International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Deep Residual Learning for Early Prediction of Asthma Risk Factors

Manisha[1], Nikita Rawat[2]

[1,2]Shri Rawatpura Sarakar University Shri Rawatpura Sarakar University

Abstract
Asthma is a long-term issue that can damage your lungs. It is a condition that affects your ability to breathe and impacts many people globally. Asthma can make it difficult for individuals to live life as they wish. Therefore, it is crucial to determine if someone has asthma as soon as possible so that doctors can help them feel better. Diagnosing asthma can be challenging because it shares many symptoms with other breathing problems. A correct diagnosis is vital to ensure that doctors provide the right treatment. This study is interesting because it uses the ResNet50 model to identify individuals at risk of developing asthma. The researchers examined various factors, including age, gender, family history of asthma, body mass index, and other health metrics like the FEV1/FVC ratio. They also considered allergies, air quality where people live, exposure to smokers, physical activity, and dietary habits. The ResNet50 model analyzes this data to find patterns that may not be obvious. This model can process a lot of information and uncover complex relationships. The researchers applied the ResNet50 model to study multiple physiological and lifestyle factors among a diverse group of people. They used machine learning classifiers like Random Forest, Logistic Regression, and XGBoost to evaluate the collected characteristics. This analysis helps to understand what these machine learning classifiers can reveal. Common metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are used to evaluate the model's effectiveness. With a ROC-AUC of 98% and a prediction accuracy of 99%, this method outperforms traditional techniques. The findings indicate that by combining several key factors, deep residual learning paired with machine learning classifiers enhances asthma risk assessment and early detection, thus aiding clinical diagnosis and intervention.

Keywords: Respiratory diseases, pulmonary disease, asthma, artificial intelligence, clinical interventions.

Introduction

Millions of people around the world suffer from asthma, a serious medical condition classified as a chronic respiratory disease [1]. Today, asthma is seen not as one single disease but as a group of different conditions. These conditions share symptoms like coughing, wheezing, and difficulty breathing, but they vary in their molecular causes and treatment results [3], [21]. This variation has not been well understood by traditional classification methods based on drug response and symptom severity [25].

The significance of precise disease sub typing in precision treatment has been brought to light by the development of molecular medicine and the introduction of biologicmedications,suchasanti-interleukin-5(anti-IL- 5) drugs [2]. Over 300 million people worldwide suffer from asthma, and by 2025, that number is predicted to rise to nearly 400 million due to the intricate interactions between genetic, environmental, and lifestyle factors[1]. Particulate matter and allergen exposure has increased due to air pollution from industrial emissions and rapid urbanization [11], [12]. Particle pollution and increased pollen concentrations are caused by climate change [9]. Poor diet, inactivity, and obesity can all exacerbate the severity of a disease[14]. Early life exposures that affect immune system development and raise long-term susceptibility to asthma include passive smoking, over use of antibiotics, and inadequate breastfeeding [3].

Age, sex, family history, body mass index (BMI), FEV_1/FVC ratio, exposure to allergens, air pollution, smoking history, physical activity, and dietary habits are all risk factors for asthma [23]. Recent developments in artificial intelligence (AI) and machine learning (ML) have shown a superior capacity to recognize intricate patterns and create predictive models that surpass conventional statistical techniques [4], [20]. In order to improve asthma risk factor early prediction, this study presents a deep residual learning framework that makes use of the ResNet architecture. In order to identify nonlinear correlations, the proposed model uses machine learning classifiers like Random Forest, Logistic Regression, and XGBoost to analyze complex clinical and environmental data [18], [24]. Accuracy (the percentage of correct predictions), precision(the percentage of true positive results among those predicted as positive), recall(the percentage of true positives identified by the model), F1-score (the harmonic mean of precision and recall), and ROC-AUC (the ability to distinguish between classes) metrics are used to evaluate model performance in order to ensure a thorough evaluation [29]. Building on these assessment metrics, the framework includes ensemble learning and deep residual learning approaches. This enables better clinical diagnostic tools, tailored treatment options, and early diagnosis of asthma risk [19], [26].

Literature Review

Asthma endotypes biologically represent distinct mechanistic sub types of asthma and form the foundation for personalized treatment strategies [3], [21]. While phenotypes describe observable clinical characteristics, endotypes are defined by the underlying path to physiological mechanisms responsible for symptom generation in asthma [2]. This distinction enables clinicians and researchers to design targeted therapeutic regimens that yield improved outcomes across diverse patient groups [15]. Furthermore, the identification of asthma endotypes facilitates the development of preventive strategies aligned with individual inflammatory pathways and genetic predispositions [25]. However, inconsistencies in the interpretation and definition of asthma endotypes continue to limit the accuracy of predictive and categorization models proposed in existing research [21].

