world

MDPI

Article

An Interpretable and Generalizable Machine Learning Model for Predicting Asthma Outcomes: Integrating AutoML and Explainable AI Techniques

Salman Mahmood [1], Raza Hasan [2,*], Saqib Hussain [3] and Rochak Adhikari [2]

[1] Department of Science and Engineering, Nazeer Hussain University, ST-2, Near Karimabad, Karachi 75950, Pakistan; salman.mahmood@nhu.edu.pk
[2] Department of Computer Science, Solent University, Southampton SO14 0YN, UK; 0adhir74@solent.ac.uk
[3] Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8QH, UK; saqib2.hussain@northumbria.ac.uk
* Correspondence: raza.hasan@solent.ac.uk

Abstract: Asthma remains a prevalent chronic condition, impacting millions globally and presenting significant clinical and economic challenges. This study develops a predictive model for asthma outcomes, leveraging automated machine learning (AutoML) and explainable AI (XAI) to balance high predictive accuracy with interpretability. Using a comprehensive dataset of demographic, clinical, and respiratory function data, we employed AutoGluon to automate model selection, optimization, and ensembling, resulting in a model with 98.99% accuracy and a 0.9996 ROC-AUC score. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) were applied to provide both global and local interpretability, ensuring that clinicians can trust and understand model predictions. Additionally, counterfactual analysis enabled hypothetical scenario exploration, supporting personalized asthma management by allowing clinicians to assess potential interventions for individual patient risk profiles. To facilitate clinical adoption, a Streamlit v1.41.0 application was developed for real-time access to predictions and interpretability. This study addresses key gaps in asthma prediction, notably in model transparency and generalizability, while providing a practical tool for enhancing personalized care. Future research could expand the validation across diverse patient populations to reinforce the model's robustness in broader clinical environments.

Keywords: asthma prediction; machine learning; AutoML; explainable AI; SHAP; LIME; counterfactual analysis; healthcare predictive modeling; personalized medicine; Streamlit application

Academic Editors: Euteum Choi, Joontaek Oh and Seongjin Lee

Received: 27 November 2024
Revised: 24 December 2024
Accepted: 9 January 2025
Published: 14 January 2025

Citation: Mahmood, S.; Hasan, R.; Hussain, S.; Adhikari, R. An Interpretable and Generalizable Machine Learning Model for Predicting Asthma Outcomes: Integrating AutoML and Explainable AI Techniques. *World* **2025**, *6*, 15. https://doi.org/10.3390/world6010015

1. Introduction

Asthma is a chronic respiratory condition that affects over 262 million people worldwide and leads to considerable morbidity, recurrent hospitalizations, and substantial healthcare costs [1]. Managing asthma effectively is challenging due to the variability in patient responses and the condition's sensitivity to both clinical and environmental factors [2]. Consequently, there is an increasing need for advanced tools that can predict asthma exacerbations and guide personalized treatment interventions [3]. Such predictive tools could significantly improve patient outcomes by enabling early intervention and customized management strategies tailored to individual risk profiles [4].

Machine learning (ML) has shown significant promise in predictive modeling for healthcare, especially for complex conditions like asthma [5]. Recent studies have demon-

strated that ML models can uncover patterns within multidimensional datasets and identify at-risk patients more accurately than traditional statistical methods [6]. For instance, predictive models have successfully utilized diverse clinical and demographic factors, such as breathing difficulties, allergies, and early-life health indicators, to forecast asthma risk in pediatric and adult populations [7]. However, most of these models prioritize accuracy without addressing interpretability, which is critical in clinical settings where transparency and trust are essential. Models that lack explainability are often regarded as "black boxes", making it challenging for clinicians to understand and validate their predictions, thus limiting their clinical acceptance and practical utility [8].

AutoML frameworks have emerged as a powerful solution, democratizing access to ML by automating model selection, hyperparameter tuning, and optimization tasks [9]. AutoML allows healthcare practitioners with limited ML expertise to develop robust models with minimal intervention, potentially enabling the broader adoption of predictive modeling in clinical environments [10]. However, while AutoML has shown substantial benefits in terms of model development speed and predictive performance, challenges persist. These include difficulties in handling imbalanced and high-dimensional healthcare data as well as limitations in the interpretability tools available within some AutoML platforms. Additionally, models trained on isolated datasets often lack generalizability and perform inconsistently across diverse patient populations and healthcare settings [11].

XAI offers a means to address interpretability challenges in predictive healthcare models, enabling clinicians to understand the rationale behind ML model predictions. XAI techniques such as SHAP and LIME have gained prominence for providing global and local insights into model decision-making processes [12]. SHAP allows for a comprehensive assessment of feature importance across a dataset, highlighting how specific clinical and demographic factors influence asthma predictions [5]. In contrast, LIME focuses on instance-specific explanations, making it possible to interpret individual predictions, which is particularly valuable in patient-centered care [12]. Recent developments in hybrid XAI methods, which combine multiple interpretability techniques, further underscore the potential of XAI to enhance the transparency of complex AI-driven models [13].

While XAI techniques are becoming more common in ML applications, the use of counterfactual analysis, a method that allows clinicians to simulate hypothetical "what-if" scenarios, remains underexplored in asthma prediction [14]. Counterfactual analysis can be instrumental in personalized medicine by enabling healthcare providers to model how specific interventions or behavioral changes might influence a patient's risk profile [15]. This approach can provide valuable insights for preventive care, especially in chronic diseases like asthma, where small adjustments in medication, lifestyle, or environmental factors can have a significant impact on patient outcomes. Thus, the integration of counterfactual analysis with predictive models could offer actionable insights, further enhancing the clinical utility of AI-driven asthma prediction [16].

