Nigerian Journal of Technology (NIJOTECH)
Vol. 44, No. 3, September, 2025, pp.450 - 460
www.nijotech.com

Print ISSN: 0331-8443
Electronic ISSN: 2467-8821
https://doi.org/10.4314/njt.v44i3.9

AN ENSEMBLE ASTHMA PREDICTION MODEL FOR EARLY DETECTION OF ASTHMA IN CHILDREN USING STACKING AND BLENDING

AUTHORS:
R. H. Ekpo[1, *], A. A. Azeta[2], O. J. Adebayo[3], A. I. Akinrinlola[4], O.S Aderibigbe[5] and B. Ewa[6]

AFFILIATIONS:
[1,4,5] Department of Computer Sciences, Lagos State University of Science and Technology, Nigeria,
[2]Department of Software Engineering, Namibia University of Science and Technology, Namibia
[3]Department of Computer and Mathematical Sciences, Crawford University, Nigeria,
[6]Landmark Hospital, Km 2 Itokin Rd, Ikorodu Lagos State, Nigeria

***CORRESPONDING AUTHOR:**
Email: ekpo.rh@lasustech.edu.ng

ARTICLE HISTORY:
Received: May 31, 2025
Revised: August 8, 2025
Accepted: August 18, 2025
Published: October 15, 2025

KEYWORDS: Asthma Prediction model, Asthma in Children, Blended Stacking, Ensemble Model, Grid Optimization.

DATA AVAILABILITY:
On request from author(s)

EDITORS:
Chidozie Nnaji

FUNDING:
None

Abstract
Asthma disease is a serious worldwide health challenge affecting every age bracket, particularly amongst children. Its widespread has extended to numerous countries. Childhood asthma remains underdiagnosed and insufficiently treated in Nigeria, affecting more than 20,000 children, including adults. It was discovered that there is a notable incidence of wheezing, with an approximation of 13 million or more individuals suffering from asthma.
The objective of this study is to build an ensemble asthma prediction model for early detection of asthma in children using stacking and blending. The methodology involved combining several techniques, including Relief feature selection, grid search optimization, and a blended stacking ensemble model. This study used the Nigerian dataset comprising the Hospital Record System (HRS) of six Nigerian hospitals and the administrative data from patients' history. The findings indicate that the proposed model attained a notably high precision of 96%, a recall rate of 94%, and an F1-score of 95% for the non-asthmatic category (0). Furthermore, it achieved a precision of 96%, a recall of 97%, and an F1-score of 97% for the asthmatic positive category (1). The study concludes that the model demonstrates effective capability in distinguishing between asthmatic and non-asthmatic patients. Consequently, this could contribute to improved patient outcomes and enhanced healthcare delivery systems. This work proposes further exploration, such as external validation, incorporating more diverse datasets and enhancing interpretability or explainability of the model.

1.0 INTRODUCTION
Asthma is a disease that consists of different components, normally distinguished by persistent inflammation of the airways. Asthma is characterized by respiratory symptoms like chest tightness, breath shortness, wheezing and coughing that undergo changes with time and potency along with fluctuating exploratory airway restriction [1]. Asthma is a serious worldwide health challenge affecting every age bracket, particularly among children. Its widespread reach has extended to numerous countries [2].

In a multi-centre research project carried out in Nigeria, more than 20,000 children, including adults, were found to have a notable incidence of wheezing,

HOW TO CITE:
Ekpo R. H., Azeta A. A, Orelaja A. E. Adebayo O. J. Akinfolajimi A. I., Adenigbe O. S., and Ewa B (M.D). "An Ensemble Asthma Prediction Model for Early Detection of Asthma in Children Using Stacking and Blending.", *Nigerian Journal of Technology* 2025; 44(3), pp 450 – 460; *https://doi.org/10.4314/njt.v44i3.9*

with an estimate of 13 million or more individuals suffering from asthma [3]. Paediatric asthma remains underdiagnosed and insufficiently treated in Nigeria. An ideal public healthcare approach is required to recognize and target disadvantaged communities that suffer from poor asthma outcomes in order to boost caregivers' asthma knowledge and management practices. [4, 5].

Also highlighted are the various challenges encountered in managing asthma in Nigeria, including diagnostic challenges that ultimately result in undiagnosed, over-diagnosis, underdiagnosis, and misdiagnosis of asthma cases. Treatment challenges ultimately result in reduced control, avoidable under-treatment, reduced quality of life, increased adverse reactions to medicine, increased illness, and in some cases, death. Follow-up challenges ultimately result in higher rates of asthma attacks and successive recurrent visits to emergency rooms, with missed follow-ups leading to reduced control of asthma.

Recent studies on asthma prediction models for children by [1] and [6] indicate that there is a moderate rise in the utilization of machine learning models for pediatric asthma prediction, and that these models have displayed higher predictive performance in childhood asthma than the traditionally available approaches. According to [7], the primary intention of machine learning is to have a greater understanding of the data, to find the connection between the features that depend on other features and independent features, and ultimately estimate a value. Although machine learning demands additional observed data to learn, it tends to be more effective and faster than conventional methods with reduced limitations based on the number of assumptions.

Furthermore, machine learning is also known to provide better performance in speed, accuracy, along with computational cost in several environmental research fields. Generally, the capability and success of machine learning results are determined by the characteristics and nature of the input data, as well as the performance of the learning algorithms. Pediatric asthma prediction models are useful for identifying probable future asthmatic patients among at-risk children. Preschool children who develop asthma symptoms may benefit from early diagnosis and intervention. [8].

According to Zhou *et al*. [9-11], machine learning also displays exceptional capability for use in severe childhood asthma attacks.

