by the Isle of Wight Health Authority. The 10-year follow-up of the IOWBC was funded by the National Asthma Campaign, UK (Grant No 364). MAAS was supported by the Asthma UK Grants No 301 (1995–1998), No 362 (1998–2001), No 01/012 (2001–2004), No 04/014 (2004–2007), BMA James Trust (2005) and The JP Moulton Charitable Foundation (2004–current), The North west Lung Centre Charity (1997–current) and the Medical Research Council (MRC) G0601361 (2007–2012), MR/K002449/1 (2013–2014) and MR/L012693/1 (2014–2018). UNICORN (Unified Cohorts Research Network): Disaggregating asthma MR/S025340/1. The authors would also like to acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

This work was supported by the National Institute for Health Research through the NIHR Southampton Biomedical Research Centre and a University of Southampton Presidential Research Studentship. Replication analysis in MAAS was supported by the Medical Research Council as part of UNICORN (Unified Cohorts Research Network): Disaggregating asthma MR/S025340/1. Angela Simpson and Clare Murray are supported by the NIHR Manchester Biomedical Research Centre.

CONFLICT OF INTEREST

The authors have no funding relationships or conflicts of interest related to this article to disclose.

AUTHOR CONTRIBUTIONS

Conceptualization and design: Dilini M. Kothalawala, S. Hasan Arshad, Faisal I. Rezwan and John W. Holloway. *Data acquisition:* John W. Holloway, Angela Simpson, Adnan Custovic, S. Hasan Arshad and STELAR/UNICORN investigators. *Analysis and interpretation:* Dilini M. Kothalawala. *Manuscript draft:* Dilini M. Kothalawala. *Manuscript revision and final approval:* all authors.

STELAR/UNICORN INVESTIGATORS

Graham C Roberts, Human Development and Health, Faculty of Medicine, University of Southampton, David Hide Asthma and Allergy Research Centre, Isle of Wight and NIHR Southampton Biomedical Research Centre, University Hospitals Southampton NHS Foundation Trust, Southampton, UK.

Steve W Turner, Child Health, University of Aberdeen, Aberdeen, UK.

Raquel Granell, MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, UK.

Sadia Haider, National Heart and Lung Institute, Imperial College London, UK.

Sara Fontanella, National Heart and Lung Institute, Imperial College London, UK.

Paul Cullinan, National Heart and Lung Institute, Imperial College London, UK.

ORCID

Dilini M. Kothalawala  <https://orcid.org/0000-0002-5804-0457>

Clare S. Murray  <https://orcid.org/0000-0002-8961-8055>

Angela Simpson  <https://orcid.org/0000-0003-2733-6666>

Adnan Custovic  <https://orcid.org/0000-0001-5218-7071>

William J. Tapper  <https://orcid.org/0000-0002-5896-1889>

S. Hasan Arshad  <https://orcid.org/0000-0001-5988-235X>

John W. Holloway  <https://orcid.org/0000-0001-9998-0464>

Faisal I. Rezwan  <https://orcid.org/0000-0001-9921-222X>

REFERENCES

1. Akdis CA, Agache I. *Global Atlas of Asthma*. European Academy of Allergy and Clinical Immunology; 2013.
2. Licari A, Castagnoli R, Brambilla I, et al. Asthma endotyping and biomarkers in childhood asthma. *Pediatr Allergy Immunol Pulmonol*. 2018;31:44–55.
3. Pavord ID, Beasley R, Agusti A, et al. After asthma: redefining airways diseases. *Lancet*. 2018;391:350–400.
4. Ullmann N, Mirra V, Di Marco A, et al. Asthma: differential diagnosis and comorbidities. *Front Pediatr*. 2018;6:276.
5. Isabella Annesi-Maesano CS, Caillaud D, de Blay F, Lavaud F, Charpin D, Raherison C. Factors related to under-diagnosis and undertreatment of childhood asthma in metropolitan France. *Multi-discip Respir Med*. 2012;7.
6. Bush A, Fleming L, Saglani S. Severe asthma in children. *Respirology*. 2017;22:886–897.
7. Martinez FD, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. *N Engl J Med*. 1995;332:133–138.
8. Kothalawala DM, Kadalayil L, Weiss VBN, et al. Prediction models for childhood asthma: a systematic review. *Pediatr Allergy Immunol*. 2020;31:616–627.
9. James G, Witten D, Hastie T, Tibshiran R. *An Introduction to Statistical Learning*. 1st ed. Springer-Verlag; 2013.
10. Singal AG, Mukherjee A, Elmunzer BJ, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol*. 2013;108:1723–1730.
11. Waljee AK, Higgins PD, Singal AG. A primer on predictive models. *Clin Transl Gastroenterol*. 2014;5:e44.
12. Prosperi MC, Marinho S, Simpson A, Custovic A, Buchan IE. Predicting phenotypes of asthma and eczema with machine learning. *BMC Med Genomics*. 2014;7(Suppl 1):S7.
13. Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann N Y Acad Sci*. 2017;1387:153–165.
14. Goto T, Camargo CA, Jr., Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *Am J Emerg Med*. 2018;36:1650–1654.
15. Patel SJ, Chamberlain DB, Chamberlain JM. A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. *Acad Emerg Med*. 2018;25:1463–70.
16. Saglani S, Custovic A. Childhood asthma: advances using machine learning and mechanistic studies. *Am J Respir Crit Care Med*. 2018;199:414–422.
17. Chatzimichail EA, Rigas AG, Paraskakis EN. An artificial intelligence technique for the prediction of persistent asthma in children. In *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*; 2010:1–4.
18. Chatzimichail E, Paraskakis E, Rigas A. An evolutionary two-objective genetic algorithm for asthma prediction. In *2013 UKSim*

- 15th International Conference on Computer Modelling and Simulation; 2013:90-94.
19. Chatzimichail E, Paraskakis E, Sitzimi M, Rigas A. An intelligent system approach for asthma prediction in symptomatic preschool children. *Computd Math Methods Med*. 2013;2013:240182.
 20. Chatzimichail E, Paraskakis E, Rigas A. Predicting asthma outcome using partial least square regression and artificial neural networks. *Adv Artif Intell*. 2013;1-7.
 21. Krautenbacher N, Flach N, Böck A, et al. A strategy for high-dimensional multivariable analysis classifies childhood asthma phenotypes from genetic, immunological, and environmental factors. *Allergy*. 2019.
 22. Fehrenbach H, Smolinska A, Klaassen EMM, et al. Profiling of volatile organic compounds in exhaled breath as a strategy to find early predictive signatures of asthma in children. *PLoS One*. 2014;9.
 23. AlSaad R, Malluhi Q, Janahi I, Boughorbel S. Interpreting patient-specific risk prediction using contextual decomposition of BiLSTMs: application to children with asthma. *BMC Med Inform Decis Mak*. 2019;19:214.
 24. Bose S, Kenyon CC, Masino AJ. Personalized prediction of early childhood asthma persistence: a machine learning approach. *PLoS One*. 2021;16:e0247784.
 25. Owora AH, Tepper RS, Ramsey CD, Becker AB. Decision tree-based rules outperform risk scores for childhood asthma prognosis. *Pediatr Allergy Immunol*. 2021.
 26. Patel D, Hall GL, Broadhurst D, Smith A, Schultz A, Foong RE. Does machine learning have a role in the prediction of asthma in children? *Paediatr Respir Rev*. 2021.
 27. Ahmad MA, Teredesai A, Eckert C. Interpretable machine learning in healthcare. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*; 2018:447.
 28. Arshad SH, Holloway JW, Karmaus W, et al. Cohort profile: the Isle of Wight Whole Population Birth Cohort (IOWBC). *Int J Epidemiol*. 2018;47:1043-4i.
 29. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20:40-49.
 30. He H, Garcia EA, Adasyn SL. Adaptive synthetic sampling approach for imbalanced learning. In *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong*; 2008:1322-1328.
 31. Custovic A, Simpson BM, Murray CS, Lowe L, Woodcock A. The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatr Allergy Immunol*. 2002;13:32-37.
 32. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. arXiv preprint. arXiv:170507874 2017.
 33. Vial Dupuy A, Amat F, Pereira B, Labbe A, Just J. A simple tool to identify infants at high risk of mild to severe childhood asthma: the Persistent Asthma Predictive Score. *J Asthma*. 2011;48:1015-1021.
 34. Biagini Myers JM, Schauburger E, He H, et al. A Pediatric Asthma Risk Score to better predict asthma development in young children. *J Allergy Clin Immunol*. 2018;143:1803-1810e2.
 35. Castro-Rodríguez JAH, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. *Am J Respir Crit Care Med*. 2000;162:1403-1406.
 36. Global Initiative for Asthma. *Global Strategy for Asthma Management and Prevention*; 2018.
 37. Isabelle Guyon JW, Barnhill S. Gene selection for cancer classification using support vector machines. *Mach Learn*. 2002;46:389-422.
 38. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinf*. 2007;8.
 39. Van Wonderen KE, Van Der Mark LB, Mohrs J, Bindels PJ, Van Aalderen WM, Ter Riet G. Different definitions in childhood asthma: how dependable is the dependent variable? *Eur Respir J*. 2010;36:48-56.
 40. Klaassen EM, van de Kant KD, Jobsis Q, et al. Exhaled biomarkers and gene expression at preschool age improve asthma prediction at 6 years of age. *Am J Respir Crit Care Med*. 2015;191:201-207.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kothalawala DM, Murray CS, Simpson A, et al. Development of childhood asthma prediction models using machine learning approaches. *Clin Transl Allergy*. 2021;e12076. <https://doi.org/10.1002/ctt2.12076>

Supplementary Material

Table of Contents

Methods	4
Datasets	4
Developmental Study Population: Isle of Wight Birth Cohort	4
Validation Study Population: Manchester Asthma and Allergy Study	4
Feature Selection	5
Pre-processing of candidate predictors	5
Recursive Feature Elimination	5
Classification Algorithms	6
Support Vector Machine	6
Decision Tree	6
Random Forest	6
Naïve Bayes	7
Multilayer Perceptron	7
K-Nearest Neighbours	7
Pre-processing of the training and test set	8
Hyperparameter tuning	8
Optimisation of the training dataset	8
Multivariate Imputation by Chain Equation (MICE)	9
Adaptive Synthetic (ADASYN) sampling	10
Random Undersampling	10
Sensitivity Analyses	10

Assessing the robustness to predict an alternative asthma definition.....	11
Assessing the resolution to predict childhood wheeze phenotypes	11
Comparison between the machine learning models and similar logistic regression models.....	12
Using SHAP to aid model interpretability	14
Software	14
Supplementary Material References	15
Supplementary Tables	17
Supplementary Figures	37
TRIPOD Checklist: Prediction Model Development and Validation.....	43

List of Supplementary Tables

Table E1 List of candidate predictors of childhood asthma.....	17
Table E2 Hyperparameter tuning criteria for each of the seven machine learning algorithms	19
Table E3 Comparability between predictors definitions in the IOWBC and MAAS cohorts.....	21
Table E4 Descriptive statistics for all candidate features in the IOWBC analysed and the subset of individuals with complete data used for feature selection for the CAPE and CAPP prediction models	23
Table E5 Distribution of CAPE and CAPP model predictors for individuals in the IOWBC and MAAS at each asthma prediction timepoint.....	30
Table E6 Model performance for predicting an alternative definition of asthma	32
Table E7 Comparison of the prediction models developed using machine learning and traditional logistic regression algorithms	33
Table E8 Reclassification table comparing predictions made by the PARS and CAPP models in MAAS	34
Table E9 Comparison of the performance of the CAPE and CAPP models with existing validated prediction models	35
Table E10 Performance of the CAPP Tool with and without the predictors of sensitisation	36

List of Supplementary Figures

Figure E1 Agreement between the original and modified asthma definitions	37
Figure E2 Global interpretation of the contribution each predictor makes to the model predictions	38
Figure E3 Local interpretation of individual predictions made by the CAPE and CAPP models.....	39
Figure E4 Parental questionnaire for collecting data needed for the CAPE tool.....	41
Figure E5 Parental questionnaire for collecting data needed for the CAPP tool.....	42

Methods

Datasets

Developmental Study Population: Isle of Wight Birth Cohort

Of the 1536 newborns born between January 1989 and February 1990, 1456 were recruited and prospectively followed up at 1, 2, 4, 10, 18 and 26 years of age¹. Individuals lost to follow-up were significantly more likely to have asthmatic siblings, be of low birthweight and have parents who smoked. At each visit, information was collected on clinical symptoms of asthma and other allergic diseases as well as lifestyle and environmental exposures through hospital records, physical examinations and study-specific parent/participant questionnaires. Skin prick testing (SPT) was performed in infants with allergy-related symptoms at the 1-year and 2-year follow-ups, and in all participants from the 4-year follow-up onwards.