Despite advancements in ML and AutoML for healthcare applications, most predictive models for asthma have been validated within isolated datasets, limiting their generalizability. Generalizable models are essential to ensure consistent performance across various populations and clinical settings [17]. For instance, ML models validated exclusively on pediatric cohorts or in specific healthcare institutions may fail to generalize to broader populations due to differences in demographic characteristics, environmental factors, or treatment protocols. Addressing this limitation requires rigorous cross-validation and external validation across diverse datasets, which can help mitigate bias and improve model robustness.

This study uniquely integrates AutoML, SHAP, LIME, and counterfactual analysis to bridge the gap between high predictive accuracy and interpretability. Unlike previous

studies, which often focused solely on statistical improvements or theoretical aspects, our approach provides actionable insights into individual patient risk profiles. For instance, counterfactual scenarios enable clinicians to explore the impact of medication adjustments on predicted outcomes, thereby fostering tailored asthma management strategies.

By focusing on the dual objectives of predictive power and transparency, this study aims to deliver a practical, clinician-friendly tool that supports proactive, patient-centered care. The integration of real-time usability through the Streamlit application further underscores the practicality of our model in clinical workflows.

The literature reveals several critical gaps that this study aims to address.

- While predictive accuracy has been the primary focus of many asthma prediction studies, few have incorporated explainable AI techniques to enhance model transparency. By integrating SHAP and LIME, this study aims to develop an interpretable model that allows healthcare providers to understand and trust predictive outputs.
- Existing models have often been validated on isolated datasets, limiting their applicability to diverse patient populations and healthcare systems. This study addresses this issue by focusing on model validation techniques and exploring ways to enhance robustness across different demographic and clinical contexts.
- Although counterfactual analysis offers valuable insights for intervention planning, its application in asthma prediction remains largely unexplored. This study will incorporate a counterfactual analysis to simulate hypothetical treatment scenarios, providing clinicians with actionable insights for personalized asthma management.

This study aims to develop an interpretable and generalizable AutoML-based predictive model for asthma outcomes, focusing on the likelihood of asthma-related hospital visits. By employing SHAP and LIME, the model will provide transparent predictions that enhance clinician trust and usability. Counterfactual analysis will enable the exploration of "what-if" scenarios, supporting personalized medicine approaches that align with patient-specific needs. By addressing these research gaps, this study contributes to the advancement of predictive modeling in asthma management, setting a foundation for the deployment of interpretable, actionable, and reliable AI tools in clinical environments.

The paper is structured as follows: Section 2 reviews current methods for asthma prediction and identifies existing gaps and opportunities. Section 3 details the methodology, including the integration of AutoML and XAI. Section 4 presents the experimental findings, followed by a discussion in Section 5, which explores the implications, limitations, and future directions. The paper concludes in Section 6 with key takeaways and suggestions for advancing interpretable and reliable predictive models for healthcare.

2. Literature Review

This literature review aims to explore key studies and advancements in the field of predictive modeling for asthma and related respiratory conditions. It highlights the role of machine learning, AutoML, and explainable AI techniques, such as SHAP, LIME, and counterfactual analysis, in improving asthma prediction models. Additionally, the review identifies gaps in the current research, including challenges related to model generalizability, interpretability, and real-world deployment. By examining these areas, this review provides a comprehensive overview of state-of-the-art methodologies and their implications for healthcare predictive modeling.

2.1. Predictive Modeling in Asthma and Respiratory Healthcare

Machine learning models have shown promise in predicting asthma outcomes, enabling a more accurate assessment of high-risk patients by identifying complex patterns in patient data. These studies highlight the critical role of ML in early intervention and

asthma management. Despite significant advances in predictive accuracy, the integration of explainability remains a challenge. Studies such as [18,19] emphasize the importance of transparent AI systems in healthcare. These models often fail to address the need for deployment-ready solutions or the integration of XAI methods, leaving gaps in their usability for clinicians. Table 1 summarizes the key studies that focus on predictive modeling of asthma, highlighting their major findings and limitations.

Table 1. Key studies on predictive modeling in asthma.

Study	Focus	Key Findings	Limitations
[20]	Pediatric asthma hospitalization prediction	Improved accuracy over traditional methods using ML	Limited interpretability; lacks transparency
[21]	Childhood asthma risk factors	Random forest model identified breathing difficulty and allergies as key predictors	Emphasis on accuracy, with limited exploration of explainability
[22]	Ensemble model for asthma exacerbation prediction	Ensemble learning showed potential for individualized asthma treatment	Complex model structure; challenges in clinical application
[23]	Systematic review of asthma prediction methods	Identified the need for standardized approaches in asthma prediction	Lack of unified methods limits model comparability across studies
[24]	Early-life non-biological factors for pediatric asthma	Non-invasive factors like maternal asthma, atopy, and antibiotic exposure can predict asthma onset	Limited generalizability; data mainly from pediatric cohort
[25]	Affinity graph-enhanced classifier using biomarkers	Blood biomarkers showed predictive potential but struggled with overfitting due to a small sample size	Limited dataset size and complex structure increase overfitting risk

2.2. *Automated Machine Learning (AutoML) for Clinical Predictions*

AutoML frameworks have democratized access to machine learning in healthcare, allowing clinicians to develop predictive models with minimal expertise. By democratizing access to machine learning, AutoML frameworks lower barriers for healthcare practitioners who lack technical expertise. This ensures that advanced predictive models can be developed and deployed without the need for in-depth programming or data science knowledge. For instance, tools like AutoGluon simplify the modeling process, enabling broader participation in AI-driven healthcare innovations. This democratization not only promotes inclusivity, but also accelerates the adoption of AI solutions in resource-limited settings, where technical expertise may be inadequate [9,10]. By automating tasks like model selection, hyperparameter tuning, and optimization, AutoML reduces the technical barrier for practitioners, fostering broader adoption of advanced machine learning tools in clinical environments. However, while AutoML has shown substantial benefits in model development speed and predictive performance, challenges persist. These include difficulties in handling imbalanced and high-dimensional healthcare data, as well as limitations in interpretability tools available within some AutoML platforms. Challenges related to data quality and computational constraints also remain. Table 2 summarizes studies that explore the role of AutoML in healthcare, with a focus on its efficiency and limitations.