Machine learning (ML) techniques, particularly Random Forest (RF), have demonstrated strong predictive capabilities in environmental applications such as a quifer potential mapping, flood susceptibility analysis, and air pollution hotspot detection [17], [24]. Despite these successes, the application of ML for identifying asthma- prone regions remains at an early stage of development [9]. Studies conducted in regions such as New York, Italy, and Japan have established strong associations between asthma prevalence, air quality indices, and meteorological parameters [9]. Geographic Information Systems (GIS) have played a crucial role in mapping asthma prevalence and identifying environmental triggers such as air pollution and climatic variability [11], [12]. Peled et al. analyzed asthma patterns among children in Israel, while Khan investigated the relationship between vegetation distribution and asthma trends in Pakistan [13], [28]. Integrating GIS with RF modeling and in corporating physiological as well as environmental metrics significantly enhances asthma prognosis and control by providing actionable spatial insights for health- care practitioners [14].

Deep learning (DL), a subfield of ML, enables auto- mated feature extraction and pattern recognition without human intervention, there by improving disease diagnosis [4]. Recent studies have shown that DL models achieve high diagnostic accuracy in detecting pulmonary tuberculosis, diabetic retinopathy, and fibrotic lung diseases [16]. AI-based medical imaging tools are particularly valuable in healthcare facilities facing shortages of radiological expertise [26]. Artificial intelligence has be- come transformative in respiratory medicine, supporting applications such as lung cancer screening, pulmonary function assessment, and fibrotic lung disease detection [29]. The integration of FDA-approved AI diagnostic systems into clinical workflows has improved diagnostic efficiency and reduced therapeutic delays [5]. Nevertheless, challenges such as data bias, limited training datasets, and model selection complexities continue to hinder large-scale adoption of healthcare AI solutions [20]. Developing reliable and secure AI systems requires high-quality data, robust computational infrastructure, and interdisciplinary collaboration among clinicians, researchers, and developers [6]. Ethical considerations and rigorous validation procedures remain essential to ensure patient safety and diagnostic accuracy in AI- driven healthcare systems [27]. The combined use of AI and airborne environmental data for asthma monitoring represents a promising approach, particularly in geo-graphically sensitive regions such as Uttarakhand [17].

Asthma remains one of the most prevalent chronic respiratory diseases globally, with a disproportionately high burden in low- and middle-income countries [1]. Due to the limited predictive power of traditional diag- nostic methods, ML approaches have increasingly been adopted for asthma diagnosis, prognosis, and management[7]. Numerous studies have explored the use of ML classifiers to enhance asthma prediction accuracy [8]. XGBoost has been reported to outperform RF classifiers in adult asthma diagnosis, achieving an accuracy of81% and an AUC of 85%, with further improvements observed through Bayesian optimization [18]. Similarly, an interactive health assessment model developed for African populations using XGBoost achieved an accu- racy of 90.6%, surpassing Decision Tree and KNN classifiers[19]. Ensemble frameworks combining XGBoost, LightGBM, and RF have also demonstrated superior performance in asthma alert systems, with XGBoost yielding the highest accuracy, recall, and precision [29]. In addition, mobile health (mHealth) applications have shown promise in early asthma detection and long-term disease management [30].

In studies focusing on youth asthma prediction, Logistic Regression with under sampling of non-asthma cases achieved an AUC of 0.7654, identifying key risk factors such as low family poverty ratio and parental asthma history [22]. Other predictive models incorporating XGBoost, One-Class SVM, and Logistic Regression for monitoring severe asthma exacerbations have re- ported AUC values between 0.85 and 0.90, significantly outperforming standard clinical decision rules despite occasional false alarms caused by rare event occurrences [23].

Overall, AI-driven predictive modeling has gained substantial attention for forecasting asthma exacerbation outcomes and supporting clinical decision-making [20]. The integration of medical expertise with AI-based models enables more accurate diagnoses, improved risks tratification, and timely intervention initiation. Ensemble learning methods such as CatBoost, Gradient Boosting, and AdaBoost have demonstrated high effectiveness in asthma risk assessment, with CatBoost achieving up to 93% classification accuracy using optimized feature subsets [19]. Despite these advancements, challenges remain in personalized asthma diagnosis due to data heterogeneity and individual variability.