Asthmatic children at age 10 were significantly more likely to be male, have a lower birthweight, be atopic and experience a range of asthma-like symptoms, both in early life and at preschool age (Table E4).

Validation Study Population: Manchester Asthma and Allergy Study

The MAAS cohort is a longitudinal, whole-population UK birth cohort established in 1995 to study the development of asthma and other atopic disorders in childhood². In brief, parents were recruited between 1995 and 1997, and 1184 children followed up at 1, 3, 5, 8, 11, 13-16 and 18 years. Medical records and validated questionnaires were used to collect data on clinical symptoms of allergy and asthma and environmental exposures. Blood samples were taken and SPT and lung function tests were performed from the three-year follow-up onwards.

Early life predictor data was collected from the 1-year and 3-year follow-ups in MAAS (1-year and 2-year follow-ups in the IOWBC), and from the 5-year follow-up for preschool predictors (4-year follow-up in the IOWBC).

Feature Selection

Pre-processing of candidate predictors

Pre-processing of all candidate predictor data included the removal of extreme outliers ($\pm 4SD$) present in each continuous variable and one-hot encoding of nominal variables into separate binary variables. Child BMI reported at 1 and 4 years old were standardised against the British 1990 growth reference³.

Recursive Feature Elimination

Feature selection was conducted using Recursive Feature Elimination (RFE) with a random forest algorithm. RFE initially builds a random forest classifier using all of the candidate features and evaluates its predictive performance^{4,5}. In an iterative process, the lowest ranking feature, in terms of Gini feature importance, is removed and the random forest classifier is rebuilt on the remaining feature subset and the predictive performance of the classifier is re-evaluated. This process was performed within a five-fold cross-validation framework whereby: i) the dataset was split into five equal folds, preserving the number of asthma cases in each fold; ii) the random forest classifier for RFE was built on 4 of the 5 folds and the balanced accuracy of predicting both asthmatic cases and non-asthmatic controls was evaluated on the remaining fold; iii) the optimal subset of predictors was identified based on the random forest model that achieved the highest cross-validation balanced accuracy score. A variation of the random forest algorithm (balanced random forest), which randomly under-samples the majority class in each bootstrap, was used to account for the class imbalance present in the dataset. For feature selection, all continuous variables were standardised to zero mean and unit variance.

Classification Algorithms

Support Vector Machine

Support vector machines aim to construct a separating hyperplane between the outcome classes. A soft margin is used to penalize misclassifications whilst trying to optimise the best classification. When dealing with high-dimensional data, SVM exploits the kernel trick to map the data onto a higher dimensional space in order to construct the separating hyperplane between the outcome classes^{6,7}. This study utilized three different kernel functions - linear, radial basis function (RBF), and polynomial - for constructing the best support vector classifier were used.

Decision Tree

Decision trees aim to stratify the predictor space using simple splitting rules. Starting at the top (root node) of the tree, the predictor space is stratified at each internal node. Predictor variables which create the best separation of the outcome classes are calculated at each node, with better splitting variables utilised at nodes higher in the tree structure. The final nodes at the bottom of the tree, at which no further separations are made, are referred to as leaves or terminal nodes and provide the final classification⁶.

Random Forest

An ensemble classifier that aggregates the decisions of multiple decision trees. For the development of each tree, a bootstrapped dataset the same size as the original dataset is created by randomly resampling the original dataset. Unlike the decision tree algorithm, each tree in a random forest only considers a random subset of variables for stratifying the predictor space. As a result, the correlation between the trees will be reduced. Although each tree will have low bias and high variance, the process of bootstrapping and aggregating decisions across the trees to make a final classification (known as bagging), will result in predictions of low variance and high accuracy⁶.

Naïve Bayes

Based on conditional probability, this algorithm is underpinned by the assumption that each feature is independent of the other in determining the outcome class. Implementation of the Naïve Bayes algorithm requires assumptions of each class' prior probability distribution, commonly Gaussian and multinomial/Bernoulli distribution for continuous and categorical features, respectively⁶.

Multilayer Perceptron

A simple feed-forward artificial neural network which is well-suited to distinguish between non-linearly separable data through a network of nodes. A MLP consists of a minimum of three layers; an input layer, at least one hidden layer and an output layer. Each neuron in the input layer is fully connected to each neuron in the next layer. These connections are weighted, with the weights optimised through a process of backpropagation which aims to minimize the error of the output layer which determines the class classification^{8,9}.

K-Nearest Neighbours

KNN is an instance-based learning algorithm which utilises the feature space to classify unknown data points based on a number of data points in close proximity (nearest neighbours) for which the class labels are known. Based on the conditional probability of the unknown data point belonging to each class, a final classification is made⁶.

Pre-processing of the training and test set

Prior to model development (and redevelopment), the continuous variables in the training dataset were standardized to a mean of zero and unit variance, and the same standardisation properties were applied to the test set.

Hyperparameter tuning

The hyperparameters of each machine learning algorithm were tuned during the training process to optimise for the model's balanced accuracy. Balanced accuracy was chosen as the optimisation metric in order to maximize the predictive accuracy of both outcome classes whilst accounting for the inherent class imbalance of the training dataset.

To identify the optimal hyperparameters, a grid search was used to systematically search over a range of each model's hyperparameters within a 5-fold cross-validation (ranges detailed in Table E2). To reduce computational time, a random search (a sample of 100 parameter settings) was performed for the SVM algorithms to narrow the hyperparameter ranges specified for evaluation in the grid search. The Naïve Bayes algorithm did not require any hyperparameter tuning (Table E2).

Optimisation of the training dataset

Due to potential bias of using only complete data for model training, the training set was optimised in an attempt to further improve the predictive performance of the classification algorithms.

First, missing predictor data for all individuals that were not allocated to the test set, (n=1185 and n=1119 for the CAPE and CAPP models, respectively), were imputed using Multivariate Imputation by Chain Equation (MICE). ADASYN (an adaptive synthetic sampling approach) was then applied to the imputed training dataset to generate new examples of the minority class (asthmatic cases), with a bias towards increasing the number of difficult to classify examples. The effect of oversampling was assessed across a range of levels – the number of asthma cases were increased by 0%, 25%, 50%, 100%, 150%, 200%, 250% and 300%. Finally, random under-sampling was applied to each of the imputed and oversampled training datasets in order to completely balance the number of asthmatic cases and non-asthmatic controls (1:1 ratio).

Multivariate Imputation by Chain Equation (MICE)

Multivariate imputation by chain equation is a type of multiple imputation used under the assumption that data is MAR. By performing multiple imputations and generating a set of plausible estimates, MICE aims to account for the statistical uncertainty associated with the imputation¹⁰. MICE can be applied to datasets containing variables of mixed datatypes. For each datatype, a different imputation model can be used^{10,11}.

In the implementation of MICE, all missing values are initially assigned a placeholder value based on mean imputation or random sampling (with replacement) of the observed data for each variable. For the first variable with missing data, x_1 , the placeholder values are removed and x_1 is regressed on the remaining variables $[x_2, x_3, \dots, x_i]$. The regression is limited to only those examples for which x_1 was observed. The missing values for x_1 are then predicted from the posterior predictive distribution generated by the imputation model. This process is repeated for the remaining missing variables, where for example, x_2 is regressed on the remaining variables $([x_3, x_4, \dots, x_i])$ and the newly imputed variable (x_1), again, limited to examples with observed data for x_2 . One cycle is complete when all of the variables with missing data have been imputed. Numerous cycles are performed in order to converge the distribution parameters of each variable and create a single dataset of stable imputation estimates. To generate multiple (m) imputed datasets, this entire process is repeated m times^{10,11}.

Following the imputation stage and the generation of multiple imputed datasets, subsequent analyses should be conducted on each of the m -imputed datasets and the results are pooled. The pooled results provide estimates with confidence intervals, addressing the statistical uncertainty of the imputation^{10,11}. However, due to the need to tune each of the machine learning algorithm to establish a single model with a single set of tuned parameters, a single imputed dataset was required for model development. To form a single imputed dataset, the imputed values generated across the m imputed datasets were averaged, with the mean and modal imputed values taken for the continuous and categorical variables, respectively.

Adaptive Synthetic (ADASYN) sampling

Adaptive Synthetic (ADASYN) sampling is an example of a synthetic data generation approach which aims to generate examples of the minority class in order to improve imbalances in data¹². The approach is based on the KNN algorithm. ADASYN can specify the construction of datasets with varying degrees of balance. In contrast to randomly oversampling examples of the minority class, the synthetic examples generated through ADASYN are informed by a density distribution of weights for examples of the minority class. The weight assigned to each example is determined by the ratio of examples belonging to the minority class in its k- nearest neighbours. These weights correspond to the learning difficulty of each example and subsequently determines the number of synthetic examples of the minority class that needs to be generated. For example, a difficult to classify example of the minority class (i.e. one that is similar to examples of the majority class) will have a small ratio of minority examples within its k nearest neighbours therefore will have a large weight. Subsequently, a greater number of synthetic examples will be generated based on this minority example. As a result, the learning model will have a greater opportunity to learn from difficult to classify examples of the minority class in addition to reducing the bias of the model by correcting for the class imbalance¹².

Random Undersampling

The number of non-asthmatic individuals in the training datasets were randomly undersampled. The training dataset was shuffled and a random subset of non-asthmatic individuals were removed in order to balance the outcomes classes in the training dataset – i.e. 1:1 ratio of asthmatic and non-asthmatics.

Sensitivity Analyses**Assessing model generalisability in high risk subgroups**

Model generalisability was assessed in MAAS among three risk groups: i) all individuals (unselected population); ii) individuals with at least one parent with allergic disease (asthma, eczema or allergic rhinitis); and iii) individuals with two parents with allergic disease.

Assessing the robustness to predict an alternative asthma definition

The robustness of the CAPE and CAPP models was evaluated using an alternative definition of school-age asthma that incorporated an objective outcome measure. Using this alternative asthma definition, a child was considered asthmatic if they presented with wheeze in the last 12 months and had bronchial hyper-responsiveness (BHR), defined as a 20% reduction in forced expiratory volume in one second (FEV₁) following a methacholine challenge test¹³.