Sills *et al*. [12] indicated that machine learning algorithms improve the predictive capability for hospitalization needs of asthmatic patients when compared to the traditional available risk scores and logistic regression approaches.

Another recent work on children's asthma also demonstrated the effectiveness of ML methods in discovering asthma phenotypes and promises to be a good predictive tool [13]. Clinical decision tools can leverage integrated machine learning (ML) models to predict asthma exacerbations using real-world outpatient EHR data. These models have the potential to enhance outpatient care and help prevent severe outcomes.

A rapid increase in medical ML applications has been stimulated by a large number of datasets obtained from electronic health records, medical records, diagnostic examinations, imaging, wearable devices, and patient inventories. As machine-driven dataset gathering becomes more common in regular healthcare, machine-learning models demonstrate the ability to predict disease progression or pathways [15]. There are few studies on asthma in Nigeria, but none of them demonstrated the machine learning approach in their work. This study is necessary because childhood asthma remains underdiagnosed and inadequately treated in Nigeria [4, 5].

A study conducted by [16] focuses on constructing predictive models for children's asthma occurrence, using a dataset of asthma and non-asthma cases. The feature selection was done using the Chi-squared test, and the predictive models of support vector machines, decision trees, logistic regression, and random forests were also utilized. The study shows the superiority of the random forest model over all the other predictive models, as well as the identification of pediatric asthma prenatal and maternal risk factors.

According to [17], the aim of the work was to present an evaluation of a predictive model for persistent asthma in children by constructing a successful and accurate prediction model. The work utilizes the blending of Principal Component Analysis and a Logistic Bayesian approach on the collected medical dataset. The authors' work was able to predict asthma with improved accuracy and, in addition, provides good information on how each asthma prediction factor influences the condition.
Hogan et al. [18] intended to utilize the customary statistical approaches of Cox proportional hazards and logistic regression, and the ANN model for

identification, and to emphasize differences in risk factors and the performance of the model for exactly 180-day childhood asthmatic readmissions obtained from a nationwide representative dataset. The results indicated that nine risk factors were linked with hospital readmissions, while four important risk factors were associated with readmissions using the customary statistical models. The authors concluded that relying solely on customary modeling leaves unnoticed the major risk factors needed for readmission, and the compound relationships recognized by the machine learning model.

Bose et al. [19] aimed to build machine-learning models capable of recognizing children diagnosed with asthma earlier than the age of five and still continuing to participate in hospital asthma-associated visits. Considering the medical information of children diagnosed at the age of 5, five machine-learning models, namely Random Forest, Naïve Bayes, XGBoost, k-nearest neighbor, and logistic regression, were trained to distinguish transient asthmatic children from persistent diagnosed ones. Among the five machine-learning models, XGBoost shows the best performance. The most significant attributes selected for the models are the total number of asthma-associated visits, eczema, self-identified Black race, age of last asthma diagnosis under 5, and allergic rhinitis.

Wang et al. [20] focused on using a dataset of Medicaid children asthma patients. The main goal was to use the claims data of asthma patients who were enrolled in Medicaid to develop an asthma prediction model using deep learning techniques to predict the emergency risk of asthma-related outpatient care before the end of three months, as well as to evaluate whether deep learning models would perform better than the traditional Lasso logistic regression model. The outcome showed that deep learning models performed better than the conventional Lasso regression model. The drawback of this study is that for the model to be used effectively in patient healthcare, there must be collaboration between clinicians and model developers for proper interpretation and understanding, so that stakeholders can have confidence in the model.

Kim et al. [21] focused on evaluating the use of deep learning models to predict classes of asthma exacerbation risk, using peak flow rate data from a total of 14 asthmatic children. The results indicated that the deep learning model outperformed the traditional multinomial logistic regression model. A drawback of the study is that the limited sample size may have introduced bias into the results.

Goto et al. [22] focused on investigating the performance of machine learning techniques in predicting health outcomes and the general attitude of children in the emergency department (ED). The study also explored the similarities between machine learning models and traditional triage methods. The models used included deep neural networks, random forest, lasso logistic regression, and gradient boosting decision tree algorithms. The results showed that machine learning approaches outperformed traditional triage methods in predicting hospitalization and critical outcomes, offering improved capability to predict hospital outcomes and the mood of children in the ED.

According to [23], this work's primary focus was to investigate the likelihood of predictive treatment success in paediatric asthmatic patients after a period of 6 months and to uncover the major attributes used to understand the essential workings of such observations by Random Forest and AdaBoost classifiers. The study results showed that machine learning predicted asthma control and FENO-based outcomes are more accurate than both FEV1 and MEF50 in paediatric asthma treatment success. Asthma severity, as well as total IgE, were the most significant predictive features for paediatric asthma patient treatment success. This study aims to build an ensemble asthma prediction model for the early detection of asthma in children using stacking and blending techniques for Nigerian children.

2.0 METHODOLOGY

2.1 Data Source

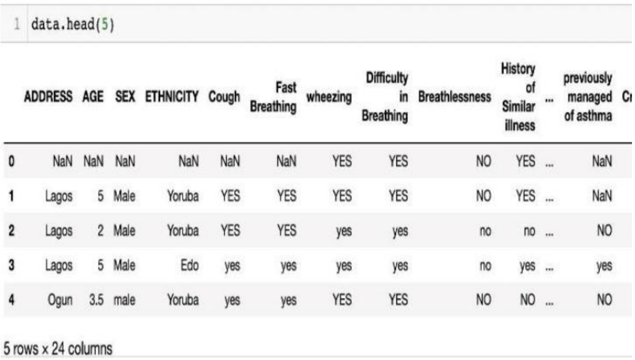
This study used the Hospital Record System (HRS) of six Nigerian hospitals and administrative data from patient histories. Records of patients aged 0 to 5 years were extracted from the HRS and administrative data. To maintain data confidentiality, patient IDs and hospital names were removed from the dataset, as shown in the exploratory section in Figure 1. The study utilized an in-depth research methodology to develop an effective asthma prediction model. This methodology combined several techniques, including Relief feature selection, grid search optimization, and a blended stacking ensemble model.