One significant method for addressing the interpretability challenges inherent in AutoML frameworks is the use of counterfactual analysis. This technique enables clinicians to explore hypothetical "what-if" scenarios, providing patient-specific insights that counterbalance the perceived opacity of AutoML models. For example, counterfactual analysis allows clinicians to simulate how changes in key variables, such as medication dosage or pulmonary function test results, might influence a patient's risk of asthma exacerbation.

This ability to test interventions virtually offers an actionable understanding of model outputs, thereby enhancing trust and usability in clinical settings.

Table 2. Key studies on AutoML applications in healthcare.

Study	Focus	Key Findings	Limitations
[26]	AutoML's accessibility in clinical settings	AutoML can rival traditional ML by automating complex tasks	Lack of interpretability tools in some AutoML systems
[10]	AutoML for model selection and hyperparameter tuning	AutoML frameworks enhance efficiency and reduce model development time	High computational demands for large datasets
[27]	AutoML applications in healthcare	Demonstrated AutoML's efficiency in model development, beneficial for non-expert clinicians	Limited data processing capabilities; may require additional data preprocessing
[28]	AutoML with large, complex healthcare datasets	AutoML can manage large datasets but faces challenges with high-dimensional data and imbalance	Requires preprocessing steps for data balancing and dimensionality reduction
[29]	AutoML for imbalanced datasets in clinical predictions	Highlighted the need for proper preprocessing for balanced datasets in AutoML	Imbalanced datasets and missing values can mislead AutoML performance
[30]	Checklist for selecting AutoML platforms in healthcare	Proposed criteria for evaluating AutoML tools, including interpretability and data quality	Checklist does not address deployment limitations in real-time clinical environments

Counterfactual scenarios provide intuitive explanations that are particularly valuable in patient-centered care. They help bridge the gap between complex black-box models and the actionable insights required for personalized treatment planning. Additionally, the ability to simulate specific interventions aligns with the goals of personalized medicine, in which treatment strategies are tailored to individual patient profiles. Recent studies have underscored the importance of counterfactual reasoning in enhancing the interpretability and clinical applicability of machine learning models, particularly in chronic disease management scenarios like asthma [14,15].

Another critical challenge in AutoML applications is the potential for bias arising from isolated datasets. Models trained on such datasets may fail to effectively generalize across diverse patient populations and healthcare settings. To address this limitation, rigorous statistical techniques are required.

Stratified sampling ensures that key subgroups within the dataset are proportionally represented in both the training and validation subsets, mitigating the risk of systematic bias. External validation—testing the model on independent datasets from diverse populations—is another essential step in evaluating robustness. For instance, this study validated its model using the Childhood Asthma Management Program (CAMP) dataset alongside the primary Kaggle dataset, ensuring its applicability across pediatric and general populations.

Future work should explore additional statistical approaches for bias mitigation, such as propensity score matching or domain adaptation methods. These techniques can further enhance the generalizability of AutoML models, ensuring consistent performance across various clinical environments.

The integration of interpretability techniques like counterfactual analysis, combined with rigorous validation protocols, highlights the potential of AutoML to deliver actionable, reliable, and generalizable AI tools for clinical decision-making. As AutoML continues

to evolve, addressing these challenges will be pivotal for translating machine learning advancements into real-world healthcare applications.

2.3. Explainable AI for Model Transparency

Explainable AI (XAI) addresses the "black-box" nature of many machine learning models by providing insights into predictions, thereby building trust in AI-driven healthcare applications. Table 3 outlines the significant studies in the area of XAI, particularly its application in clinical predictions and asthma.

Table 3. Key studies on XAI in healthcare.

Study	Focus	Key Findings	Limitations
[31]	Overview of XAI applications	XAI provides transparency by explaining AI decision-making processes	XAI implementation in clinical models remains limited
[32]	XAI for understanding model behavior	XAI helps predict model strengths, weaknesses, and future behavior	Emphasis on theoretical aspects with limited practical applications
[33]	Interpretability in clinical predictions	Interpretability builds trust, enabling clinicians to validate predictions	Focuses on local interpretability; lacks global feature insights
[34]	Rule-based explanations for asthma prediction	Rule-based models guide resource allocation in asthma care	Limited integration of XAI with more complex ML algorithms
[35]	XAI techniques: Saliency maps and SHAP	Categorizes XAI into interpretable models and post-hoc explanations for black-box models	Does not address hybrid XAI methods for deep learning
[36]	Hybrid explanation methods	Combining saliency maps, LIME, and SHAP offers comprehensive insights into complex models	Hybrid methods are computationally intensive, limiting their application in real-time predictions

The integration of XAI techniques like SHAP and LIME has proven to be highly beneficial in elucidating model decisions. For instance, SHAP offers global insights into feature importance, enabling clinicians to understand the overall influence of specific variables on model predictions. Conversely, LIME provides instance-level interpretability, helping explain individual predictions. However, studies such as [37] underscore the computational challenges associated with these methods, particularly in real-time applications.

Counterfactual analysis, although less commonly employed in asthma prediction, holds significant potential for personalized medicine. By enabling "what-if" scenario simulations, counterfactual analysis provides actionable insights that support clinicians in tailoring interventions to individual patient profiles.

2.4. Counterfactual Analysis in Healthcare Predictions

Counterfactual analysis provides actionable insights by simulating hypothetical scenarios, a valuable tool for understanding potential interventions in chronic conditions like asthma. Table 4 summarizes the studies on counterfactual analysis and its potential for asthma prediction and personalized medicine.

Table 4. Key studies on counterfactual analysis in healthcare.