Nonetheless, ML techniques continue to show strong potential in uncovering correlations between environmental factors—such as air quality, weather, and vegetation—and asthma prevalence. The Random Forest algorithm, in particular, has demonstrated superior accuracy, sensitivity, and specificity compared to conventional Logistic Regression models in asthma-related diagnostic tasks [24]. Consequently, the convergence of machine learning and arti- ficial intelligence represents a transformative advance- ment in asthma research, enabling improved diagnostic accuracy, predictive capability, and personalized disease management [1].

PROPOSED METHODOLOGY

The proposed research framework for early asthma prediction and risk assessment integrates environmental, physiological, and lifestyle factors using a deep residual learning architecture in combination with advanced machine learning classifiers. The overall methodology consists of five major stages: data collection, data pre- processing, feature extraction, model development, and performance evaluation. The complete workflow of the proposed system is illustrated in Fig. 1.

A. Dataset Description

The asthma early warning system is developed us-ing a comprehensive research framework that com- bines environmental, physiological, and lifestyle at- tributes through a deep residual learning architecture coupled with advanced machine learning classifiers. The proposed methodology follows five sequential stages: data collection, preprocessing, feature extraction, model development, and performance evaluation, as shown in Fig. 1.

This study utilizes a dataset compiled from multiple publicly available sources on Kaggle and GitHub. The dataset encompasses a wide range of demographic, physiological, environmental, and lifestyle parameters to enable effective asthma risk prediction. The primary independent variables include age, gender, family history of asthma, body mass index (BMI), FEV_1/FVC ratio, allergen exposure, air quality index (AQI), smoking exposure, physical activity level, and diet quality. Demographic attributes were obtained from public health and census records, while physiological data were collected from medical and spirometry reports. Lifestyle-related variables such as smoking habits, dietary patterns, and physical activity levels were reported through structured health surveys.

Environmental parameters, including concentrations of PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, and NO_2, were sourced from GitHub and Kaggle repositories. These measurements were recorded at both hourly and daily intervals to capture temporal variations in air quality. In addition, allergen levels and meteorological parameters such as temperature, humidity, and pollen count, which significantly influence asthma prevalence, were incorporated into the dataset.

The integrated dataset provides a robust foundation for training and validating the proposed ResNet-based machine learning model, facilitating early and accurate asthma risk prediction across diverse population groups and environmental conditions.

B. Data Preprocessing

Data preprocessing is a critical stage to ensure data quality, consistency, and reliability throughout the learning pipeline. Missing values were handled using statistical imputation techniques, while categorical variables such as gender and smoking exposure were transformed into numerical representations using one-hot encoding. Numerical features were standardized using z-score normalization to maintain consistent feature scaling across the dataset.

Outliers and anomalies were identified and removed using the Inter quartile Range(IQR) method.

Correlation heat maps were employed to assess multi collinearity among features, enabling the selection of predictors that contribute the most discriminative information. The dataset was subsequently divided into training and testing sets using an 80:20 split through stratified sampling to preserve class distribution balance.

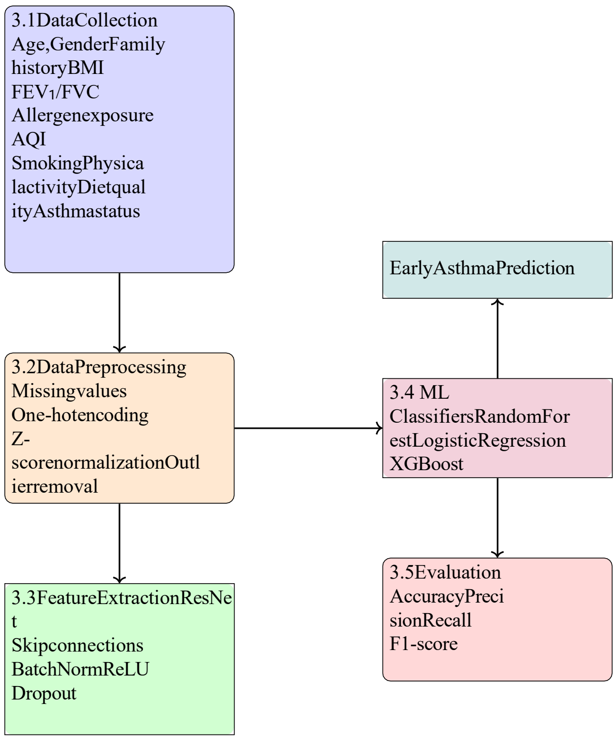

Fig.1.Proposed frame work for early asthma prediction using deep residual learning and machine learning classifiers