Following the data preprocessing phase, the Relief feature selection technique was employed to identify the most relevant variables for asthma prediction. This technique assigned weights to features based on their

https://doi.org/10.4314/njt.v44i3.9

discriminatory power in classifying asthmatic and non-asthmatic patients, thereby streamlining the feature selection process. Once the subset of informative features was determined, the ensemble was built using base models such as Decision Tree (DT), Random Forest (RF), and XG Boost

Each base model underwent grid search optimization to fine-tune its hyperparameters, with performance enhanced through iterative cross-validation. The final ensemble model was constructed using a blended stacking technique, in which a meta-learner integrated the predictions from the optimized base models. This meta-learner learned to effectively combine the predictions to produce the final output, as shown in Figure 2. The proposed methodology ensured the development of a robust and efficient asthma prediction model by leveraging the strengths of various machine learning techniques while effectively addressing challenges such as feature selection and hyperparameter optimization.



```
1 data.head(5)
```

	ADDRESS	AGE	SEX	ETHNICITY	Cough	Fast Breathing	wheezing	Difficulty in Breathing	Breathlessness	History of Similar illness	...	previously managed of asthma	C
0	NaN	NaN	NaN	NaN	NaN	NaN	YES	YES	NO	YES	...	NaN	
1	Lagos	5	Male	Yoruba	YES	YES	YES	YES	NO	YES	...	NaN	
2	Lagos	2	Male	Yoruba	YES	YES	yes	yes	no	no	...	NO	
3	Lagos	5	Male	Edo	yes	yes	yes	yes	no	yes	...	yes	
4	Ogun	3.5	male	Yoruba	yes	yes	YES	YES	NO	NO	...	NO	

5 rows x 24 columns

Figure 1: Head section of the Nigerian Hospital Record System dataset

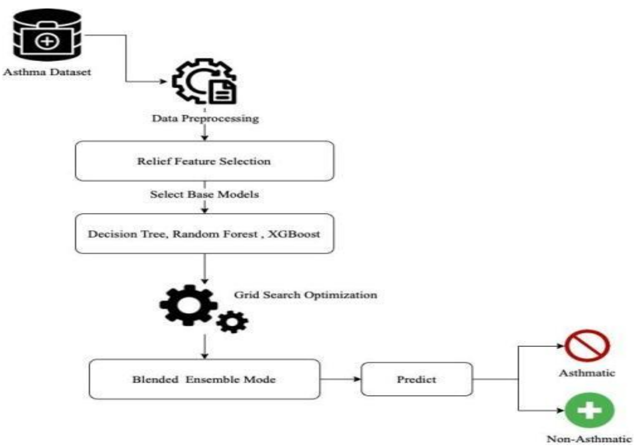


Figure 2: The proposed ensemble asthma prediction model.

The datasets obtained from six hospitals in Lagos, Nigeria, contain twenty-four (24) features—comprising twenty-three attributes and one class label (asthma or non-asthma)—with a total of 2,436 samples. Table 1 presents the features of the dataset.

Table 1: Features of the Dataset

S/N	Features
1	Address
2	Age
3	Sex
4	Ethnicity
5	Coughing
6	Fast Breathing
7	Wheezing
8	Difficulty in Breathing
9	Breathlessness
10	History of Similar illness
11	Family history of Asthma
12	Nebulization with Salbutamol
13	Rhonchi
14	Nasal flaring
15	previously managed of asthma
16	Crepitation
17	Atopy
18	Fever
19	Triggers
20	Exposure to dust
21	Exposure to cold
22	Exposure to strenuous exercise
23	Sneezing
24	Label

2.2 Data Processing Phase of the Proposed System

The data preprocessing phase includes the following steps:

a. ***Data Cleaning:*** This involves detecting and handling missing values and outliers, as well as resolving any inconsistencies or errors present in the raw asthma dataset.

b. ***Data Integration:*** This fundamental step combines data from various sources into a unified and larger data repository, such as a data warehouse. In this study, integrating asthma datasets from multiple healthcare institutions was essential to create a comprehensive dataset. During this phase, redundant attributes were eliminated from all data sources, and efforts were made to identify and resolve conflicts in data values.

c. ***Data Transformation:*** After data cleaning and integration, it was necessary to convert the high-quality data into a consistent and standardized format through data normalization. This process involves transforming data values, structure, or

https://doi.org/10.4314/njt.v44i3.9

format to ensure uniformity. Normalization allows for fair comparisons and helps eliminate biases associated with differing measurement scales.

In this study, numerical attributes were rescaled to fit within a predefined range using min-max normalization. This technique transforms each feature's minimum value to 0 and the maximum value to 1, scaling all other values proportionally between 0 and 1. For example, consider a feature with a minimum value of 50 and a maximum value of 70. A value of 60 would be transformed to approximately 0.5 because it is halfway between the minimum and maximum values. The formula used for this transformation is provided in Equation 1.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

d. ***Splitting the Dataset:*** The asthma dataset was partitioned into a ratio of 70% for the training set and 30% for testing set. A major part of the data was assigned to the training set. The training set was used for training and initial validation while the testing set was kept for testing or generalization purposes [24-25]

2.3 Feature Selection

In this study, the Relief feature selection approach was employed to identify the most relevant features from the dataset based on their importance. This technique assigns weights to features according to their ability to distinguish between classes in the data. It evaluates feature importance by examining how relevant each feature is to the target variable (asthma prediction).
For every instance in the dataset, the Relief method compares feature values between instances belonging to the same class (asthmatic or non-asthmatic) and those from different classes. Features that consistently show significant differences between these classes are considered more relevant and are assigned higher weights.