Study	Focus	Key Findings	Limitations
[38]	Counterfactuals in asthma prediction	Hypothetical scenarios assist clinicians in evaluating potential interventions for reducing asthma risks	Limited data scope for counterfactual simulations
[39]	Counterfactuals in precision medicine	Allows for personalized treatment by exploring how changes in features might impact outcomes	Primarily theoretical; limited applications in practical healthcare

2.5. Deployment and Generalizability Challenges

Generalizability remains a challenge, as many ML models in healthcare have been developed using isolated datasets, which limits their applicability across different populations. Table 5 presents studies that highlight the challenges related to the generalizability of healthcare models across different populations and clinical environments.

Table 5. Key studies on generalizability in healthcare.

Study	Focus	Key Findings	Limitations
[40]	Generalizability in SEER-based analysis for metastasis prediction	Retrospective data analysis requires cross-validation for broader applicability	Dependence on imputed values may introduce bias
[41]	Generalizability challenges in acute kidney injury prediction	Highlights the need for diverse datasets to ensure robustness across healthcare settings	Retrospective design risks overfitting and lacks prospective validation
[42]	Early-life respiratory prediction	Early predictors must be validated across various demographics to account for genetic/environmental variations	Limited ethnic diversity in early prediction studies
[43]	Deployment issues for respiratory disease classification	Data security, privacy, and system integration are essential for deployment in real-world settings	Deployment readiness hindered by data security and integration challenges
[44]	Enhancing generalizability with preprocessing	Proposed methods to improve AutoML robustness in ICU mortality predictions through feature selection and SMOTE	Limited focus on cross-institutional model validation

AI adoption in clinical settings is often hindered by skepticism from healthcare providers and infrastructure limitations. Resource constraints, training requirements, and integration with existing electronic health record systems pose significant challenges. This study addresses these issues by developing a user-friendly Streamlit application and incorporating features to ensure its seamless integration into clinical workflows.

2.6. Contribution of the Study

This study surpasses prior work by combining the predictive capabilities of AutoML with XAI methods like SHAP and counterfactual analysis, providing a transparent and generalizable model for asthma management. The use of dual datasets—the Kaggle dataset for its extensive demographic and clinical details and the CAMP dataset for its longitudinal pediatric focus—ensures robustness across diverse populations, enabling the model to address variations in patient profiles and clinical settings. This comprehensive approach further enhances the model's clinical relevance and applicability in real-world healthcare scenarios.

3. Methodology

This study aims to develop an interpretable predictive model for asthma outcomes using AutoML and XAI techniques. The methodology consists of several steps, including data collection and preprocessing, model development using AutoML, and evaluation of model performance. Generalization is performed to ensure that the model can perform reliably across different patient populations and in real-world clinical settings. The integration of interpretability techniques further enhances transparency and clinical utility, while counterfactual analysis supports personalized intervention planning.

3.1. Data Collection and Preprocessing

3.1.1. Data Source

The dataset used in this study is publicly available on Kaggle (https://www.kaggle.com/datasets/jassonus122332/budesonide-nedocromil (accessed on 10 October 2024)), consisting of anonymized patient records related to asthma outcomes. It was selected for its comprehensive demographic and clinical data, which included a balanced sample of patients with and without asthma exacerbations. This balance ensures that that the model can learn from both positive and negative cases, contributina g to a robust prediction capability.

The Kaggle dataset includes several key features relevant to asthma management.

- Demographics: Data on age, gender, ethnicity, and socio-economic status.
- Clinical Measurements: Pulmonary function test results, such as pre-forced vital capacity (PREFVC), pre-forced expiratory volume (PREFEV), and post-forced peak flow (POSFP).
- Medication History: Information on dosages and frequencies of medications like Budesonide and Nedocromil.
- Environmental Factors: Limited data on exposure to allergens and air quality indices.

While the dataset offers valuable insights, it has some limitations.

- Urban Bias: The dataset is primarily derived from urban healthcare systems, which may not fully represent rural or underserved populations.
- Lack of Longitudinal Data: The absence of longitudinal data limits the ability to analyze temporal patterns of asthma exacerbations.
- Missing Environmental and Socio-economic Variables: The dataset lacks detailed data on environmental and socio-economic factors, which could impact the model's predictions.

To ensure the generalizability of the model, we complemented the Kaggle dataset with the Childhood Asthma Management Program (CAMP) dataset for external validation. The CAMP dataset provides longitudinal data on pediatric patients with asthma, offering a valuable dimension of diversity. It covers a broader range of demographic and clinical characteristics, allowing the model to be tested on pediatric cases, which introduces different clinical variables compared to the adult population in the Kaggle dataset. Integrating both datasets improves the model's reliability and accuracy across diverse patient groups.

While the Kaggle and CAMP datasets provide a solid foundation, expanding the dataset diversity is crucial. Incorporating additional data sources can enhance the model's generalizability and fairness in diverse clinical settings. Potential sources include the following:

- International Repositories: Datasets from regions like Asia, Africa, and South America to provide a global perspective on asthma outcomes.
- Rural Healthcare Records: Data from rural healthcare systems to address existing biases and improve the model's applicability to underserved populations.
- Specialized Cohorts: Datasets with specific populations, such as those with genetic predispositions or unique environmental exposures, would provide further insights into how these factors influence asthma outcomes.

Expanding the dataset diversity would allow the model to capture a broader range of socio-economic, environmental, and genetic factors, ensuring more equitable and accurate predictions. This approach is essential for the model's widespread clinical adoption, making it relevant to diverse patient populations and improving its overall fairness and effectiveness.