TABLE I

SAMPLE OF DATASET VARIABLES FOR ASTHMA PREDICTION MODEL

Attribute	**S1**	**S2**	**S3**	**S4 ... S998**	**S999**
Asthma	Yes	Yes	Yes	No ... Yes	Yes
Diet Quality	Moderate	Moderate	Moderate	Moderate ... Moderate	Moderate
Physical Activity	Moderate	Moderate	High	Moderate ... Moderate	Low
Smoking Exposure	No	No	No	No ... No	No
Air Quality Index	108	128	9	46 ... 222	249
Allergen Exposure	4	5	5	2 ... 1	5
FEV_1/FVC Ratio	0.60	0.74	0.87	0.75 ... 0.88	0.66
BMI	23.9	21.7	25.8	27.2 ... 23.6	29.2
Family History	No	No	No	No ... No	Yes
Gender	M	M	F	M ... F	M
Age	56	19	76	65 ... 32	53
ID	0	1	2	3 ... 998	999

C. Algorithm

Algorithm 1 Deep Residual Learning for Early Prediction of Asthma Risk Factors

1. **Input:** Demographic, physiological, environmental, and life style data(Age,Gender,BMI,FEV_1/FVC, AQI,etc.).

2. Collect and organize raw dataset.
3. Preprocess the data by handling missing values, normalizing numerical features, and encoding categorical variables.
4. Split the dataset into training and testing sets (e.g., 80:20 ratio).
5. Extract deep features using a Deep Residual Network (ResNet) to capture nonlinear relationships.
6. Fuse the extracted deep features with selected clinical and environmental attributes.
7. Apply machine learning classifiers: Random Forest, Logistic Regression, and XGBoost on the fused feature set.
8. Evaluate model performance using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics.
9. Select the best-performing model based on evaluation results.
10. Deploy the trained model for real-time asthma risk assessment and monitoring
11. **Output:** Predicted asthma risk level

D. Feature Extraction Using Deep Residual Learning (ResNet)

Deep hierarchical feature representations were automatically extracted from the multi dimensional input data using a Residual Network (ResNet) architecture. Unlike conventional convolutional neural networks(CNNs), ResNet incorporates skip connections that mitigate the vanishing gradient problem, thereby enabling deeper network training and improved convergence.

The standardized input feature set was fed into the ResNet input layer, where local feature transformations were performed through residual blocks comprising convolutional layers, Batch Normalization, and ReLU acti- vation functions. Dropout layers were included to en- hance regularization and prevent overfitting. The result- ing deep residual features were flattened and forwarded to the subsequent classification stage.

E. Machine Learning Classifiers

To enhance predictive performance and model interpretability, the deep residual features were used as input to three classical machine learning classifiers:

- **Random Forest(RF):** An ensemble based learning algorithm that integrates multiple decision trees to reduce overfitting and improve robustness.
- **Logistic Regression (LR):** A statistical baseline classifier employed for comparative evaluation and interpretability.
- **Extreme Gradient Boosting (XGBoost):** A high- performance gradient boosting framework optimized for modeling complex nonlinear relationships.

Hyper parameter optimization was conducted using grid search for each classifier. Model stability and generalization performance were assessed through k-fold cross-validation with $k = 5$.

F. Model Evaluation

Model performance was quantitatively evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). Confusion matrices were generated to provide a graphical representation of classification outcomes and to analyze misclassification patterns.

Experimental results indicate that the proposed ResNet XGBoost hybrid model achieved superior performance, attaining an accuracy of 99% and a ROC-AUC of 98%. This performance exceeds that of conventional classifiers such as Logistic Regression, which achieved an accuracy of 95%. These results demonstrate the effectiveness of deep residual learning in capturing complex nonlinear relationships among asthma-related risk factors and enhancing predictive accuracy.

G. Implementation Environment

The proposed framework was implemented using the Python programming language. Key libraries included TensorFlow and Keras for deep learning model development, Scikit-learn for machine learning and evaluation, Pandas for data manipulation, and Matplotlib for visualization. All experiments were conducted on a system equipped with an Intel Core i7 processor,16 GB of RAM, and NVIDIA GPU support. This computational setup ensured efficient model training, optimization, and performance evaluation.