Once the importance scores are calculated, Relief ranks the features according to their weights. The top-ranked features—considered most informative for asthma prediction—are then selected for inclusion in the proposed ensemble prediction model designed for early detection of asthma in children using stacking and blending techniques.

The Relief approach is not only computationally efficient but also sensitive to subtle patterns in the data. This makes it particularly effective in retaining informative attributes that might otherwise be discarded prematurely in downstream modeling processes [26].

In a recent study by [27], a filter-based feature selection method—despite involving a larger number of attributes—was shown to outperform other methods in predicting heart disease. It achieved this using standard machine learning metrics such as accuracy, F-measure, and precision. The Relief feature was selected for this study because of its ability to enhance the effectiveness of the proposed ensemble asthma prediction model. By isolating the most relevant features from the dataset, it reduces noise, improves model interpretability, and boosts the overall performance and usefulness of the model in healthcare prediction scenarios.

The selected feature indices and corresponding attributes are presented in Table 2.

Table 2: Selected Features and Indices

Indices	**Features**
1	Age
2	Sex
3	Coughing
4	Fast Breathing
5	Difficulty in Breathing
6	Wheezing
7	Breathlessness
8	History of Similar illness
9	Family history of Asthma
10	Nebulization with Salbutamol
11	Rhonchi
12	Crepitation
13	Label

2.4 Base Model of the Proposed System

The study employed three machine learning algorithms to implement the proposed asthma prediction model. The proposed ensemble asthma prediction model is based on Decision Tree (DT), Random Forest (RF), and XGBoost.

https://doi.org/10.4314/njt.v44i3.9

Decision Tree: This is one of the straightforward machine learning techniques that categorizes data points derived from sequences of decision rules. It is suitable for disease continuation and treatment follow-up prediction in healthcare [28]. This is also a supervised learning technique that demonstrates improved efficiency and effectiveness when applied to large datasets. It achieves this by evaluating each feature's information gain and entropy within the dataset. DT optimizes its decision-making process by splitting the features based on greater information gain and lesser entropy. Information gain and the entropy of every feature are computed using Equation 2.

$$E = H(T) = I_E(k_1, k_2, \dots, k_n) = -\sum_{i=1}^{n} k_i k_i \quad (2)$$

In Equation 2, k denotes the probability of class i in the dataset, and *log log 2* represents the logarithm with base 2. The entropy denoted as E or *H(T)* is a measure of impurity that quantifies the level of uncertainty in a dataset. A lower entropy value indicates a less impure or more homogeneous dataset. An efficient feature exhibits high information gain, which leads to a decrease in entropy. The uncertainty within a set of instances of the asthma dataset is represented by the entropy, denoted as E or *H(T)*. The entropy is determined by summing the products of probabilities k_i and their logarithms k_i for all possible outcomes i within the feature. The summation is performed from $I = 1$ to n, where n represents the number of distinct outcomes or classes within the feature.

$$IG(F_i) = H(C) - H(F_i) \quad (3)$$

where $IG(F_i)$is the information gain, H(C)represents the entropy of the class, and $H(F_i)$ is the conditional entropy of class. Equation 3 is used to calculate the information gain of feature F_i within the asthma dataset. The equation involves the classification classes, denoted as C, and the features, represented by F_i. The base classifier utilized in this case was the DT. The cluster size was set to 100. The DT possesses important parameters, including the confidence factor and the choice between "pruned" or "unpruned" trees. In this study, the confidence factor was set to 2.5, the certainty factor to 2.5, and an "unpruned" tree was used as a parameter. The remaining parameters were left as their default values. Those are the DT Scikit-learn implementation pre-set hyperparameters that were not adjusted during the process.

Random Forest: This is an influential supervised ML algorithm consisting of various decision trees used for both classification and regression problems. The majority of votes from decision tree predictions are used by Random Forest when solving a classification problem. Conversely, for regression problems, the final result is obtained by finding the average of the individual tree outputs. [29]. In the asthma prediction model training phase, a distinct training asthma dataset T_i is created for every tree using samples from the original asthma training dataset, T.

To produce all tree splits, a subset of m attributes is randomly selected, and a measurement is applied to decide which attributes should be utilized for the split. Due to this randomization, several trees are generated, which typically lead to better prediction performance when combined.
Random forest provides outstanding performance on several real-world problems, it performs fast and when compared with other algorithms based on trees, it shows significant performance progress. Random Forest provides outstanding performance on several real-world problems. It performs quickly, and when compared with other algorithms based on trees, it shows significant performance improvements. Random Forest is not prone to overfitting and is not affected by noise present in the dataset used. These characteristics make it useful in some application areas [30-31] and in numerous agricultural applications, as indicated by [32-33], including predicting the performance of path loss in the metropolitan environment [34].

In the machine learning domain, Random Forest has become a widely used classification technique applied across several problem areas, including variable importance analysis, classification, prediction, outlier detection, and feature selection. Its performance can be enhanced by combining attribute evaluator methods with instance filter methods, which improves the accuracy of each individual classifier while reducing the correlation between them [35–36].