3.1.2. Feature Engineering

Feature engineering plays a critical role in enhancing the predictive performance of a model by selecting, transforming, and preparing relevant features for training. The process involves identifying clinically significant variables, addressing issues like class imbalance, and ensuring compatibility with machine learning algorithms. The following steps were undertaken to engineer the features of this study:

- Feature Selection: Initial exploratory data analysis was performed to identify clinically relevant features and exclude non-informative columns, such as unique identifiers, that do not contribute to the predictive power of the model. This step ensured that only the relevant variables related to asthma outcomes were included in the model.
- Clinical and Demographic Variables: The features retained for analysis included important clinical and demographic variables, such as age, gender, pulmonary function test results (e.g., pre-forced vital capacity [PREFVC] and pre-forced expiratory volume [PREFEV]), medication type and dosage (with a focus on Budesonide dosage), and clinical history. These features have been established as significant predictors in asthma-related studies, contributing to the model's ability to capture key factors influencing asthma outcomes.
- Encoding Categorical Variables: To prepare the categorical data for model training, Label Encoding was applied to the Ethnic variable using LabelEncoder from the sklearn library. This encoding transformed categorical ethnicity data into numerical values, enabling compatibility with machine learning algorithms. Other categorical variables, such as medication type, were similarly prepared to ensure seamless integration with the AutoML framework.
- Synthetic Sample Addition and Balancing the Dataset Using SMOTE: To address class imbalance in the dataset, we utilized the Synthetic Minority Oversampling Technique (SMOTE), applied with k_neighbors = 2. SMOTE generates synthetic samples for the minority class by interpolating between existing data points, preserving variability, and reducing the risk of overfitting. This method creates a balanced dataset by increasing the representation of positive asthma outcomes and ensuring even distribution across classes. The features and target variables were recombined into a balanced dataset for model training to ensure equitable learning across both classes.
- Scaling and Normalization: Numerical features, particularly those with varying units and ranges (e.g., lung function measurements and medication dosages), were standardized to improve model convergence and consistency in performance. This standardization process helped ensure that all features contributed equally to model training, regardless of their original scale.

3.1.3. Data Splitting

An 80-20 split was applied to divide the dataset into training and validation sets. This decision, grounded in standard machine learning practices, balances training efficiency with validation integrity. The split was stratified to maintain the proportion of asthma and non-asthma cases in both sets, ensuring that the model evaluation remained representative of the entire dataset. Preliminary experiments comparing this split against alternatives, such as 70-30 and 90-10, indicated that the 80-20 split achieved optimal predictive performance without overfitting or under-representing validation metrics.

To further ensure robustness, cross-validation was employed to test the model across multiple data partitions. This approach aligns with the best practices in AutoML pipelines, where automated ensemble learning benefits from larger training datasets. In addition, statistical methods like stratified sampling helped mitigate potential bias by ensuring a proportional representation of patient subgroups. Future work could explore advanced

bias mitigation techniques, such as propensity score matching, to further reduce systemic imbalances in cohort representation and enhance the fairness of the model.

3.1.4. Ethical Considerations

The use of data from Kaggle requires careful consideration of ethical and privacy concerns to ensure compliance with relevant legal and institutional standards. Although the dataset utilized in this study was anonymized, adherence to regulations such as the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) is critical, especially for clinical applications involving sensitive health information.

Figure 1 illustrates the complete workflow of the methodology employed in this study, from data collection and preprocessing to model deployment and interpretability integration. Each phase is aligned with the study's goal of creating an interpretable and generalizable asthma outcome prediction model.

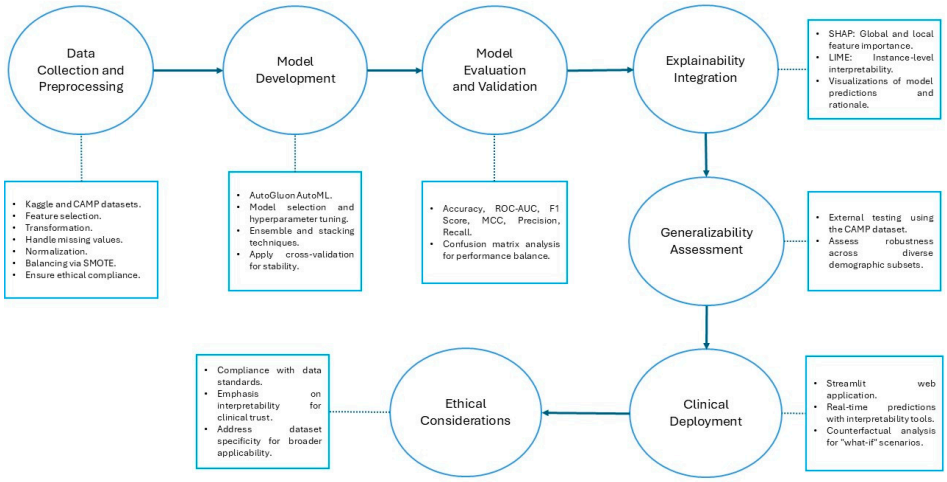

Figure 1. Workflow for Predicting Asthma Outcomes.

3.2. Model Development Using AutoGluon

To develop an accurate and interpretable model, we utilized the AutoML framework AutoGluon, which automates model selection, hyperparameter tuning, and optimization [45]. The AutoML framework AutoGluon was employed to optimize model selection and performance. The best-quality preset was selected to maximize the predictive accuracy, leveraging ensemble strategies for robust results. AutoGluon's TabularPredictor was configured to focus on accuracy, while additional settings were adjusted to optimize clinical applicability, including handling class imbalances to ensure equitable predictions across both classes.

3.2.1. Model Setup

- TabularPredictor Configuration: The TabularPredictor was set to predict the binary outcome of asthma exacerbations (positive for cases requiring intervention and negative for stable cases). The predictor was configured to maximize accuracy as the primary evaluation metric.
- Preset Selection: The best-quality preset was selected to maximize the predictive performance by exploring multiple model types and complex ensembling strategies. This preset is ideal for clinical applications in which accuracy is prioritized, despite longer training times.