Asthma status, based on the alternative asthma definition, was available for 1312 of the 1368 individuals analysed in the IOWBC (prevalence 8.61%). Despite an overall 92.3% agreement, there was a statistically significant difference between the two asthma definitions ($p < 0.01$). This stemmed from a 97.6% agreement for labelling non-asthmatics but only a 53.8% agreement for labelling asthmatics (Figure E1).

A labelled asthma status using the alternative asthma definition incorporating BHR was available for 248 out of 255 individuals in the CAPE test dataset (20 asthmatic) and 179 out of 183 individuals in the CAPP test dataset (18 asthmatic). The CAPE and CAPP models were less robust to predict the alternative asthma outcome (CAPE AUC=0.67 vs 0.71 and CAPP AUC=0.79 vs 0.82). Both models demonstrated an increased sensitivity to predict asthmatics, but the corresponding increase in false positive predictions resulted in PPV reducing by approximately 50% (Table E7).

Assessing the resolution to predict childhood wheeze phenotypes

The resolution of the asthma predictions to distinguish between individuals presenting with distinct wheeze phenotypes throughout childhood and adolescence was assessed. The

identification and assignment of individuals in the IOWBC and MAAS into one of five distinct wheeze phenotypes through a latent class analysis has been previously described¹⁴.

Briefly, using wheeze data available across five time-points, a latent class analysis of 7,719 individuals from five UK birth cohorts (including the IOWBC and MAAS) identified five distinct phenotypes of wheeze - never/infrequent wheeze, early onset preschool remitting, early onset mid-childhood remitting, persistent, and late-onset wheeze (full details on the analysis can be found in reference ¹⁴). Based on the latent class analysis, each individual was provided with a probability for belonging to each wheeze phenotype. In this analysis, individuals were categorised into their most probable wheeze phenotype.

In the CAPE and CAPP test datasets, 213 and 167 individuals had a defined wheeze phenotype, respectively. In MAAS, 237 and 216 individuals with complete predictor and school-age asthma data for the CAPE and CAPP models, respectively, also had a defined wheeze phenotype.

Comparison between the machine learning models and similar logistic regression models

To evaluate the hypothesis that machine learning methods may offer greater predictions of childhood asthma than regression-based methods, the performance of the CAPE and CAPP machine learning models were directly compared against i) equivalent logistic regression models developed using the same predictors; and ii) their current regression-based benchmark models (see Supplementary Methods).

Derivation of equivalent logistic regression models for the CAPE and CAPP models

To directly compare if the use of more complex machine learning algorithms offer greater predictions of childhood asthma than regression-based methods, CAPE and CAPP equivalent logistic regression models were constructed. The logistic regression equivalent models used the same predictors and training datasets that were used to construct the CAPE and CAPP machine learning models. For the CAPE-logistic regression model, the model was trained on the complete training dataset which was undersampled to obtain 1:1 class ratio. For the

CAPP-logistic regression model, the complete training dataset, oversampled by 300% and undersampled to obtain 1:1 class ratio, was used. To construct the logistic regression models, the scikit learn logistic regression model was used, with the 'lbfgs' solver¹⁵. No further regularisation of the predictor coefficients were applied during the construction of the logistic regression model.

Comparison of the CAPE and CAPP models with current benchmark models

The developed machine learning models were compared against current published models. The API, the most widely known asthma prediction tool, was unable to be replicated due to the absence of eosinophil data in the IOWBC. Of the remaining validated regression-based models, the PAPS (Persistent Asthma Predictive Score)¹⁶ and PARS (Paediatric Asthma Risk Score)¹⁷ were considered the best performing models comparable with the CAPE and CAPP models, offering predictions in early life and at preschool age, respectively. However, PAPS was also unable to be replicated as specific-IgE tests were not performed in the IOWBC. PARS was able to be replicated in both the IOWBC and MAAS.

For replication of the PARS model in the IOWBC, race was not reported for each individual; reported cohort demographics suggest the cohort is 96% Caucasian, hence all individuals were assumed to be Caucasian in this replication. All individuals with complete data for the PARS predictors and the asthma outcome were included in the analysis (predicting asthma in the IOWBC at age 10: n=913, in MAAS at age 8 years: n=552, in MAAS at age 11 years: n=487). Models were compared against the CAPP machine learning models based on AUC.

Individuals with predictions from both the CAPP and PARS models were used to compare the differences in individual predictions between the two models. In the IOWBC, only individuals in the test set (i.e. not used to train the model) were included. Reclassification tables were used to evaluate the differences in predictions on an individual level, for asthmatics and non-asthmatic individuals separately¹⁸. The table presents the differences in prediction categorization using the CAPP model compared to the PARS model. The net proportion of individuals reclassified by the CAPP model to a more appropriate prediction

group was summarized using the net reclassification indices for true future asthmatics and non-asthmatics separately (NR_{levent} and NR_{Inon-event}, respectively)¹⁸:

$$NR_{levent} = P(up|event) - P(down|event)$$

$$NR_{Inon-event} = P(down|non-event) - P(up|non-event)$$

Using SHAP to aid model interpretability

SHapley Additive exPlanations (SHAP) is a tool¹⁹, which unifies a number of model interpretability methods, capable of explaining the decisions of any machine learning model. SHAP is often considered an extension of Shapley values, a concept rooted in game theory, which aims to determine the average contribution of each feature in deriving a certain prediction.

The average contribution of each predictor was approximated based on the difference in the predictions obtained the inclusion and exclusion of the predictor, averaged across all possible permutations of the model. For the subset of predictors identified from the feature selection process, SHAP values were used to infer feature importance and effect (direction of risk for developing asthma).

Based on the global explanations of each model, as a sensitivity analysis, the CAPE and CAPP models were redeveloped using the subset of features shown to offer the greatest contribution to model predictions. Examples of how SHAP values can be used locally to explain individual predictions were also provided.

Software

Data cleaning, pre-processing and imputation were performed using R statistical programming language (version 3.5.1)²⁰. All other data manipulation and modelling was performed using Python scripting language (version 3.6.8) and the machine learning libraries Scikit-learn¹⁵ and imbalanced-learn²¹. The SHAP package was used compute the SHAP values used to enhance model interpretability¹⁹. SHAP TreeExplainer was used to calculate SHAP values for the specific random forest algorithm used during RFE²². SHAP KernelExplainer was used compute SHAP values to explain individual predictions made by the final CAPE and CAPP models^{19,23}.

Supplementary Material References

1. Arshad SH, Holloway JW, Karmaus W, et al. Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC). *International Journal of Epidemiology* 2018;47:1043-4i.
2. Custovic A, Simpson BM, Murray CS, Lowe L, Woodcock A. The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatric Allergy and Immunology : Official Publication of the European Society of Pediatric Allergy and Immunology* 2002;13:32-7.
3. Cole T. The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition* 1990;44:45-60.
4. Isabelle Guyon JW, Stephen Barnhill. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 2002;46:389-422.
5. Granitto PM, Furlanello C, Biasioli F, Gasperi F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems* 2006;83:83-90.
6. James G, Witten D, Hastie T, Tibshiran R. *An Introduction to Statistical Learning*. 1 ed: Springer-Verlag New York; 2013.
7. Ben-Hur A, Ong CS, Sonnenburg S, Scholkopf B, Ratsch G. Support vector machines and kernels for computational biology. *PLOS Computational Biology* 2008;4:e1000173.
8. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*: Springer; 2001.
9. M.W.Gardner SRD. Artificial neural networks (the multilayer perceptron) - A review of applications in the atmospheric sciences. *Atmospheric Environment* 1998;32:2627-36.
10. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* 2011;30:377-99.
11. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research* 2011;20:40-9.
12. Haibo He YB, Eduardo A. Garcia, and Shutao Li. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. Hong Kong 2008:1322-8.
13. Scottish Intercollegiate Guidelines Network BTS. *British guideline on the management of asthma: A national clinical guideline* 2019.
14. Oksel C, Granell R, Haider S, et al. Distinguishing Wheezing Phenotypes from Infancy to Adolescence. A Pooled Analysis of Five Birth Cohorts. *Annals of the American Thoracic Society* 2019;16:868-76.
15. Fabian Pedregosa GV, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research* 2011;12:2825–30.
16. Vial Dupuy A, Amat F, Pereira B, Labbe A, Just J. A Simple Tool to Identify Infants at High Risk of Mild to Severe Childhood Asthma: The Persistent Asthma Predictive Score. *Journal of Asthma* 2011;48:1015-21.
17. Biagini Myers JM, Schauburger E, He H, et al. A Pediatric Asthma Risk Score to better predict asthma development in young children. *Journal of Allergy and Clinical Immunology* 2018;143:1803-10.e2.
18. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* 2014;25:114-21.
19. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *arXiv preprint arXiv:170507874* 2017.
20. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013.

21. Guillaume Lemaître FN, Christos K. Aridas. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research* 2017;18:1-5.
22. Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell* 2020;2:56-67.
23. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018;2:749-60.

Supplementary Tables

Table E1 List of candidate predictors of childhood asthma

Candidate predictor	Definition
Family History	
Maternal smoking at birth	Maternal smoking status during pregnancy
Paternal smoking at birth	Paternal smoking status during pregnancy
Maternal asthma	Maternal asthma status
Maternal eczema	Maternal eczema status
Maternal hay fever	Maternal hay fever status
Paternal asthma	Paternal asthma status
Paternal eczema	Paternal eczema status
Paternal hay fever	Paternal hay fever status
Parity	Position of child in the family
SES	Maternal socioeconomic status
Prenatal/ at birth	
Maternal age	Maternal age at pregnancy
Prematurity	Gestation age
Caesarean delivery	Child birth through caesarean delivery
Total breastfeeding	Total breastfeeding duration
Exclusive breastfeeding	Exclusive breastfeeding duration
Solid food introduction	Age, in months, at which solid foods were introduced to the child's diet
Birthweight	Birth weight (kg)
Sex	Child's gender
Season of birth	Season of birth
Dog	Household pet dog during pregnancy
Cat	Household pet cat during pregnancy
Furry pet	Household furry pet during pregnancy - dog, cat or other animal
Early life (1 and 2 year follow-up)	
SDS BMI	BMI at age 1, standardised against the British 1990 growth reference
Wheeze	Occurrence of wheezing before age 2
Wheeze without cold	Likely occurrence of wheezing in the absence of a cold before age 2
Cough	Occurrence of cough before age 2
Nasal symptoms	Occurrence of nasal symptoms before age 2
Chest infection	Occurrence of chest infections before age 2
Nocturnal symptoms	Occurrence of nocturnal asthma symptoms before age 2
Eczema	Eczema status by age 2
Hay fever	Hay fever status by age 2
Atopy	Atopy status (sensitisation to one or more allergens) by age 2
Monosensitisation	Sensitisation to one allergen by age 2
Polysensitisation	Sensitisation to two or more allergens by age 2
Parental smoking	Household parental smoking status by age 2
Dog	Household pet dog by age 2
Cat	Household pet cat by age 2
Furry pet	Household furry pet (dog, cat or other animal) by age 2