XGBoost: This is a machine learning algorithm designed to optimize gradient tree boosting by developing decision trees in a stepwise manner. It demonstrates faster computational efficiency across various computing environments. XGBoost is a boosting algorithm that utilizes Classification and Regression Trees (CART), primarily for classification tasks. Unlike traditional classification methods, the leaf nodes of CART trees in XGBoost represent actual scores rather than fixed categories. This feature supports the implementation of an efficient optimization process.

https://doi.org/10.4314/njt.v44i3.9

Beyond its popularity, XGBoost is recognized for its strong performance in a wide range of regression and classification tasks. When using K CART trees, the final classification result is derived from the combined contributions of all the trees. XGBoost also offers several configurations to minimize overfitting and enhance the overall performance of asthma prediction models. One such configuration involves adding regularization to the loss function, resulting in the objective function shown in Equation 4.

$$X^{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (4)$$

The objective function in the t^{th} iteration is expressed in Equation 4. In this Equation, $X^{(t)}$ denotes the objective function, while n represents the number of predicted asthma instances, $L(y_i, \hat{y}_i)$ represents the training error of the i^{th} sample, and Ω represents the regularization function.

2.5 Grid Optimization

The grid search hyperparameter tuning is used to tune the hyperparameters of the proposed model by exhaustively searching through a specified subset of the hyperparameter space. When applied to a blended stacking model, grid search optimization helps to find the optimal combination of hyperparameters for each base model as well as for the meta-learner. For each of the base models (DT, RF, XGBoost), a grid of hyperparameters is defined. For instance, parameters like maximum depth, number of estimators, and learning rate may vary for DT, RF, and XGBoost, respectively. The grid search optimization then exhaustively evaluates the performance of each model across different combinations of hyperparameters using cross-validation.

According to Rashidi et al. [23], cross-validation provides an approximation of the model's performance on an unseen dataset and is a basic step in model selection and the process of hyperparameter tuning in the machine learning domain. Cross-validation also allows us to recognize if a particular model is underfitting or overfitting by comparing the performance on the training set and validation test set. Such cross-validation provides a more rigorous approximation of the true capability of the model by averaging its statistical performance across various train-test splits of the dataset. Hyperparameter tuning using cross-validation activities reduces the chance of overfitting and enhances the generalization performance of the models. Hence, the outcome is high accuracy across the models.

Moreover, according to [37], grid optimization search was applied to Random Forest and Support Vector algorithms, along with the Relief feature selection approach. The study demonstrated the effectiveness of feature selection on combined algorithms using grid search optimization, and the performance was good.

The combination of hyperparameters that yields the best performance is then selected as the optimal configuration for each base model. Similarly, hyperparameters for the meta-learner are defined, and grid search is employed to explore the hyperparameter space for the meta-leaner using the predictions from the base models as input features. Once the optimal hyperparameters for each base model and the meta-learner have been identified, the ensemble asthma prediction model is constructed using these configurations.

The predictions from the base models are combined according to the weights determined by the meta-learner, resulting in the final blended prediction. Grid search optimization is employed in this study because it ensures that the blended stacking ensemble model is fine-tuned to achieve the best possible performance by systematically exploring the hyperparameter space. This optimization process helps maximize the model's predictive accuracy, generalization to new data, and overall effectiveness in enhancing asthma prediction and patient care outcomes

2.6 Proposed Ensemble Asthma Prediction Model

The proposed ensemble asthma prediction model is a sophisticated ensemble learning approach that merges the predictions of several single models to generate a more accurate and robust prediction. This approach is utilized in this study because it is effective in dealing with complex healthcare datasets. In the ensemble asthma prediction model, several machine learning algorithms are included in the ensemble. In this study, DT, RF, and XGBoost were included in the ensemble. Once the base models are trained, their predictions are used as input features for a meta-learner, often known as a blender or meta-model.

This meta-learner takes the predictions of the base models as input and learns to combine them to make a final asthma prediction. The proposed ensemble asthma prediction models offer several advantages including increased predictive accuracy, improved generalization to new data, and better resilience to overfitting. Hence, the ensemble asthma prediction model can improve asthma prediction and ultimately enhance patient care and outcomes.

https://doi.org/10.4314/njt.v44i3.9

3.0 RESULTS AND DISCUSSION

This section discusses the results of the proposed ensemble asthma prediction diagnosis model. To evaluate the performance of the proposed model, standard evaluation indicators must be established to assess the significance of the implemented model. Hence, standard machine learning evaluation techniques (i.e., accuracy, misclassification rate, precision, recall, F1-score, confusion matrix, Area Under the Curve (AUC)) were used to evaluate the model. A. Classification Report of the Model.

The classification report of the proposed blended stacking model is shown in Table 3. The classification report displays the performance evaluation metrics, which include the model's precision, recall, and F1 score. The results in Table 4 reveal that the proposed model achieved considerably high precision (96%), recall (94%), and F1-score (95%) for the non-asthmatic class (0), and a precision of (96%), recall of (97%), as well as an F1-score of (97%) for the asthmatic positive class (1).
The classification report also shows that 194 patients belong to the non-asthmatic class, while 294 belong to the asthmatic class. These results suggest that the model performs well in accurately identifying both asthmatic and non-asthmatic patients.

Table 3: Classification Report

	Precision	Recall	F1-score	Support
0	96%	94%	95%	294
1	96%	97%	97%	294
accuracy			96%	488
macro avg	96%	96%	96%	488
weighted avg	96 %	96%	96%	488

The confusion matrix in Figure 3 reveals that the model correctly predicts non-asthmatic patients as non-asthmatic, accurately identifying 183 such cases. However, it incorrectly classifies 11 non-asthmatic patients as having asthma. Additionally, 8 asthmatic patients are misclassified as non-asthmatic. On the other hand, the model correctly identifies 286 patients with asthma.

The confusion matrix shows that the model performs reasonably well in diagnosing asthma. However, there is still room for improvement in reducing the misclassification rate. Enhancing the model's accuracy and sensitivity is essential for ensuring precise diagnoses and optimal asthma management, ultimately leading to better patient outcomes and more efficient healthcare delivery.