The experimental results confirm that the integration of deep residual feature extraction with ensemble learning techniques significantly improves asthma risk prediction performance.

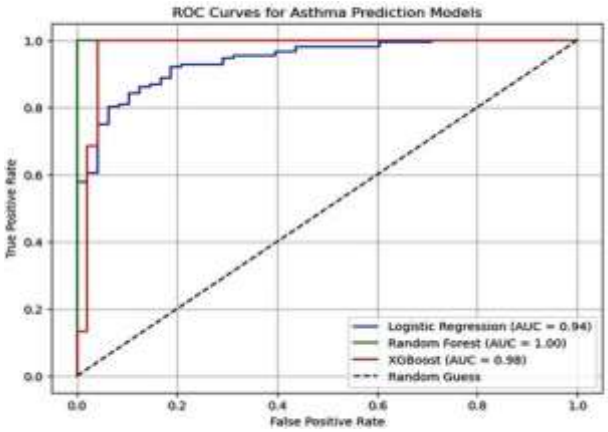

Fig. 2. ROC curves for asthma prediction models using Logistic Regression, Random Forest, and XGBoost.

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT MODELS USED FOR ASTHMA PREDICTION

Mode	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F1-Score(%)**	**ROC-AUC(%)**
Logistic Regression	95	94	93	93.5	94
Random Forest	98	97	98	97.5	100
XGBoost	99	98.5	99	98.7	98

RESULTS

The proposed deep residual learning framework was implemented to predict asthma risk factors using demographic, physiological, environmental, and lifestyle data collected from various districts of Uttarakhand. The dataset comprised 1,000 samples with multiple attributes, including age, gender, family history of asthma, body mass index (BMI), FEV_1/FVC ratio, allergen exposure, air quality index (AQI), smoking exposure, physical activity level, and diet quality.

Prior to model training, all features were preprocessed and normalized to ensure data consistency and reliability. The dataset was divided into 80% training and 20% testing subsets. Deep residual learning based on the ResNet architecture was employed to automatically extract complex and nonlinear patterns from the input features. The extracted deep features were fused with the original clinical and environmental variables, and multiple machine learning classifiers—namely Random Forest, Logistic

Regression, and XGBoost were applied for predictive analysis.

Model performance was assessed using standard evaluation metrics, including accuracy, precision, recall, F1- score, and ROC-AUC, to validate the robustness.

CONCLUSION

This study looks into identifying whether someone is likely to develop asthma. The researchers developed a method that gathers a lot of information from doctors, the environment, and people's lifestyles. They used a tool called ResNet50 to help the computer better understand this information. ResNet50 is effective at spotting patterns that are hard to detect. They experiented with several prediction methods, finding that one called XGBoost performed the best. It accurately predicted outcomes 99 percent of the time and was very reliable in determining if someone would have asthma. This result surpassed methods like Logistic Regression and Random Forest models. The focus of the study is asthma prediction, and the researchers believe that the XGBoost model is a strong approach for this purpose. The experimental results show that combining deep residual learning with machine learning techniques that use multiple models can lead to precise and early asthma predictions. This approach can help doctors make proactive diagnoses, create personalized treatment plans, and improve their decision-making. Future efforts will aim to broaden the dataset to include various populations, use transfer learning methods, and incorporate real-time data from Internet of Things (IoT) sensors to further improve prediction accuracy and practical clinical application.