Candidate predictor	Definition
Early-life living on a farm	Main residence on a farm in the first year of life
Preschool age (4 year follow-up)	
BMI	BMI at age 4, standardised against the British 1990 growth reference
Wheeze	Occurrence of wheezing at age 4
Wheeze without cold	Likely occurrence of wheezing in the absence of a cold at age 4
Cough	Occurrence of cough at age 4
Nasal symptoms	Occurrence of nasal symptoms at age 4
Nocturnal symptoms	Occurrence of nocturnal asthma symptoms at age 4
Eczema	Eczema status at age 4
Hay fever	Hay fever status at age 4
Atopy	Atopy status (sensitisation to one or more allergens) at age 4
Monosensitisation	Sensitisation to one allergen at age 4
Polysensitisation	Sensitisation to two or more allergens at age 4
Parental smoking	Household parental smoking status at age 4
Dog	Household pet dog at age 4
Cat	Household pet cat at age 4
Furry pet	Household furry pet (dog, cat or other animal) at age 4

Table E2 Hyperparameter tuning criteria for each of the seven machine learning algorithms

Algorithm	Hyperparameters	Description	Search range
Support Vector Machine	Cost	Regularisation term	100 values between 10^{-3} and 10^{2^a}
	Gamma	Scalar tem for the RBF and polynomial kernels	100 values between 10^{-2} and 10^{2^a}
	Degree	Degree term for the polynomial kernel	1,2,3,...,10
Decision Tree	Max tree depth	The maximum depth each tree should be constructed to	1,2,3,...,32 or None
	Min samples split	The minimum number of samples needed to split a node	2,3,4,...,11
	Max features	The maximum number of features to consider to find the best split	'log2', 'sqrt', None
	Splitter	Criteria used to choose the split at a node	'best', 'random'
	Criterion	Criteria used to determine the quality of a node split	Gini, entropy
Random forest	N estimators (trees)	The number of trees used to construct the forest	1,2,4,8,16,32,64,100,200
	Max tree depth	The maximum depth each tree should be constructed to	1,2,3,...,32
	Min samples split	The minimum number of samples needed to split a node	2,3,4,...,11
	Max features	The maximum number of features to consider to find the best split	'log2', 'sqrt', None
	Criterion	Criteria used to determine the quality of a node split	Gini, entropy
	Bootstrap	Determines whether bootstrapping with replacement should be used to build the trees	True, False
Multilayer Perceptron	Hidden layers	The number of neurons in each hidden layer	(1,),(2,...(11, (1,1),(2,2),...(11,11) ^b
	Activation	The activation function for the hidden layers	'relu', 'identity', 'tanh', 'logistic'
	Solver	Criteria used to optimise the weights of the connections	'lbfgs', 'sgd', 'adam'
	Alpha	Regularisation term	10^{-1} , 10^{-2} , 10^{-6}
	Learning rate	The rate at which to update the weights	'constant', 'invscaling', 'adaptive'
	Initial learning rate	The initial learning rate	0.1,0.2,...,0.9

Algorithm	Hyperparameters	Description	Search range
KNN	Number of neighbours (k)	The number of neighbours	1,2,3,...,100
	Weight	Determines whether each neighbour should be weighted equally or based on their distance	Uniform, distance
	Power	Specifies the distance measure to use	Manhattan, Euclidean
Naïve Bayes	Distribution	Determines which distribution each feature is assumed to follow	Continuous features = Gaussian distribution. Categorical features = multinomial distribution

^a Specifies the parameter space for the random search strategy. Based on the results of the random search, a refined grid search across 500 steps was specified.

^b Number of neurons in each hidden layer, where (1,) represents 1 neuron in the first hidden layer, with no further hidden layers; and (1,2) represents 1 neuron in the first hidden layer and 2 in the second hidden layer.

^c The naïve Bayes algorithm did not undergo any hyperparameter grid search but instead required variables to be specified as either continuous and categorical at the time of model development.

Table E3 Comparability between predictor definitions in the IOWBC and MAAS cohorts

Variable	IOWBC definition	MAAS definition	Comparability ^a
Maternal age	Maternal age at booking	Maternal age at birth of child	
Birthweight	Birth weight (kg)	Birthweight (kg)	
Total breastfeeding	Total breastfeeding duration	Breast feeding duration	
Age of solid food introduction	Age of introduction of cereals/solids (weeks)	At what age did your baby begin solid foods? (Weeks)	
Early life BMI	BMI at age 1, standardised against the British 1990 growth reference	SDS BMI at age 1, standardised against the British 1990 growth reference	
Early life wheeze	Frequency of asthma wheezing episodes at either 1 or 2 years	If no to: has or does your child's chest ever wheeze or whistle, what best describes your child's wheezing (at either 1 or 3 years)	IOWBC: categorised as no wheeze, occasional, frequent MAAS: categorised as no wheeze, 1-2 times or from time to time (occasional), every day (frequent)
Early life cough	Asthmatic cough at either 1 or 2 years	Does your child usually have a cough apart from with colds at 1 or 3 years	
Preschool BMI	SDS BMI at age 4	SDS BMI at age 5	
Preschool wheeze	Frequency of wheezing at 4YR	Current wheeze age 5 years	
Preschool cough	Any asthmatic cough at 4 YR	Does your child usually have a cough during the day apart from with colds?	
Preschool nocturnal symptoms	Any nocturnal symptoms at 4YR	Does your child usually have a cough at night apart from with colds? Or, in the last 12 months how often - on average - has your child's sleep been disturbed by wheezing	
Preschool atopy	Sensitisation (+SPT) to one or more allergens at age 4	Sensitisation (+SPT) to one or more allergens at age 5	IOWBC: tested allergens included house dust mite, milk, egg, cat, dog, grass, wheat, soya, peanut, cod, Cladosporium,

Variable	IOWBC definition	MAAS definition	Comparability ^a
Preschool polysensitisation	Sensitisation (+SPT) to two or more allergens by age 4	Sensitisation (+SPT) to two or more allergens by age 5	Alternaria MAAS: tested allergens included house dust mite, cat, dog, pollen, mould, milk, egg IOWBC: tested allergens included house dust mite, milk, egg, cat, dog, grass, wheat, soya, peanut, cod, Cladosporium, Alternaria MAAS: tested allergens included house dust mite, cat, dog, pollen, mould, milk, egg
Maternal socioeconomic status	Maternal socioeconomic status	Maternal socioeconomic status	IOWBC: categorised into the following income strata: very low, low, low-middle, middle and high. MAAS: categorised as routine (low), intermediate (low-middle), managerial (middle) and professional (high).
School-age asthma	Doctor diagnose asthma PLUS wheeze in the last 12 month AND/OR asthma treatment	Doctor diagnose asthma PLUS wheeze in the last 12 month AND/OR asthma treatment	IOWBC: evaluated at age 10 MAAS: evaluated at ages 8 and 11

^a MAAS variable categorisations are given as: original categorisation of the MAAS variables (IOWBC equivalent used in the validation analysis).

Table E4 Descriptive statistics for all candidate features in the IOWBC analysed and the subset of individuals with complete data used for feature selection for the CAPE and CAPP prediction models

		Total IOWBC (n=1368)		CAPE complete dataset (n=490)		CAPP complete dataset (n=373)	
		Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=70)	Non-asthmatic (n=420)	Asthmatic (n=55)	Non-asthmatic (n=318)
Family history	Maternal smoking at birth						
	No	152 (75.62)	877 (75.15)	55 (78.57)	345 (82.14)	43 (78.18)	257 (80.82)
	Yes	47 (23.38)	276 (23.65)	15 (21.43)	75 (17.86)	12 (21.82)	61 (19.18)
	Paternal smoking at birth						
	No	119 (59.20)	714 (61.18)	46 (65.71)	283 (67.39)	34 (61.82)	206 (64.46)
	Yes	79 (39.30)	440 (37.70)	24 (34.29)	137 (32.62)	21 (38.18)	112 (35.22)
	Maternal asthma						
	No	170 (84.58)	1047 (89.72)	65 (92.86)	381 (90.71)	52 (94.55)	293 (92.14)
	Yes	29 (14.43)	113 (9.68)	5 (7.14)	39 (9.29)	3 (5.45)	25 (7.86)
	Maternal eczema						
	No	170 (84.58)	1025 (87.83)	59 (84.29)	371 (88.33)	45 (81.82)	284 (89.31)
	Yes	28 (13.93)	133 (11.40)	11 (15.71)	49 (11.67)	10 (18.18)	34 (10.69)
	Maternal hay fever						
	No	149 (74.13)	941 (80.63)	57 (81.43)	335 (79.76)	43 (78.18)	256 (80.50)
	Yes	50 (24.88)	219 (18.77)	13 (18.57)	85 (20.24)	12 (21.82)	62 (19.50)
	Paternal asthma						
	No	171 (85.07)	1049 (89.89)	60 (85.71)	387 (92.14)	47 (85.45)	291 (91.51)
	Yes	27 (13.43)	104 (8.91)	10 (14.29)	33 (7.86)	8 (14.55)	27 (8.49)
	Paternal eczema						
	No	179 (89.05)	1082 (92.72)	60 (85.71)	395 (94.05)	48 (87.27)	298 (93.71)
	Yes	19 (9.45)	70 (6.00)	10 (14.29)	25 (5.95)	7 (12.73)*	20 (6.29)*
	Paternal hay fever						
	No	163 (81.09)	987 (84.58)	56 (80.00)	366 (87.14)	42 (76.36)	278 (87.42)
	Yes	35 (17.41)	166 (14.22)	14 (20.00)	54 (12.86)	13 (23.64)	40 (12.58)

		Total IOWBC (n=1368)		CAPE complete dataset (n=490)		CAPP complete dataset (n=373)	
		Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=70)	Non-asthmatic (n=420)	Asthmatic (n=55)	Non-asthmatic (n=318)
Parity							
	No	78 (38.81)	415 (35.56)	28 (40.00)	180 (42.86)	19 (34.55)	138 (43.4)
	Yes	95 (47.26)	573 (49.10)	42 (60.00)	240 (57.14)	36 (65.45)	180 (56.60)
Maternal socioeconomic status							
	Very low	25 (12.44)	163 (13.97)	11 (15.71)	43 (10.24)	10 (18.18)	30 (9.43)
	Low	35 (17.41)	199 (17.05)	12 (17.14)	81 (19.29)	8 (14.55)	63 (19.81)
	Low-Mid	62 (30.85)	334 (28.62)	18 (25.71)	128 (30.48)	14 (25.45)	103 (32.39)
	Mid	52 (26.37)	320 (27.42)	21 (30.00)	135 (32.14)	16 (29.09)	99 (31.13)
	High	13 (6.47)	96 (8.23)	8 (11.43)	33 (7.86)	7 (12.73)	23 (7.23)
Prenatal/at birth	Maternal age	201	1167	70	420	55	318
		(26.61, 5.44)	(27.04, 5.26)	(27.44, 5.32)	(27.60, 4.91)	(27.98, 5.37)	(27.69, 4.95)
Prematurity							
	Pre-term	9 (4.48)	32 (2.74)	1 (1.43)	7 (1.67)	1 (1.82)	4 (1.26)
	Term	184 (91.54)	1103 (94.52)	67 (95.71)	411 (97.86)	53 (96.36)	312 (98.11)
	Post-term	3 (1.49)	12 (1.03)	2 (2.86)	2 (0.48)	1 (1.82)	2 (0.63)
Caesarean delivery							
	No	150 (74.63)	857 (73.44)	62 (88.57)	381 (90.71)	48 (87.27)	287 (90.25)
	Yes	18 (8.96)	86 (7.37)	8 (11.43)	39 (9.29)	7 (12.73)	31 (9.75)
Total breastfeeding							
	Never	46 (22.89)	267 (22.88)	18 (25.71)	95 (22.62)	15 (27.27)	73 (22.96)
	<3months	66 (32.84)	352 (30.16)	29 (41.43)	137 (32.62)	22 (40.00)	105 (33.02)
	3-6 months	22 (10.95)	164 (14.05)	4 (5.71)	69 (16.43)	3 (5.45)	56 (17.61)
	>6 months	37 (18.41)	264 (22.62)	19 (27.14)	119 (28.33)	15 (27.27)	84 (26.42)
Exclusive breastfeeding							
	Never	55 (27.36)	334 (28.62)	25 (35.71)	125 (29.76)	21 (38.18)	91 (28.62)
	<3 months	85 (42.29)	489 (41.90)	35 (50.00)	191 (45.48)	27 (49.09)	146 (45.91)
	>3 months	31 (15.42)	224 (19.19)	10 (14.29)	104 (24.76)	7 (12.72)	81 (25.47)