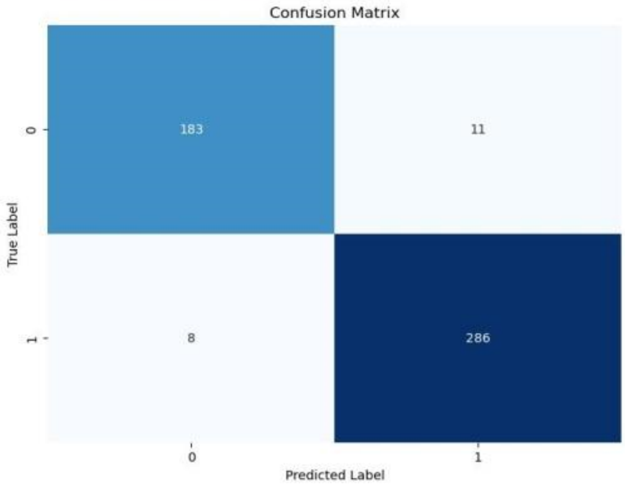


Figure 3: Confusion matrix

The ROC curve in Figure 4 graphically represents the model's performance across various classification thresholds, plotting the true positive rate against the false positive rate. The Area Under the Curve (AUC) quantifies this performance, with values ranging from 0 to 1. The model achieved an AUC score of 96%, indicating excellent discriminatory ability. A score closer to 1.0 suggests strong performance, while a score near 0.5 implies random guessing.

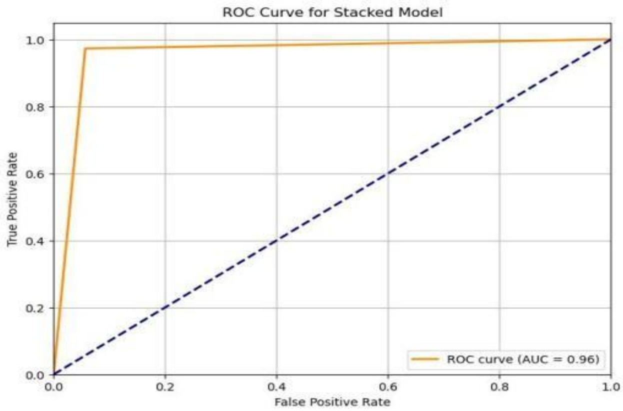


Figure 4: ROC-AUC Score

An AUC of 96% means the model is highly effective at distinguishing between asthmatic and non-asthmatic patients. In other words, it has a high probability of ranking a randomly selected asthmatic patient higher than a non-asthmatic one. This high ROC-AUC score underscores the model's robustness and reliability, boosting confidence in its diagnostic capabilities. It supports healthcare professionals in making informed decisions regarding patient care, treatment strategies, and resource allocation. Overall,

https://doi.org/10.4314/njt.v44i3.9

the ROC-AUC result confirms the model's efficacy and dependability in asthma classification tasks, providing valuable insights for clinical decision-making.

Table 4 presents the results of each base model, benchmarking the proposed model against the baseline models.

Table 4: Benchmarking of the Proposed Model with Baseline Models

Baseline Models	**Accuracy**	**Precision**	**Recall**	**F1-score**
DT	96.1%	96 %	96%	96%
RF	96.1%	96%	96%	96 %
KNN	95.0%	94%	96%	95 %
NB	76%	77%	78%	70%
XGBoost	96.1%	96%	96%	96%
Proposed Model	96.1%	96%	96%	96%

4.0 CONCLUSION

This paper presents an ensemble asthma prediction model for the early detection of asthma in children using stacking and blending. It utilizes the Hospital Record System (HRS) of six hospitals in Lagos, Nigeria, and the administrative data from patients' histories. Records of patients between 0 to 5 years old were extracted from the HRS and administrative data from patients' histories. Data exploration was conducted on the collected asthma dataset. Data preparation, such as data cleaning, data integration, and transformation, was performed to detect and handle missing features, outliers, and eliminate redundant features, as well as to improve the quality of the dataset.

The dataset was subjected to a relief feature selection technique to select a relevant feature dataset based on their importance. The study employed three machine learning algorithms to implement the proposed asthma prediction model. The proposed ensemble asthma prediction model for the early detection of asthma in children, using stacking and blending, was based on Decision Tree (DT), Random Forest (RF), and XGBoost. The grid search hyperparameter tuning is used to tune the hyperparameters of the proposed model by exhaustively searching through a specified subset of the hyperparameter space. When applied to an ensemble asthma prediction model, grid search optimization helps to find the optimal combination of hyperparameters for each base model, as well as for the meta-learner.

Standard machine learning evaluation techniques (i.e., accuracy, misclassification rate, precision, recall, F1-score, confusion matrix, Area Under the Curve (AUC)) were used to evaluate the model. The study concludes that the model performs exceptionally well in distinguishing between asthmatic and non-asthmatic patients, and that the ensemble blended stacking asthma prediction model worked better for patient outcomes and healthcare efficiency.

According to Tohka and Van Gills [38], the work enumerated customer usability, approach, explainability, etc. as factors that influence algorithm acceptance in healthcare practice. This study acknowledged these concerns, but the aim of this study was to build an ensemble asthma prediction model for the early detection of asthma in children using stacking and blending. This work is significant because there are few studies on asthma in Nigeria, but none of them demonstrated the use of a machine learning approach. This work proposed further exploration, such as external validation, incorporating more diverse datasets, or enhancing the interpretability and explainability of the asthma models.

Authors' Contribution: Conceptualization, investigation and writing-original draft was done by RE. Supervision was done by AA. Methodology was done by OA. Visualization was done by AA. Methodology was done by OA, and data curation was done by BE. Writing - review and editing of the work was done and approved by all the authors.