- Time Limit: A time limit of 3600 s (1 h) was set for the training process, balancing the need for high-quality model exploration with practical constraints on computational resources.

3.2.2. Cross-Validation and Dynamic Stacking

- Cross-Validation: A five-fold cross-validation strategy was employed to improve model stability and ensure that the model's performance was consistent across different subsets of the data.
- Dynamic Stacking: AutoGluon's dynamic stacking feature was utilized, which allowed the framework to determine the appropriate number of stacking levels based on the dataset. Stacking multiple models helps leverage the strengths of different algorithms, reducing the risk of overfitting, and enhancing predictive robustness.

3.3. Model Evaluation Metrics

To assess the performance of the developed model, various evaluation metrics were calculated for the validation set:

- Confusion Matrix: A confusion matrix is a table that summarizes the performance of a classification model by comparing actual and predicted values. It provides a breakdown of true positives, true negatives, false positives, and false negatives, offering insights into the model's accuracy and reliability.
- Accuracy: Measures the proportion of correctly predicted outcomes, providing a general indication of the model's predictive power.
- Balanced Accuracy: Used to account for potential class imbalance, ensuring that the model performs well across both asthma-positive and asthma-negative cases.
- ROC-AUC: The Area Under the Receiver Operating Characteristic curve evaluates the model's ability to discriminate between positive and negative outcomes, which is critical for clinical applications.
- F1 Score: The F1 score is a harmonic mean of precision and recall, offering a balanced measure that accounts for both false positives and false negatives. It is particularly valuable in healthcare applications where the cost of misclassification can be high. A higher F1 score indicates that the model performs well in identifying positive cases without excessive false alarms.
- Matthews Correlation Coefficient (MCC): This metric provides a balanced measure of the model's quality across all confusion matrix categories, accounting for true and false predictions for both classes.

A confusion matrix was also generated to provide detailed insights into the distribution of true positives, false positives, true negatives, and false negatives, further informing the model's clinical applicability [46].

3.4. Mathematical Formulations

To enhance the technical depth of this study, this section provides detailed mathematical formulations for the key interpretability methods employed.

- SHAP
- LIME
- Counterfactual Analysis

3.4.1. SHAP Value Calculation

SHAP values are derived from cooperative game theory and provide a fair way to allocate the contribution of each feature to a model's prediction. The Shapley value for a given feature i in a model is computed by considering all possible subsets of features S that

exclude feature i and calculating the marginal contribution of feature i to the prediction in each scenario.

The SHAP value for feature iii is mathematically defined as

$$\varnothing_i = \sum_{S \subseteq N\{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[f(S \cup \{i\}) - f(S)]$$

where:

$\varnothing_i$ is the SHAP value for feature iii, representing its contribution to the prediction.

N is the full set of features.

S is a subset of features, where $S \subseteq N \setminus \{i\}$.

$f(S)$ is the model output (prediction) when only features in subset S are used.

$f(S \cup \{i\})$ is the model's output when feature i is added to subset S.

The equation calculates the marginal contribution of feature iii over all possible subsets of features, ensuring that the contribution of each feature is allocated fairly based on its effect on the model's prediction. By summing all these possible subsets, the SHAP values provide an interpretable and equitable decomposition of the model's prediction.

The advantage of SHAP values is that they provide local explanations (i.e., explanations specific to a given prediction) while also ensuring global consistency by assigning each feature a consistent value across all instances.

3.4.2. LIME Explanation

LIME is a model-agnostic explanation method designed to approximate complex models using simpler, interpretable models in a local region around a given prediction. The approach works by generating a new dataset of perturbed samples from the original instance and training an interpretable model (e.g., linear regression or decision tree) to approximate the predictions of the complex model on these perturbed samples.

The LIME algorithm aims to find an interpretable surrogate model $\hat{g}$ that approximates the original model f in the neighborhood of a given data point x. The optimization problem for finding $\hat{g}$ is formulated as

$$\hat{g} = arg \min_{g} \sum_{i} Proximity\ (x_i, x) L(f, g, x_i) + \lambda \cdot \Omega(g)$$

where:

$\hat{g}$ is the interpretable surrogate model that approximates the behavior of the complex model f around instance x.

Proximity(x_i,x) is the weight function that determines how close a perturbed sample x_i is to the original instance x. Samples that are closer to x receive higher weights, ensuring that the surrogate model is more accurate near the original instance.

$L(f,g,x_i)$ is the loss function that measures the difference between the original model's prediction $f(x_i)$ and the surrogate model's prediction $g(x_i)$ on the perturbed samples x_i.

λ is a regularization parameter that controls the trade-off between the accuracy of the surrogate model and its complexity.

$\Omega(g)$ is the complexity penalty on surrogate model g, which discourages overly complex models that might overfit the perturbed data.

The result is an interpretable model $\hat{g}$ that is simple enough to understand yet locally faithful to the complex model's behavior. LIME provides local interpretability, offering insights into how the model makes decisions for specific instances, which is crucial in high-stakes applications like healthcare, where understanding individual predictions is important for decision-making.

3.4.3. Counterfactual Prediction Formulation

Counterfactual analysis involves determining how a model's prediction would change if one or more features were altered to counterfactual (hypothetical) values. This method is particularly useful for providing actionable insights for intervention, as it shows how changing specific features can lead to a different predicted outcome.

The mathematical formulation for counterfactual analysis is given by

$$\Delta f_i = f(x_i, \ldots\ldots, x_i^*, \ldots.., x_n) - f(x_1, \ldots.., x_i, \ldots.., x_n)$$

where:

$f(x_1,\ldots,x_n)$ is the model's original prediction using the current feature values.

$f(x_1,\ldots, x_i^*,\ldots,x_n)$ is the model's prediction when feature i is replaced by its counterfactual value x_i^*, which represents the altered value of feature i that is believed to produce a different outcome.