		Total IOWBC (n=1368)		CAPE complete dataset (n=490)		CAPP complete dataset (n=373)	
		Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=70)	Non-asthmatic (n=420)	Asthmatic (n=55)	Non-asthmatic (n=318)
Early life	Age of solid food introduction in weeks	168 (14.36, 4.51)	1026 (14.34, 4.12)	70 (13.96, 4.08)	420 (14.45, 4.08)	55 (13.96, 4.24)	318 (14.59, 4.07)
	Birthweight	199 (3.34, 0.52)*	1142 (3.44, 0.50)*	70 (3.45, 0.51)	420 (3.47, 0.52)	55 (3.48, 0.56)	318 (3.45, 0.49)
	Sex	*	*				
	Male	118 (58.71)	578 (49.53)	40 (57.14)	191 (45.48)	31 (56.36)	143 (44.97)
	Female	83 (41.29)	589 (50.47)	30 (42.86)	229 (54.52)	24 (43.64)	175 (55.03)
	Season of birth						
	Autumn	38 (18.91)	243 (20.82)	13 (18.57)	101 (24.05)	12 (21.82)	86 (27.04)
	Winter	64 (31.84)	382 (32.73)	19 (27.14)	117 (27.86)	14 (25.45)	77 (24.21)
	Spring	51 (25.37)	274 (23.48)	19 (27.14)	100 (23.81)	15 (27.27)	80 (25.16)
	Summer	48 (23.88)	268 (22.96)	19 (27.14)	102 (24.29)	14 (25.45)	75 (23.58)
	Dog						
	No	149 (74.13)	815 (69.84)	55 (78.57)	301 (71.67)	41 (74.55)	226 (71.07)
	Yes	51 (25.37)	346 (29.65)	15 (21.43)	119 (28.33)	14 (25.45)	92 (28.93)
	Cat						
	No	143 (71.14)	764 (65.47)	46 (65.71)	274 (65.24)	35 (63.64)	205 (64.47)
	Yes	57 (28.36)	397 (34.02)	24 (34.29)	146 (34.76)	20 (36.36)	113 (35.53)
	Furry pet						
	No	105 (52.24)	525 (44.99)	39 (55.71)	194 (46.19)	29 (52.73)	143 (44.97)
	Yes	95 (47.26)	636 (54.50)	31 (44.29)	226 (53.81)	26 (47.27)	175 (55.03)
Early life	BMI	135 (-0.15, 1.15)	851 (-0.16, 1.22)	70 (-0.13, 1.12)	420 (-0.17, 1.19)	55 (-0.20, 1.13)	318 (-0.18, 1.19)
	Wheeze	*	*	*	*	*	*
	Never	78 (38.81)	739 (63.32)	39 (55.71)	341 (81.19)	30 (54.55)	259 (81.45)
	Occasional	14 (6.97)	63 (5.40)	6 (8.57)	28 (6.67)	5 (9.09)	20 (6.29)
	Frequent	56 (27.86)	124 (10.63)	25 (35.71)	51 (12.14)	20 (36.36)	39 (12.26)

		Total IOWBC (n=1368)		CAPE complete dataset (n=490)		CAPP complete dataset (n=373)	
		Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=70)	Non-asthmatic (n=420)	Asthmatic (n=55)	Non-asthmatic (n=318)
Wheeze without cold		*	*	*	*	*	*
	No	89 (44.28)	796 (68.21)	45 (64.29)	369 (87.86)	35 (63.64)	279 (87.74)
	Yes	56 (27.86)	124 (10.63)	25 (35.71)	51 (12.14)	20 (36.36)	39 (12.26)
Cough		*	*	*	*	*	*
	No	78 (38.81)	749 (64.18)	39 (55.71)	346 (82.38)	30 (54.55)	263 (82.7)
	Yes	70 (34.83)	174 (14.91)	31 (44.29)	74 (17.62)	25 (45.45)	55 (17.30)
Nasal symptoms		*	*	*	*	*	*
	No	101 (50.25)	757 (64.87)	45 (64.29)	331 (78.81)	36 (65.45)	252 (79.25)
	Yes	60 (29.85)	239 (20.48)	25 (35.71)	89 (21.19)	19 (34.55)	66 (20.75)
Chest infection		*	*	*	*	*	*
	No	103 (51.24)	830 (71.12)	51 (72.86)	375 (89.29)	40 (72.73)	284 (89.31)
	Yes	54 (26.87)	139 (11.91)	19 (27.14)	45 (10.71)	15 (27.27)	34 (10.69)
Nocturnal symptoms		*	*	*	*	*	*
	No	80 (39.80)	756 (64.78)	41 (58.57)	347 (82.62)	32 (58.18)	264 (83.02)
	Yes	68 (33.83)	161 (13.80)	29 (41.43)	73 (17.28)	23 (41.82)	54 (16.98)
Eczema		*	*	*	*		
	No	95 (47.26)	740 (63.41)	46 (65.71)	320 (76.19)	35 (63.64)	246 (77.36)
	Yes	66 (32.84)	247 (21.17)	24 (34.29)	100 (23.81)	20 (36.36)	72 (22.64)
Hay fever							
	No	123 (61.19)	827 (70.87)	53 (75.71)	353 (84.05)	42 (76.36)	268 (84.28)
	Yes	36 (17.91)	165 (14.14)	17 (24.29)	67 (15.95)	13 (23.64)	50 (15.72)
Atopy		*	*	*	*	*	*
	No	95 (47.26)	834 (71.47)	53 (75.71)	392 (93.33)	43 (78.18)	304 (95.6)
	Yes	44 (21.89)	58 (4.97)	17 (24.29)	28 (6.67)	12 (21.82)	14 (4.40)
Monosensitisation		*	*	*	*	*	*
	No	99 (49.25)	840 (71.98)	55 (78.57)	394 (93.81)	44 (80)	305 (95.91)
	Yes	37 (18.41)	51 (4.37)	15 (21.43)	26 (6.19)	11 (20.00)	13 (4.09)

		Total IOWBC (n=1368)		CAPE complete dataset (n=490)		CAPP complete dataset (n=373)	
		Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=70)	Non-asthmatic (n=420)	Asthmatic (n=55)	Non-asthmatic (n=318)
Polysensitisation		*	*	*	*	*	*
	No	119 (59.20)	876 (75.06)	67 (95.71)	417 (99.29)	53 (96.36)	317 (99.69)
	Yes	10 (4.98)	9 (0.77)	3 (4.29)	3 (0.71)	2 (3.64)	1 (0.31)
Parental smoking							
	Never	69 (34.33)	470 (40.27)	33 (47.14)	231 (55.00)	24 (43.64)	168 (52.83)
	Ex-smoker	5 (2.49)	54 (4.63)	2 (2.86)	29 (6.90)	2 (3.64)	23 (7.23)
	Current	93 (46.27)	488 (41.82)	35 (50.00)	160 (38.10)	29 (52.73)	127 (39.94)
Dog							
	No	118 (58.71)	656 (56.21)	55 (78.57)	294 (70.00)	41 (74.55)	219 (68.87)
	Yes	41 (20.40)	327 (28.02)	15 (21.43)	126 (30.00)	14 (25.45)	99 (31.13)
Cat							
	No	26 (12.94)	176 (15.08)	11 (15.71)	71 (16.90)	9 (16.36)	54 (16.98)
	Yes	131 (65.17)	821 (70.35)	59 (84.29)	349 (83.10)	46 (83.64)	264 (83.02)
Furry pet							
	No	6 (2.99)	19 (1.63)	4 (5.71)	8 (1.90)	3 (5.45)	7 (2.20)
	Yes	155 (77.11)	1001 (85.78)	66 (94.29)	412 (98.10)	52 (94.55)	311 (97.80)
Early life residence on a farm							
	No	174 (86.57)	990 (84.83)	66 (94.29)	398 (94.76)	52 (94.55)	305 (95.91)
	Yes	6 (2.99)	43 (3.68)	4 (5.71)	22 (5.24)	3 (5.45)	13 (4.09)
Preschool age	BMI	146 (0.21, 1.03)	855 (0.23, 1.04)	-	-	55 (0.28, 0.88)	318 (0.28, 0.95)
	Wheeze	*	*			*	*
	Never	85 (42.29)	879 (75.32)	-	-	28 (50.91)	281 (88.36)
	Occasional	18 (8.96)	34 (2.91)	-	-	7 (12.73)	10 (3.14)
	Frequent	70 (34.83)	75 (6.43)	-	-	20 (36.36)	27 (8.49)
	Wheeze without cold	*	*			*	*
	No	103 (51.24)	913 (78.23)	-	-	35 (63.64)	291 (91.51)
	Yes	70 (34.83)	75 (6.43)	-	-	20 (36.36)	27 (8.49)

		Total IOWBC (n=1368)		CAPE complete dataset (n=490)		CAPP complete dataset (n=373)	
		Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=70)	Non-asthmatic (n=420)	Asthmatic (n=55)	Non-asthmatic (n=318)
Cough		*	*			*	*
	No	74 (36.82)	860 (73.69)	-	-	23 (41.82)	279 (87.74)
	Yes	99 (49.25)	128 (10.97)	-	-	32 (58.18)*	39 (12.26)*
Nasal symptoms		*	*			*	*
	No	112 (55.72)	852 (73.01)	-	-	34 (61.82)	275 (86.48)
	Yes	61 (30.35)	136 (11.65)	-	-	21 (38.18)*	43 (13.52)*
Nocturnal symptoms		*	*			*	*
	No	79 (39.30)	860 (73.69)	-	-	25 (45.45)	277 (87.11)
	Yes	94 (46.77)	128 (10.97)	-	-	30 (54.55)*	41 (12.89)*
Eczema		*	*				
	No	126 (62.69)	901 (77.21)	-	-	45 (81.82)	290 (91.19)
	Yes	47 (23.38)	87 (7.46)	-	-	10 (18.18)	28 (8.81)
Hay fever		*	*			*	*
	No	144 (71.64)	953 (81.66)	-	-	45 (81.82)	303 (95.28)
	Yes	29 (14.43)	35 (3.00)	-	-	10 (18.18)*	15 (4.72)*
Atopy		*	*			*	*
	No	67 (33.33)	670 (57.41)	-	-	29 (52.73)	275 (86.48)
	Yes	72 (35.82)	94 (8.05)	-	-	26 (47.27)*	43 (13.52)*
Monosensitisation		*	*				
	No	115 (57.21)	708 (60.67)	-	-	46 (83.64)	292 (91.82)
	Yes	22 (10.95)	53 (4.54)	-	-	9 (16.36)	26 (8.18)
Polysensitisation		*	*			*	*
	No	93 (46.27)	874 (74.89)	-	-	38 (69.09)	301 (94.65)
	Yes	48 (23.88)	38 (3.26)	-	-	17 (30.91)*	17 (5.35)*