REFERENCES

[1] Ekpo, R. H., Osamor, V. C., Azeta, A. A., Ikeakanam, E. and Amos, B. O. "Machine learning classification approach for asthma prediction models in children," *Health and Technology*, 13 (1), pp. 1–10, 2023. 10.1007/s12553-023-00732-8.

[2] Reddel, H. K, Leonard, B. B, Batman, D. et al. "Global Initiative for Asthma Strategy 2021: Executive Summary and Rationale for Key changes," *American Journal of Respiratory and Critical Care Medicine*, 205 (1), 17–35, 2021. 10.1164/rccm.202109-2205pp.

[3] "The Global Asthma Report," *Global Asthma Network*, 2022, [Online]. Available: http://www.globalasthmareport.org/

[4] Akinso, O., Adhikari, A., Yin, J., Chopak-Foss J. and Shah G. "Childhood Asthma-Management Practices in Rural Nigeria: Exploring the

knowledge, attitude, and practice of caregivers in Oyo State," *Children*, 10 (6) p 1043, 2023. 10.3390/children10061043.

[5] Onyedum, C. C., Ukwaja, K. N., Desalu O. O. and Ezeudo C. "Challenges in the management of bronchial asthma among adults in Nigeria: a Systematic review," *Annals of Medical and Health Sciences Research,* 3(3), 324-329, 2013.

[6] Patel, D., Hall, G. L., Broadhurst, D., Smith, A., Schultz, A.and Foong, R. E. "Does machine learning have a role in the prediction of asthma in children?," *Paediatric Respiratory Reviews*, 41, pp. 51–60, 2021. 10.1016/j.prrv.2021.06.002

[7] Razavi-Termeh, S. V., Sadeghi-Niaraki, A. and Choi, S. M. "Asthma-prone areas modeling using a machine learning model," *Scientific Reports*, 11 (1), 2021. 10.1038/s41598-021-81147-1.

[8] De Jesus Romero-Tapia, S., Becerril-Negrete,J. R., Castro-Rodriguez, J . A and Del-Río-Navarro, B. E. "Early prediction of asthma," *Journal of Clinical Medicine*, 12 (16), p. 5404, 2023. 10.3390/jcm12165404

[9] Zhou, C., Shuai, L., Hu, H., Ung, C.O.L., Lai, Y., Fan, L., Du, W., Wang, Y. and Li, M. "Applications of machine learning approaches for pediatric asthma exacerbation management: a systematic review". *BMC Medical Informatics and Decision Making*, 25(1), p.170, 2025. https://doi.org/10.1186/s12911-025-02990-0

[10] Ojha, T., Patel, A., Sivapragasam, K., Sharma, R., Vosoughi, T., Skidmore, B., Pinto, A.D. and Hosseini, B. "Exploring machine learning applications in pediatric asthma management: scoping review". *Jmir Ai*, *3*, p.e57983, 2024. https://dx.doi.org/10.13005/bpj/3089

[11] Darsha Jayamini, W.K., Mirza, F., Asif Naeem, M. and Chan, A.H.Y. " Investigating machine learning techniques for predicting risk of asthma exacerbations: a systematic review." *Journal of Medical Systems*, *48*(1), p.49, 2024. https://doi.org/10.1007/s10916-024-02061-3

[12] Sills, M.R., Ozkaynak, M. and Jang, H. "Predicting Hospitalization of Pediatric Asthma Patients in Emergency Departments using Machine Learning". *International Journal of Medical Informatics*, 151, p.104468. Jul. 1, 2021. https://doi.org/10.1016/j.ijmedinf.2021.104468.

[13] Cilluffo, G., Fasola, S., Ferrante, G., Licari, A., Marseglia, G.R., Albarelli, A., Marseglia, G.L. and La Grutta, S. "Machine learning: A modern approach to pediatric asthma," *Pediatric Allergy and Immunology*, 33 (S27), pp. 34–37, 2022. 10.1111/pai.13624

[14] Zein, J.G., Wu, C.P., Attaway, A.H., Zhang, P. and Nazha, A. "Novel machine learning can predict acute asthma exacerbation," *CHEST Journal*, 159 (5), pp. 1747–1757, 2021. 10.1016/j.chest.2020.12.051.

[15] Filipow, N., Main, E., Sebire, N.J., Booth, J., Taylor, A.M., Davies, G. and Stanojevic, S "Implementation of prognostic machine learning algorithms in paediatric chronic respiratory conditions: a scoping review," *BMJ Open Respiratory Research*, 9(1) p e001165, 2022. 10.1136/bmjresp-2021-001165.

[16] Jeddi, Z., Gryech, I., Ghogho, M., El Hammoumi, M. and Mahraoui, C. ."Machine learning for predicting the risk for childhood asthma using prenatal, perinatal, postnatal and environmental factors," *Healthcare*, 9(11) p. 1464, 2021. 10.3390/healthcare9111464.

[17] Spyroglou, I.I., Spöck, G., Chatzimichail, E.A., Rigas, A.G. and Paraskakis, E.N. "A Bayesian Logistic Regression approach in Asthma Persistence Prediction," *Epidemiology Biostatistics and Public Health*, 15 (1), 2022. 10.2427/12777.

[18] Hogan, A.H., Brimacombe, M., Mosha, M. and Flores, G."Comparing artificial intelligence and traditional methods to identify factors associated with pediatric asthma readmission," *Academic Pediatrics*, 22(1) pp. 55–61, 2021. 10.1016/j.acap.2021.07.015

[19] Bose, S., Kenyon, C.C.and Masino, A.J. "Personalized prediction of early childhood asthma persistence: A machine learning approach," *PLoS ONE*, 16 (3), p. e0247784, 2021. 10.1371/journal.pone.0247784

[20] Wang, X., Wang, Z., Pengetnze, Y.M., Lachman, B.S. and Chowdhry, V. "Deep learning models to predict pediatric asthma emergency department visits," *arXiv (Cornell University)*, 2019, 10.48550/arxiv.1907.11195.