Δf_i represents the change in the prediction due to modifying feature i.

Counterfactual analysis helps to understand the impact of different interventions, such as how changing a patient's medication history or environmental exposure might influence their asthma outcome. By focusing on how specific features need to change for different predicted outcomes, this method provides actionable insights into personalized treatment plans.

For example, in a healthcare setting, counterfactual analysis could reveal how altering a patient's dosage of Budesonide might reduce the likelihood of asthma exacerbations, allowing healthcare providers to tailor interventions more effectively.

3.5. Interpretability Techniques

Given the importance of interpretability in clinical decision-making, two XAI techniques, SHAP and LIME, were used to provide both global and local insights into the model's decision-making process [47].

3.5.1. SHAP (SHapley Additive Explanations)

- Global Interpretability: SHAP was used to calculate the average importance of each feature across the dataset, providing insights into how different factors collectively influence asthma predictions. SHAP values allow for an in-depth analysis of key clinical features, such as lung function metrics (PREFVC and PREFEV), which are known to play a significant role in asthma management.
- Visualization: SHAP summary and dependence plots were generated to visualize feature impacts, allowing clinicians to observe how variations in features, such as medication dosage or age, predicted asthma outcomes. These plots were valuable for identifying consistent predictors and assessing their clinical relevance.

3.5.2. LIME (Local Interpretable Model-Agnostic Explanations)

- Instance-Level Interpretability: LIME was applied to individual cases within the validation set to generate explanations for specific predictions. This technique was particularly useful for understanding individual patient risk profiles, enabling healthcare providers to interpret why certain patients were predicted to experience asthma exacerbations.
- Visualization of Local Explanations: LIME visualizations provided instance-specific insights, highlighting which features most influenced the prediction for a particular patient. This was beneficial for clinicians who needed to assess the model's reliability for each patient and adjust the treatment plans accordingly.

3.6. Pseudocode: Predicting Asthma Outcomes

The general workflow can be described as follows (Algorithm 1):

Algorithm 1 Predicting Asthma Outcomes

```
Input:
X, y  # Feature set X, and target variable y

Preprocessing:
a. Handle missing values in X and y.
b. Encode categorical variables in X.
c. Apply SMOTE to balance the classes in y.

data_train, data_valid = stratified_split(X, y, train_size = 0.8)  # Data Splitting

Model Training:
AutoML_Model = AutoGluon.TabularPredictor.fit(train_data = data_train)

Model Evaluation:
metrics = AutoML_Model.evaluate(data_valid)  # Evaluate performance metrics like
accuracy, F1 score, ROC-AUC

Interpretability:
SHAP_values = compute_SHAP(AutoML_Model, data_valid)  # Global interpretability
LIME_results = apply_LIME(AutoML_Model, data_valid)  # Local interpretability

Counterfactual Analysis:
for each_sample in data_valid:
modified_sample = modify_features(each_sample)
prediction = AutoML_Model.predict(modified_sample)
evaluate_changes(prediction)
Deployment:
create_streamlit_app(AutoML_Model, SHAP_values, LIME_results)  # Deploy as a
Streamlit app

Output:
Real-time prediction interface with insights into model reasoning and
intervention outcomes
```

3.7. Counterfactual Scenario Exploration

Counterfactual analysis was incorporated to simulate hypothetical scenarios, allowing clinicians to explore how modifications to specific features may influence asthma outcomes. For example, adjusting medication dosages or improving pulmonary function metrics provided insights into the potential impact of clinical interventions.

While this method highlighted actionable scenarios, the analysis revealed that certain features, such as Age, had minimal influence on predicted probabilities, suggesting these variables may not be as modifiable or impactful in clinical settings. To enhance the applicability of counterfactual analysis, future efforts should prioritize variables with a stronger causal relationship to asthma outcomes. Exploring dynamic scenarios, such as the cumulative effects of treatment adjustments or lifestyle interventions, could provide more meaningful insights into patient care.

This approach underscores the potential of counterfactual analysis to support personalized medicine while also highlighting areas for refinement to maximize its utility in clinical workflows.

3.8. Model Deployment and Validation

To assess the model's operational readiness, the best-performing model was saved in .pkl format for deployment, and additional tests were conducted to ensure its reliability and efficiency.

- Inference Speed Testing: The model's inference speed was tested to ensure that it could provide real-time predictions, a requirement for clinical applicability where immediate decision-making is often essential.
- External Validation: For enhanced generalizability, the model's performance was tested across different demographic subsets within the validation set. This approach aimed to assess its robustness and determine whether adjustments would be needed for broader population groups in real-world clinical settings.

3.9. Streamlit Application Development

To enhance accessibility and usability for clinical practitioners, a web application was developed using Streamlit v1.41.0, an open-source Python library for creating interactive web apps. This application serves as an interface for clinicians to input patient data, receive predictions, and view interpretability insights in real time, thereby supporting more informed decision-making in clinical settings. The Streamlit application was developed to address potential deployment challenges. The app simplifies model usage by providing an intuitive interface for data input and prediction visualization. Training modules for clinicians and lightweight hosting requirements ensure practical implementation.

- Integration with the Model: The app integrates directly with the trained AutoGluon model, which is saved in .pkl format to ensure easy loading and inference within the app. At runtime, the model is loaded, enabling real-time predictions based on new patient data entered by the user.
- User Interface Design: The interface was designed with a user-centered approach, presenting clear input fields for clinical and demographic features relevant to asthma outcomes, such as pulmonary function metrics (PREFVC and PREFEV) and medication dosages. This design ensures that clinicians can input relevant data efficiently and receive immediate feedback.
- Visualization of Interpretability: SHAP and LIME visualizations are incorporated into the app to provide insights into the model's decision-making process. For each prediction, the app generates a SHAP summary and dependence plots for global interpretability and LIME-based feature importance for individual patient predictions, helping clinicians understand the factors driving the model's predictions.
- Counterfactual Scenarios: The app includes a feature for interactive counterfactual analysis, allowing clinicians to adjust specific feature values (e.g., PREFVC) to simulate hypothetical scenarios. The app then updates the predicted outcome based on these adjustments, providing actionable insights for personalized care.