		Total IOWBC (n=1368)		CAPE complete dataset (n=490)		CAPP complete dataset (n=373)	
		Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=70)	Non-asthmatic (n=420)	Asthmatic (n=55)	Non-asthmatic (n=318)
Parental smoking							
	Never	62 (30.85)	424 (36.33)	-	-	22 (40.00)	157 (49.37)
	Ex-smoker	19 (9.45)	122 (10.45)	-	-	6 (10.91)	47 (14.78)
	Current	78 (38.81)	357 (30.59)	-	-	27 (49.09)	114 (35.85)
Dog							
	No	125 (62.19)	710 (60.84)	-	-	40 (72.73)	227 (71.38)
	Yes	50 (24.88)	280 (23.99)	-	-	15 (27.27)	91 (28.62)
Cat							
	No	115 (57.21)	620 (53.13)	-	-	40 (72.73)	198 (62.26)
	Yes	60 (29.85)	370 (31.71)	-	-	15 (47.27)	120 (37.74)
Furry pet							
	No	75 (37.31)	386 (33.08)	-	-	18 (32.73)	123 (38.68)
	Yes	96 (47.76)	586 (50.21)	-	-	37 (67.27)	195 (61.32)

Summary data is reported as the number of individuals with data, with the mean and standard deviation ($\bar{x}$, s) for the continuous features of: maternal age, birthweight, age of solid food introduction, early life BMI and preschool BMI; or proportions for the remaining categorical features (%). Where the number of individuals with data for a variable does not equal the total number of individuals detailed in the column, the difference indicates the number of individuals with missing data.

* Indicates a statistically significant difference for the variable between asthmatic and non-asthmatic children at age 10 ($p < 0.05$), assessed using an independent two sample t-test or Pearson's Chi-square test for independence for continuous and categorical features, respectively.

Table E5 Distribution of CAPE and CAPP model predictors for individuals in the IOWBC and MAAS at each asthma prediction time point.

	IOWBC 10YR (n=1368)		MAAS 8YR (n=1018)		MAAS 11YR (n=898)	
	Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=144)	Non-asthmatic (n=874)	Asthmatic (n=116)	Non-asthmatic (n=782)
Maternal age	201 (26.61, 5.44)	1167 (27.04, 5.26)	116 (30.53, 5.09)	842 (20.66, 4.67)	94 (29.59, 4.94))*	762 (30.88, 4.61)*
Birthweight	199 (3.34, 0.52)*	1142 (3.44, 0.50)*	132 (3.44, 0.50)	845 (3.49, 0.49)	107 (3.41, 0.51)	757 (3.49, 0.49)
Age of solid food introduction	168 (14.36, 4.51)	1026 (14.34, 4.12)	51 (14.88, 3.83)	392 (14.67, 3.52)	44 (14.93, 5.03)	351 (14.69, 3.34)
Breastfeeding duration						
Never	46 (22.89)	267 (22.88)	47 (23.64)	281 (32.15)	32 (27.59)	236 (30.18)
<3months	66 (32.84)	352 (30.16)	33 (22.92)	214 (24.49)	30 (25.86)	190 (24.30)
3-6 months	22 (10.95)	164 (14.05)	24 (16.67)	162 (18.54)	16 (12.79)	157 (20.08)
>6 months	37 (18.41)	264 (22.62)	22 (15.28)	194 (22.20)	23 (19.83)	181 (23.15)
Early life BMI	135 (-0.15, 1.15)	851 (-0.16, 1.22)	49 (-0.18, 1.00)	387 (-0.25, 1.11)	43 (-0.04, 1.09)	347 (-0.32, 1.12)
Preschool BMI	146 (0.21, 1.03)	855 (0.23, 1.04)	134 (0.57, 0.95)	804 (0.46, 0.94)	113 (0.65, 0.90)*	731 (0.42, 0.94)*
Early life wheeze	*	*	*	*	*	*
Never	78 (38.81)	739 (63.32)	8 (5.56)	212 (24.26)	9 (7.76)	189 (24.17)
Occasional	14 (6.97)	63 (5.40)	93 (64.58)	339 (38.79)	67 (57.76)	299 (38.24)
Frequent	56 (27.86)	124 (10.63)	4 (2.78)	6 (0.69)	4 (3.45)	5 (0.64)
Early life cough	*	*	*	*	*	*
No	78 (38.81)	749 (64.18)	31 (21.53)	318 (36.38)	28 (24.14)	281 (35.93)
Yes	70 (34.83)	174 (14.91)	34 (23.61)	135 (15.45)	32 (27.59) *	117 (14.96) *
Preschool wheeze	*	*	*	*	*	*
Never	85 (42.29)	879 (75.32)	44 (30.56)	721 (82.49)	40 (34.48)	655 (83.76)
Occasional	18 (8.96)	34 (2.91)	89 (61.81)	116 (13.27)	66 (56.90)	105 (13.43)
Frequent	70 (34.83)	75 (6.43)	5 (3.47)	6 (0.69)	6 (5.17)	3 (0.38)
Preschool cough	*	*	*	*	*	*
No	74 (36.82)	860 (73.69)	70 (48.61)	711 (81.35)	62 (53.45)	642 (82.10)
Yes	99 (49.25)	128 (10.97)	68 (47.22)	132 (15.10)	50 (43.10)	121 (15.47)

	IOWBC 10YR (n=1368)		MAAS 8YR (n=1018)		MAAS 11YR (n=898)	
	Asthmatic (n=201)	Non-asthmatic (n=1167)	Asthmatic (n=144)	Non-asthmatic (n=874)	Asthmatic (n=116)	Non-asthmatic (n=782)
Preschool nocturnal symptoms	*	*	*	*	*	*
No	79 (39.30)	860 (73.69)	49 (34.03)	700 (80.09)	47 (40.52)	629 (80.43)
Yes	94 (46.77)	128 (10.97)	89 (61.81)	143 (16.36)	65 (56.03)	134 (17.14)
Preschool atopy status	*	*	*	*	*	*
No	67 (33.33)	670 (57.41)	52 (36.11)	573 (65.56)	38 (32.76)	530 (67.77)
Yes	72 (35.82)	94 (8.05)	76 (52.78)	197 (22.54)	69 (59.48)	169 (21.61)
Preschool polysensitisation status	*	*	*	*	*	*
No	93 (46.27)	874 (74.89)	77 (53.47)	675 (77.23)	62 (53.45)	612 (78.26)
Yes	48 (23.88)	38 (3.26)	47 (32.64)	77 (8.81)	41 (35.34)	73 (9.34)
Maternal socioeconomic status						
Very low	25 (12.44)	163 (13.97)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Low	35 (17.41)	199 (17.05)	11 (7.64)	66 (7.55)	9 (7.76)	54 (6.91)
Low-Mid	62 (30.85)	334 (28.62)	15 (10.42)	137 (15.68)	10 (8.62)	129 (16.50)
Mid	52 (26.37)	320 (27.42)	41 (28.47)	388 (44.39)	30 (25.86)	375 (47.95)
High	13 (6.47)	96 (8.23)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

Summary data for predictors included in the CAPE and CAPP models is reported for all individuals with a reported asthma status in the IOWBC or MAAS at each prediction time point.

The distribution of predictors is reported as the number of individuals, with mean and standard deviation ($\bar{x}$, s) for the continuous features of: maternal age, birthweight, age of solid food introduction, early life BMI and preschool BMI; or as proportions for the remaining categorical features (%). Where the number of individuals with data for a variable does not equal the total number of individuals detailed in the column, the difference indicates the number of individuals with missing data.

* Statistically significant difference between asthmatic and non-asthmatic children ($p < 0.05$), assessed using an independent two sample t-test or Pearson's Chi-square test for independence for continuous and categorical features, respectively.

Table E6 Model performance for predicting an alternative definition of asthma

	Asthma definition [†] (% asthmatic)	Balanced accuracy	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-	F1 score
CAPE Model	IOWBC Original (13.3%)	0.71 (0.62, 0.78)	0.71 (0.61, 0.80)	0.74 (0.56, 0.88)	0.68 (0.62, 0.74)	0.26 (0.21, 0.32)	0.94 (0.91, 0.97)	2.29 (1.69, 3.01)	0.39 (0.18, 0.63)	0.38 (0.31, 0.46)
	IOWBC Alternative (8.1%)	0.70 (0.62, 0.76)	0.67 (0.56, 0.78)	0.90 (0.75, 1.00)	0.49 (0.43, 0.56)	0.13 (0.11, 0.16)	0.98 (0.96, 1.00)	1.77 (1.44, 2.12)	0.20 (0.00, 0.49)	0.23 (0.20, 0.27)
	MAAS 8YR Alternative (5.0%)	0.61 (0.51, 0.68)	0.69 (0.55, 0.82)	0.87 (0.67, 1.00)	0.34 (0.29, 0.40)	0.07 (0.05, 0.08)	0.98 (0.95, 1.00)	1.32 (1.02, 1.57)	0.39 (0.00, 0.97)	0.12 (0.09, 0.14)
	MAAS 11YR Alternative (3.0%)	0.60 (0.45, 0.69)	0.58 (0.37, 0.75)	0.86 (0.57, 1.00)	0.33 (0.28, 0.39)	0.03 (0.02, 0.04)	0.99 (0.97, 1.00)	1.29 (0.84, 1.61)	0.43 (0.00, 1.34)	0.06 (0.04, 0.08)
	IOWBC Original (13.7%)	0.80 (0.70, 0.89)	0.82 (0.71, 0.91)	0.72 (0.52, 0.88)	0.88 (0.83, 0.92)	0.47 (0.38, 0.62)	0.95 (0.92, 0.98)	5.99 (3.79, 10.11)	0.32 (0.13, 0.54)	0.56 (0.45, 0.70)
	IOWBC Alternative (10.1%)	0.77 (0.68, 0.85)	0.79 (0.67, 0.89)	0.83 (0.67, 1.00)	0.71 (0.64, 0.78)	0.25 (0.19, 0.31)	0.97 (0.95, 1.00)	2.92 (2.11, 4.07)	0.23 (0.00, 0.48)	0.38 (0.30, 0.46)
CAPP Model	MAAS 8YR Alternative (5.3%)	0.68 (0.56, 0.78)	0.70 (0.57, 0.82)	0.79 (0.57, 1.00)	0.57 (0.51, 0.63)	0.09 (0.07, 0.12)	0.98 (0.96, 1.00)	1.83 (1.29, 2.39)	0.38 (0.00, 0.77)	0.17 (0.12, 0.21)
	MAAS 11YR Alternative (2.4%)	0.71 (0.53, 0.81)	0.68 (0.40, 0.87)	0.83 (0.50, 1.00)	0.58 (0.52, 0.64)	0.05 (0.03, 0.06)	0.99 (0.98, 1.00)	1.98 (1.15, 2.63)	0.29 (0.00, 0.88)	0.09 (0.05, 0.12)

Performance measures are reported, with 95% confidence intervals from 2000 bootstrapped samples in brackets.