[21] Kim, D., Cho, S., Tamil, L., Song, D.J. and Seo, S. "Predicting Asthma attacks: Effects of indoor PM concentrations on peak expiratory flow rates of asthmatic children," *IEEE Access*, 8, pp 8791–8797, 2019. 10.1109/access.2019.2960551.

[22] Goto, T., Camargo, C.A., Faridi, M.K., Freishtat, R.J. and Hasegawa, K. "Machine learning–based prediction of clinical outcomes for children during emergency department triage.," *JAMA Network Open*, 2(1), pp. e186937- e186937, 2019, 10.1001/jamanetworkopen.2018.6937

[23] Lovrić, M., Banić, I., Lacić, E., Pavlović, K., Kern, R. and Turkalj, M. ."Predicting Treatment Outcomes Using Explainable Machine Learning

in Children with Asthma," *Children*, 8 (5) p. 376, 2021. 10.3390/children8050376

[24] Rashidi, Hooman H., Samer Albahra, Scott Robertson, Nam K. Tran, and Bo Hu. "Common statistical concepts in the supervised Machine Learning arena." *Frontiers in Oncology,* 13, p.1130229, 2023. 10.3389/fonc.2023.113022

[25] Rainio, O., Teuho, J. and Klén, R. "Evaluation metrics and statistical tests for machine learning." *Scientific Reports*, 14(1), p.6086, 2024. https://doi.org/10.1038/s41598-024-56706-x

[26] Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S. and Moore, J.H. "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, 85, pp 189–203, 2018. 10.1016/j.jbi.2018.07.014.

[27] Noroozi, Z., Orooji, A. and Erfannia, L. "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," *Scientific Reports*, 13 (1), pp. 22588, 2023. 10.1038/s41598-023-49962-w.

[28] Olalekan Kehinde, A. "Leveraging machine learning for predictive models in healthcare to enhance patient outcome management." *International Research Journal of Modernization in Engineering Technology and Science*7(1),p.1465.2025. https://www.doi.org/10.56726/IRJMETS66198

[29] Belyadi, H. and Haghighat, A. "Machine learning guide for oil and gas using Python: A step-by-step breakdown with data, algorithms, codes, and applications. Supervised Learning Gulf Professional Publishing 2021 . https://doi.org/10.1016/B978-0-12-821929-4.00004-4.

[30] Emanet, N., Öz, H.R., Bayram, N. and Delen, D. , "A comparative analysis of machine learning methods for classification type decision problems in healthcare," *Decision Analytics*, 1 (1), pp.1-6, 2014. 10.1186/2193-8636-1-6.

[31] Patil, M.S. and Patil, H.D., "Towards an automatic pain intensity levels evaluation from multimodal physiological signal using machine learning approaches" *Nigerian Journal of Technology*, 43(4), pp.763-771, 2024. 10.4314/njt.v43i4.16

[32] Nandom, S.S., Abe, G.T. and Gambo, I. P. "Application of random forest and hierarchical clustering models for crop and fertilizer recommendation to farmers," *Nigerian Journal of Technology*, 44 (1), pp. 114–122, 2025. https://doi.org/10.4314/njt.v44i1.13.

[33] Vemulapalli, S., Sri, M.S., Varshitha, P., Kumar, P.P. and Vinay, T., "An experimental analysis of machine learning techniques for crop recommendation." *Nigerian Journal of Technology* , 43(2), pp.301-308, 2024. 10.4314/njt.v43i2.13

[34] Shaibu, F.E., Onwuka, E.N., Salawu, N. and Oyewobi, S.S. "Evaluating the effectiveness of machine learning models for path loss prediction at 3.5 GHz with focus on feature prioritization." *Nigerian Journal of Technology*, *43*(4), pp.754-762., 2024. 10.4314/njt.v43i4.15

[35] Yao, D., Yang, J. and Zhan, X. "A novel method for disease prediction: hybrid of random forest and multivariate adaptive regression SPlines," *Journal of Computers*, 8 (1), 2013. 10.4304/jcp.8.1.170-177.5z7

[36] Chaudhary, A., Kolhe, S. and Kamal, R. "An improved random forest classifier for multi-class classification," *Information Processing in Agriculture*, 3 (4), pp. 215–222, 2016. 10.1016/j.inpa.2016.08.002

[37] Simbolon, I.N. and Naibaho, R. "Influence of relief feature selection on random forest and support vector machine classification algorithm," *International Workshop on Big Data and Information Security,* pp. 27–32, 2022. 10.1109/iwbis56557.2022.9924782.

[38] Tohka, J. and Van Gils, M., "Evaluation of machine learning algorithms for health and wellness applications: A tutorial." *Computers in Biology and Medicine*, *132*, p.104324, 2021 https://doi.org/10.1016/j.compbiomed.2021.104324

[39] Bari, P. and Ragha, L. "Predicting wheat yield in the agricultural industry using deep learning techniques: a review". *Nigerian Journal of Technology*, *43*(4), pp.716-737.,2024, doi: https://doi.org/10.4314/njt.v43i4.12

[40] Ajayi, O.T. and J. O. Olusina. "Vehicle detection, tracking, counting and classification using deep learning." *Nigerian Journal of Technology 43*(4), pp.738-742. 2024. 10.4314/njt.v43i4.12

https://doi.org/10.4314/njt.v44i3.9