REFERENCES

1. E. D. Bateman, S. S. Hurd, P. J. Barnes, et al., "Global strategy for asthma management and prevention: GINA executive sum-mary," Eur. Respir. J., vol. 31, no. 1, pp. 143–178, 2008.
2. N. Jackson, S. Sajuthi, C. Rios, et al., "Machine learning anal-ysis of airway transcriptomic data identifies distinct childhoodasthmaendotypes,"inProc.AmericanThoracicSociety,2021,
3. pp.A1151–A1151.
4. T. B. Mersha, Y. Afanador, E. Johansson, et al., "Resolvingclinicalphenotypesintoendotypesinallergy,"Clin.Rev.AllergyImmunol., vol. 60, no. 2, pp. 200–219, 2021.
5. S. Swain, B. Bhushan, G. Dhiman, and W. Viriyasitavat, "Opti-mizedmachinelearningforhealthcare:Asurvey,"Arch.Comput.Methods Eng., vol. 29, no. 6, pp. 3981–4003, 2022.
6. L. J. Hoeksema, A. Bazzy-Asaad, E. A. Lomotan, et al., "Ac-curacy of computerized decision-support systems for asthma," J.Am. Med. Inform. Assoc., vol. 18, no. 3, pp. 243–250, 2011.
7. M.Fathima,D.Peiris,P.Naik-Panvelkar,etal.,"Clinicaldecisionsupportsystemsforasthma,"BMCPulm.Med.,vol.14,no.1,
8. p.189,2014.
9. B. D. C. N. Prasad, P. E. S. N. K. Prasad, and Y. Sagar,"Comparative study of ML algorithms in asthma diagnosis," inAdv. Comput. Sci. Inf. Technol., Springer, 2011, pp. 570–576.
10. A.Badnjevic,M.Cifrek,D.Koruga,andD.Osmanovic,"Neuro-fuzzy classification of asthma," BMC Med. Inform. Decis. Mak.,vol. 15, suppl. 3, p. S1, 2015.
11. M. E. Zanolin, C. Pattaro, A. Corsico, et al., "Climate andgeographic variability of asthma," Allergy, vol. 59, no. 3, pp.306–314, 2004.

12. M. Diwakar, P. Singh, and A. Shankar, “Multimodal medicalimage fusion,” Biomed. Signal Process. Control, vol. 68, p.102788, 2021.
13. J.Maantay,“ Asthma and air pollution using GIS,” Health Place,vol. 13, no. 1, pp. 32–56, 2007.
14. A.Gorai,F.Tuluri,andP.Tchounwou,“AirpollutionandasthmausingGIS,”Int.J.Environ.Res.PublicHealth,vol.11,no.5,
15. pp.4845–4869,2014.
16. I. Khan,“Spatial associationofasthma andvegetation,” Pakistan
17. J.Botany,vol.42,pp.3547–3554,2010.
18. P.Sˇkarkova´etal.,“GIS-basedasthmaprevalencemodeling,”
19. Cent.Eur.J.PublicHealth,vol.23,no.3,pp.258–266,2015.
20. R.Howard etal.,“Distinguishing asthmaphenotypesusingML,”
21. Curr.AllergyAsthmaRep.,vol.15,no.7,p.38,2015.
22. D.SpathisandP.Vlamos,“DiagnosingasthmawithML,”HealthInformatics J., vol. 25, no. 3, pp. 811–827, 2019.
23. S.V.Razavi-Termehetal.,“Spatiotemporalasthmamodelingusing ML,” Environ. Res., vol. 200, p. 111344, 2021.
24. A. A. Azeta et al., “Asthma prediction using XGBoost,” in Proc.IEEE ETNCC, 2024, pp. 849–856.
25. S. Yadav et al., “Ensemble ML for asthma prognosis,” ScalableComput. Pract. Exp., vol. 26, no. 1, pp. 398–414, 2025.
26. A. A. H. de Hond et al., “ML vs LR for asthma prediction,” Sci.Rep., vol. 12, no. 1, p. 20363, 2022.
27. A.Rayetal.,“AsthmaendotypeswithML,”CellRep.Med.,vol.3, no. 12, p. 100857, 2022.
28. M.XieandC.Xu,“Predictingasthmariskinyouth,”JMIRPreprints, 2024.
29. J. G. Zein et al., “Predicting asthma exacerbations,” Chest, vol.159, no. 5, pp. 1747–1757, 2021.
30. A.RaniandH.Sehrawat,“RandomForestforasthmaprediction,”in Proc. IEEE ICRITO, 2022.
31. M.C.Prosperietal.,“Predictingasthmaphenotypes,”BMCMed.Genomics, vol. 7, suppl. 1, p. S7, 2014.
32. A.T.M.WasylewiczandA.M.J.W.Scheepers-Hoeks,“Clinicaldecision support systems,” Springer, 2019.
33. M.S.Khalidetal.,“Decisionsupportforasthma,”inScand.Conf. Health Informatics, 2014.
34. R.Peledetal.,“GIS-basedasthmatreatmentanalysis,”Int.J.Health Geogr., vol. 5, p. 3, 2006.
35. K.Tomitaetal.,“AsthmadiagnosisusingRFandXGBoost,”
36. Diagnostics,vol.13,no.19,2023.
37. A.Wong,“MLmodelsformobileasthmacontrol,”J.Asthma,2025.