[†] The outcome of school-age asthma was defined as follows: original asthma definition= doctor diagnosis of asthma ever plus the presence of wheeze or use of asthma medication in the last 12 months; alternative asthma definition= current wheeze and bronchial hyper-responsiveness. Both asthma outcomes were evaluated at age 10 in the IOWBC, in individuals in the respective test sets for each model.

Table E7 Comparison of the prediction models developed using machine learning and traditional logistic regression algorithms

	Algorithm (dataset)	Balanced accuracy	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-	F1 score
CAPE Model	SVM [†]	0.71	0.71	0.74	0.68	0.26	0.94	2.29	0.39	0.38
	(IOWBC test set)	(0.62, 0.78)	(0.61, 0.80)	(0.56, 0.88)	(0.62, 0.74)	(0.21, 0.32)	(0.91, 0.97)	(1.69, 3.01)	(0.18, 0.63)	(0.31, 0.46)
	Logistic regression [‡]	0.62	0.59	0.44	0.80	0.25	0.90	2.22	0.70	0.32
	(IOWBC test set)	(0.54, 0.71)	(0.48, 0.70)	(0.27, 0.59)	(0.75, 0.85)	(0.17, 0.35)	(0.88, 0.93)	(1.33, 3.43)	(0.49, 0.91)	(0.21, 0.43)
	Logistic regression (MAAS 8YR)	0.60	0.47	0.39	0.80	0.21	0.91	2.00	0.75	0.28
	(MAAS 11YR)	(0.52, 0.68)	(0.36, 0.59)	(0.24, 0.55)	(0.75, 0.85)	(0.13, 0.29)	(0.89, 0.93)	(1.15, 3.02)	(0.55, 0.96)	(0.17, 0.38)
CAPP Model	Logistic regression (MAAS 8YR)	0.58	0.49	0.38	0.79	0.18	0.91	1.79	0.79	0.24
	(MAAS 11YR)	(0.50, 0.67)	(0.36, 0.61)	(0.22, 0.53)	(0.74, 0.84)	(0.10, 0.25)	(0.89, 0.94)	(0.96, 2.84)	(0.57, 1.01)	(0.14, 0.34)
	SVM [§]	0.80	0.82	0.72	0.88	0.47	0.95	5.99	0.32	0.56
	(IOWBC test set)	(0.70, 0.89)	(0.71, 0.91)	(0.52, 0.88)	(0.83, 0.92)	(0.38, 0.62)	(0.92, 0.98)	(3.79, 10.11)	(0.13, 0.54)	(0.45, 0.70)
	Logistic regression [¶]	0.77	0.76	0.80	0.74	0.33	0.96	3.08	0.27	0.47
	(IOWBC test set)	(0.68, 0.85)	(0.63, 0.88)	(0.64, 0.96)	(0.67, 0.80)	(0.26, 0.41)	(0.93, 0.99)	(2.24, 4.34)	(0.05, 0.50)	(0.38, 0.56)
	Logistic regression (MAAS 8YR)	0.72	0.77	0.82	0.61	0.22	0.96	2.12	0.30	0.35
	(MAAS 11YR)	(0.64, 0.78)	(0.67, 0.85)	(0.70, 0.94)	(0.55, 0.67)	(0.18, 0.26)	(0.94, 0.99)	(1.68, 2.66)	(0.10, 0.52)	(0.29, 0.41)
	Logistic regression (MAAS 11YR)	0.71	0.76	0.79	0.62	0.20	0.96	2.10	0.33	0.32
	(MAAS 11YR)	(0.62, 0.78)	(0.64, 0.86)	(0.66, 0.93)	(0.56, 0.68)	(0.17, 0.25)	(0.93, 0.99)	(1.62, 2.67)	(0.11, 0.59)	(0.26, 0.39)

The CAPE and CAPP machine learning and equivalent logistic regression models were evaluated at age 10 in the IOWBC, in individuals in the respective test sets for each model. Validation in MAAS was performed to evaluate the prediction of asthma at 8 years (MAAS 8YR) and 11 years (MAAS 11YR).

Performance measures are reported, with 95% confidence intervals from 2000 bootstrapped samples in brackets.

Performance measures were evaluated at the optimal model thresholds, which maximised the Youden's Index: [†]=0.42, [‡]=0.48, [§]=0.73, [¶]=0.42.

Table E8 Reclassification table comparing predictions made by the PARS and CAPP models in MAAS

		Predicted risk (CAPP model)			Reclassified by CAPP (%)			
	Predicted risk (PARS model)	No asthma	Asthma	Total	Increased risk	Decreased risk	Correctly reclassified	NRI
MAAS 8YR (PARS AUC=0.86 vs CAPP=0.83)								
No asthma (n=213)	No asthma	173	14[‡]	187				
	Asthma	21⁺	5	26	14(7%)	21(10%)	21(10%)	0.03
	Total	194	19	213				
Asthma (n=28)	No asthma	5	7⁺	12				
	Asthma	7[‡]	9	16	7(25%)	7(25%)	7(25%)	0.00
	Total	12	16	28	21	28	28	
MAAS 11YR (PARS AUC=0.78 vs CAPP=0.79)								
No asthma (n=215)	No asthma	170	14[‡]	184				
	Asthma	24⁺	7	31	14(7%)	24(11%)	24(11%)	0.05
	Total	194	21	215				
Asthma (n=26)	No asthma	8	7⁺	15				
	Asthma	4[‡]	7	11	7(27%)	4(15%)	7(27%)	0.12
	Total	12	14	26	21	28	21	

Reclassification table comparing the change in individual asthma predictions with the CAPP model instead of the PARS model (reference model). For the PARS model, categorisations of predictions as asthmatic and non-asthmatic was based on the optimal threshold (cutoff=7) as defined in their original publication. Results are presented separately for individuals who were asthmatic and non-asthmatic at ages 8 and 11 in MAAS. Values in bold identify the number of individuals who were reclassified into a more appropriate (⁺) or less appropriate ([‡]) risk group by the CAPP model with respect to the risk classifications made by the PARS model. NRI=net reclassification index is given separately for true asthmatics and non-asthmatics.

Table E9 Comparison of the performance of the CAPE and CAPP models with existing validated prediction models

Model	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-
CAPE	0.71	0.74	0.68	0.26	0.94	2.29	0.39
CAPP	0.82	0.72	0.88	0.47	0.95	5.99	0.32
Loose API	-	0.42	0.85	0.59	0.73	2.72 [†]	0.69 [†]
Stringent API	-	0.16	0.97	0.76	0.68	6.04 [†]	0.87 [†]
PIAMA	0.74	0.19	0.97	0.42	0.91	6.33 [†]	0.84 [†]
PAPS	0.66	0.42	0.90	0.67	0.76	4.06	0.64
PARC	0.74	0.72	0.71	0.49	0.86	2.47	0.40
PARS	0.80	0.68	0.77	0.37	0.93	3.02	0.41

Predictive performance of the published validated studies compared to the machine learning CAPE and CAPP models developed in this study. None of the machine learning models are reported in this table as none of them were externally validated. For each model, performance measures are reported based on the optimal threshold specified in their original papers. The shaded rows indicate the models which were developed in this study.

[†] Performance measures were not reported in the original study so were calculated to enable comparison.

Table E10 Performance of the CAPP Tool with and without the predictors of sensitisation

	Dataset	Balanced Accuracy	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-	F1 Score
CAPP Model[†]	IOWBC	0.80 (0.70, 0.89)	0.82 (0.71, 0.91)	0.72 (0.52, 0.88)	0.88 (0.83, 0.92)	0.47 (0.38, 0.62)	0.95 (0.92, 0.98)	5.99 (3.79, 10.11)	0.32 (0.13, 0.54)	0.56 (0.45, 0.70)
	MAAS 8YR	0.73 (0.64, 0.81)	0.83 (0.75, 0.90)	0.55 (0.36, 0.70)	0.91 (0.88, 0.95)	0.45 (0.33, 0.59)	0.94 (0.92, 0.96)	6.17 (3.64, 10.69)	0.50 (0.33, 0.69)	0.49 (0.36, 0.62)
	MAAS 11YR	0.73 (0.63, 0.82)	0.79 (0.68, 0.88)	0.55 (0.38, 0.72)	0.90 (0.87, 0.94)	0.41 (0.29, 0.55)	0.94 (0.92, 0.96)	5.71 (3.44, 9.85)	0.50 (0.30, 0.71)	0.47 (0.33, 0.62)
CAPP Model without sensitisation[‡]	IOWBC	0.75 (0.64, 0.84)	0.72 (0.58, 0.85)	0.64 (0.44, 0.80)	0.85 (0.80, 0.91)	0.41 (0.30, 0.53)	0.94 (0.91, 0.97)	4.40 (2.71, 7.22)	0.42 (0.22, 0.66)	0.50 (0.36, 0.63)
	MAAS 8YR	0.67 (0.59, 0.76)	0.79 (0.70, 0.87)	0.47 (0.31, 0.64)	0.87 (0.83, 0.91)	0.33 (0.23, 0.44)	0.92 (0.90, 0.95)	3.69 (2.23, 5.91)	0.61 (0.41, 0.80)	0.39 (0.27, 0.51)
	MAAS 11YR	0.65 (0.56, 0.74)	0.70 (0.57, 0.81)	0.42 (0.26, 0.58)	0.87 (0.83, 0.91)	0.29 (0.19, 0.40)	0.92 (0.90, 0.95)	3.32 (1.87, 5.55)	0.66 (0.46, 0.87)	0.34 (0.21, 0.47)

The CAPP machine learning model was evaluated with and without predictors needing SPTs to be performed at age 10 in the IOWBC test set and at ages 8 (MAAS 8YR) and 11 (MAAS 11YR) in MAAS.

Performance measures are reported, with 95% confidence intervals from 2000 bootstrapped samples in brackets.

Performance measures were evaluated at the optimal model thresholds, which maximised the Youden's Index: [†]=0.73, [‡]=0.45.

Supplementary Figures

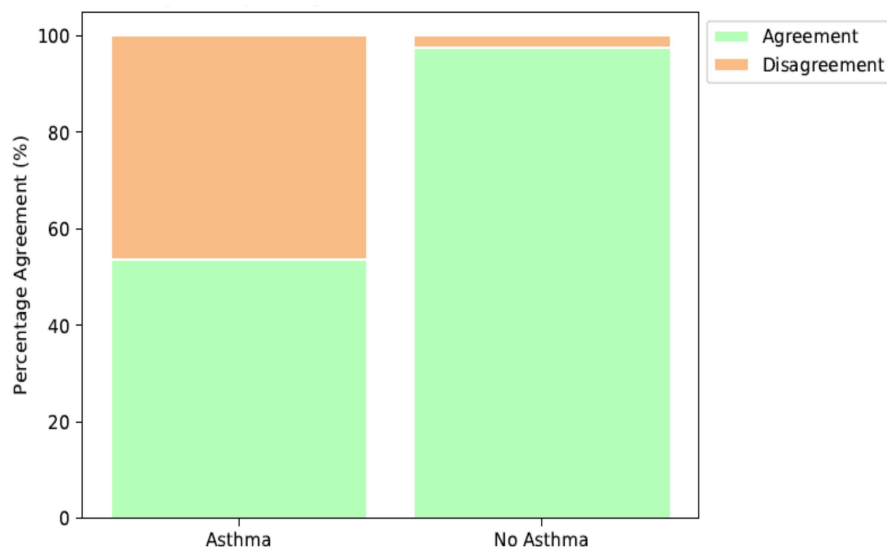


Figure E1 Agreement between the original and modified asthma definitions

Of the 1368 individuals in the IOWBC included in the main study, 1312 individuals had their asthma status defined using the two asthma definitions: original definition used in the analysis (doctor diagnosis asthma ever and wheeze or use of asthma medication in the last 12 months) and a modified definition (wheeze in the last 12 months and BHR). Each stacked bar represents the classification of individuals as asthmatic (left, n=160) or non-asthmatic (right, n=1152) based on the original asthma definition. Each bar shows the proportion of individuals for whom the modified asthma definition assigned the same asthma status (green stacks) or opposing asthma status (orange stacks) compared to the original asthma definition.

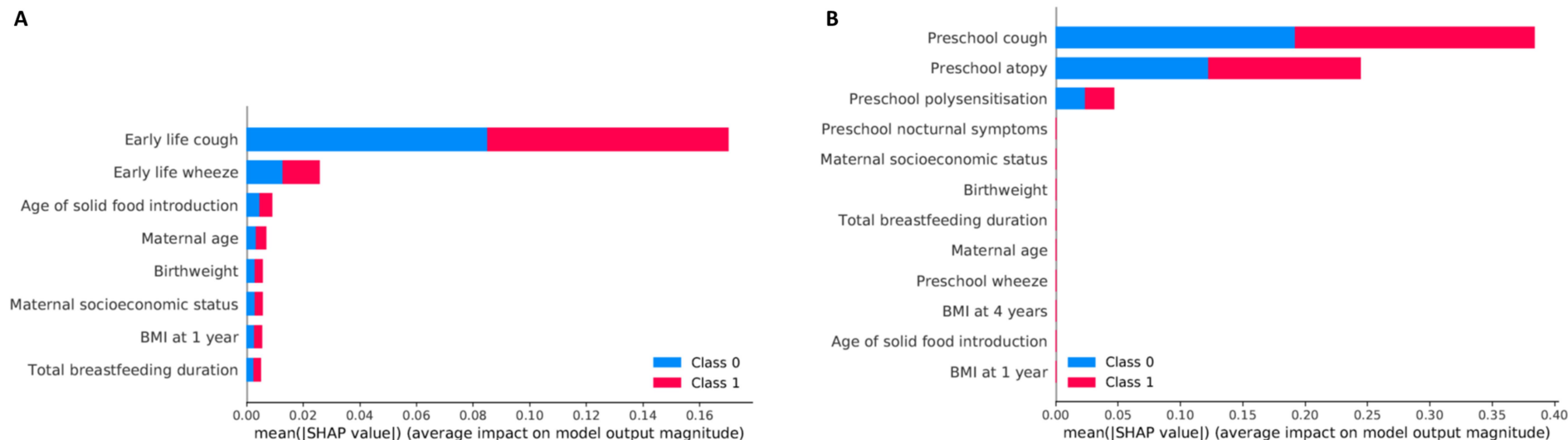


Figure E2 Global interpretation of the contribution each predictor makes to the model predictions

The stacked bar plot shows the mean absolute SHAP value of each feature across all samples in the IOW test set for the CAPE (A) and CAPP (B) models. The bars show the impact each feature has on the model offering a prediction of asthma (Class 1 – red bars) and no asthma (Class 0 – blue bars). The order of the bars corresponds to the contribution (feature importance) each predictor makes in determining the output of the models (top to bottom = highest to lowest contribution).

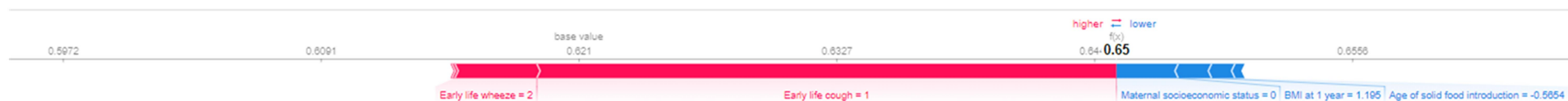
A**B**

Figure E3 Local interpretation of individual predictions made by the CAPE and CAPP models

Example explanations of the predictions offered by the CAPE (A) and CAPP (B) models for two randomly selected individuals are shown. (A) The individual was offered a predicted probability ($f(x)$) of 0.65 for developing school-age asthma at age 10. The plot shows that frequent early life wheeze and early life cough positively contributed to increase the probability of the individual being classified as asthmatic. Low maternal socioeconomic status, higher than average BMI and earlier than average introduction of solid foods into the diet negatively impacted a prediction of asthma, contributing to a reduction in the overall predicted probability of asthma. (B) The individual depicted was offered a predicted probability of 0.72 for developing school-age asthma by the CAPP model. Having preschool cough and being both atopic and polysensitised at age 4, all contributed to an increase in the predicted probability of asthma. The base value is the reference point the feature contributions take effect from, and is defined as the prediction that would be made for an asthmatic individual if none of the features of the current output were known.

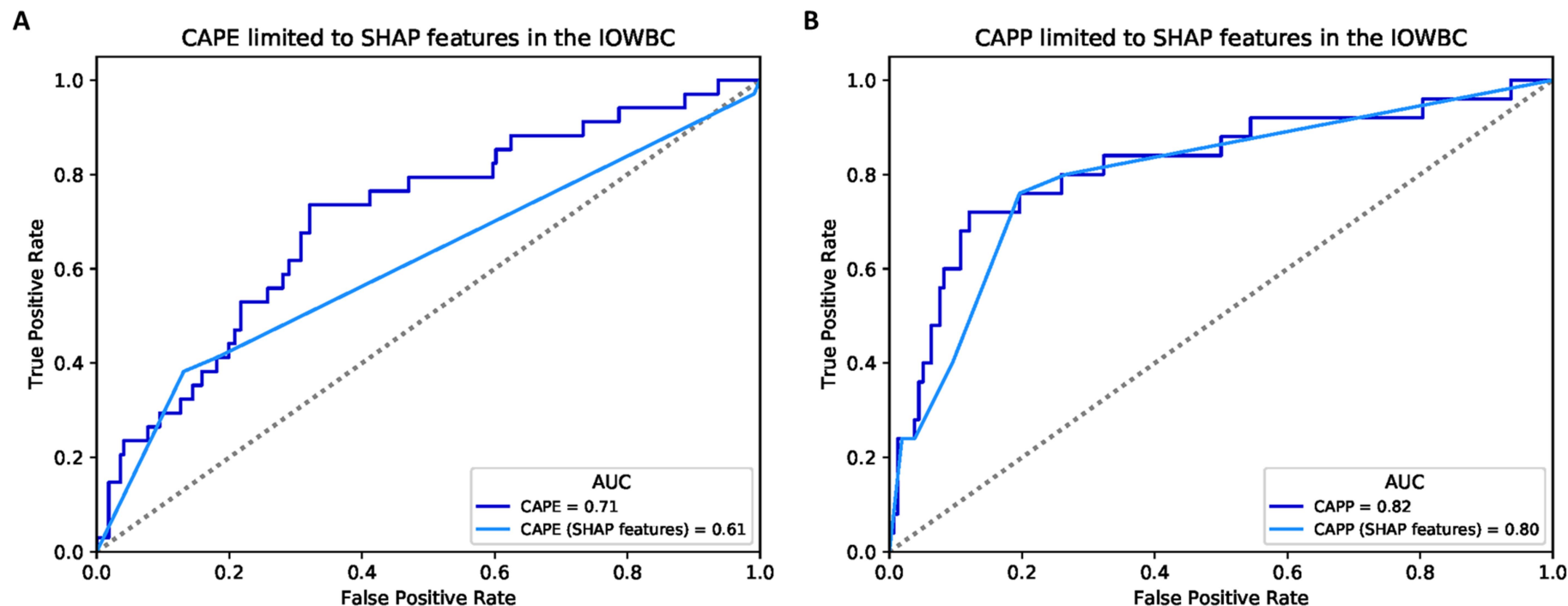


Figure E4 ROC curves comparing the performance of the original and SHAP feature restricted CAPE and CAPP models

Performance is presented for the CAPE (A) and CAPP (B) models in the IOWBC using all features selected through RFE (dark blue line) and the subset of predictors identified as offering the greatest contributions to model predictions (light blue line). Highly contributing predictors: CAPE= early life cough and wheeze, CAPP= preschool cough, atopy and polysensitisation.

CAPE Tool Questionnaire (use if the child is an infant)	
What was your child's birthweight?	___ kg
At what age was your child introduced to solid foods?	___ weeks
Did you breastfeed your child? If yes, how long did you breastfeed for?	___ months
What was your child's BMI when they were one year old?	___
How often did your child experience any wheezing or whistling in their first two years of life?	☐ Never ☐ 1-3 times per year ☐ 3+ times per year
Did your child experience coughing in the absence of a cold in their first two years of life?	Yes/ No
How old were you (the mother) at the time of your child's birth?	___ years
How would you best describe your socioeconomic class?	☐ Very low ☐ Low ☐ Lower-Middle ☐ Middle ☐ Upper

Figure E5 Parental questionnaire for collecting data needed for the CAPE tool

CAPP Tool Questionnaire (use if the child is of preschool age, 3-5 years old)	
What was your child's birthweight?	__ kg
At what age was your child introduced to solid foods?	__ weeks
Did you breastfeed your child? If yes, how long did you breastfeed for?	__ months
What was your child's approximate height and weight when they were one year old?	__ cm __ kg
How old were you (the mother) at the time of your child's birth?	__ years
How would you best describe your socioeconomic class?	☐ Very low ☐ Low ☐ Lower-Middle ☐ Middle ☐ Upper
Current status:	
What is your child's approximate height and weight?	__ cm __ kg
How often does your child experience any wheezing or whistling?	☐ Never ☐ 1-3 times per year ☐ 3+ times per year
Does your child experience coughing in the absence of a cold?	Yes/ No
Does your child experience asthma-like symptoms (wheeze/cough) at night?	Yes/ No
If your child has undergone a recent skin prick test, how many food/aeroallergens were they sensitised (positive) to?	__ (number of allergens or NA if test not done)

Figure E6 Parental questionnaire for collecting data needed for the CAPP tool

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	D;V Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	3
Abstract	2	D;V Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3
Introduction			
Background and objectives	3a	D;V Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6-7
	3b	D;V Specify the objectives, including whether the study describes the development or validation of the model or both.	7
Methods			
Source of data	4a	D;V Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	8
	4b	D;V Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	SM-4
Participants	5a	D;V Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7, SM-4
	5b	D;V Describe eligibility criteria for participants.	8
	5c	D;V Give details of treatments received, if relevant.	NA
Outcome	6a	D;V Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	8
	6b	D;V Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	7a	D;V Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	8, SM-14
	7b	D;V Report any actions to blind assessment of predictors for the outcome and other predictors.	8
Sample size	8	D;V Explain how the study size was arrived at.	8-10
Missing data	9	D;V Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9-10
Statistical analysis methods	10a	D Describe how predictors were handled in the analyses.	SM-5
	10b	D Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	9-10
	10c	V For validation, describe how the predictions were calculated.	11
	10d	D;V Specify all measures used to assess model performance and, if relevant, to compare multiple models.	10
	10e	V Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D;V Provide details on how risk groups were created, if done.	NA
Development vs. validation	12	V For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	11, SM-4,18

Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	13, SM-4
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	SM-20
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	SM-28
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	13-15
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	29
	15b	D	Explain how to use the prediction model.	2
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	28-30
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	NA
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	18,20
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	20-21
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	21
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	2
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

SM – supplementary material, NA – not applicable.