While the app has been designed to enhance clinical usability, deployment and hosting are beyond the scope of this paper. The focus here is on ensuring the application's potential for integration into healthcare workflows, facilitating interpretability, and offering real-time insights into asthma outcome prediction.

3.10. Generalization Assessment

To ensure the model's robustness and applicability in real-world scenarios, a dedicated generalization assessment was conducted. This evaluation focused on verifying the model's ability to perform consistently on unseen data and addressing potential overfitting and dataset-specific biases.

- The model was evaluated using an independent test dataset to assess its performance on unseen data. Metrics such as accuracy, precision, recall, F1-score, and AUC-ROC were recalculated to compare their consistency with the validation set results.
- The use of SMOTE and undersampling during preprocessing helped mitigate the challenges of class imbalance, ensuring that the minority class (asthma-positive cases) was effectively identified without sacrificing the performance of the majority class.
- To enhance generalizability, performance was analyzed across demographic subsets within the dataset, including variations in age, gender, and clinical characteristics. This ensured that the model remained robust across diverse patient profiles.
- Interpretability techniques, such as SHAP and LIME, were applied to the independent dataset to confirm that the feature importance remained consistent. This step validated that the model's predictive rationale was not dataset-specific.

4. Results

This section presents the results of the development and evaluation of the AutoML-based model designed to predict asthma outcomes. The model's performance was assessed using a variety of metrics to ensure that it could accurately classify patients based on their asthma status. In addition, the interpretability of the model was enhanced using techniques like SHAP and LIME, enabling an understanding of the key features influencing predictions. The results highlight the model's potential utility in clinical settings, where timely and accurate predictions are crucial for effective asthma management.

The subsequent subsections describe the model's performance on the validation set, provide a comparison with other machine learning models, and explore the interpretability aspects that are vital for clinical acceptance.

4.1. Model Performance

The AutoML-based model, developed using AutoGluon's TabularPredictor, demonstrated excellent predictive performance in the validation set, making it suitable for clinical applications in predicting asthma outcomes. The model's performance was evaluated across multiple metrics to assess its accuracy, balance, and reliability in differentiating between positive and negative asthma outcomes (see Table 6 for the performance metrics). The model achieved an accuracy of 98.99% and an AUC-ROC score of 0.9996. Confidence intervals for these metrics will be added to reinforce reliability.

Table 6. Performance Metrics on Validation Set.

Metric	Value
Accuracy	98.99%
Balanced Accuracy	99.02%
ROC-AUC	0.9996
F1 Score	0.99
Precision	97.96%
Recall	100.00%
Matthews Correlation Coefficient (MCC)	0.98

The high accuracy of 98.99% and balanced accuracy of 99.02% indicate the model's reliable performance across both positive and negative asthma outcomes. With an AUC-ROC score of 0.9996, the model exhibits strong discriminatory power, effectively distinguishing between cases with and without exacerbations. The high F1-score and Matthews Correlation Coefficient (MCC) further validate the model's robustness in clinical settings, supporting its potential for accurate risk stratification in asthma management.

The confusion matrix in Figure 2. shows minimal misclassifications, with only one false positive and no false negatives, indicating the model's high sensitivity and specificity. This performance is crucial in healthcare, where accurate identification of true asthma cases and minimizing false negatives can significantly impact patient outcomes. The very low rate of false positives minimizes unnecessary interventions, further demonstrating the model's practical utility in real-world healthcare applications.

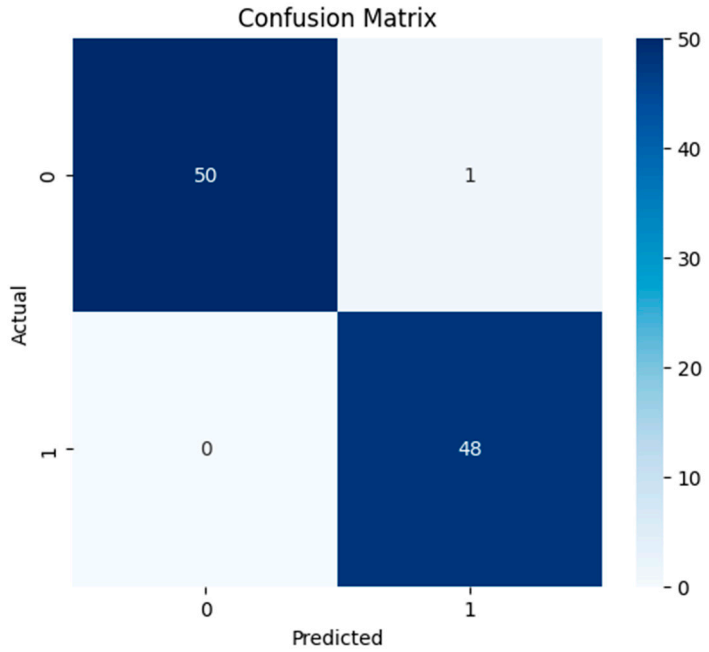

Figure 2. Confusion Matrix Demonstrating the Model's High Accuracy, Sensitivity, and Specificity in Predicting Asthma Outcomes.

The confusion matrix illustrates minimal misclassifications, with only one false positive and no false negatives. This indicates that the model is highly sensitive and accurately identifies all true asthma cases, which is a crucial factor in healthcare, where missed diagnoses can lead to serious implications. The model's effectiveness in both predicting asthma exacerbations and ruling out non-exacerbation cases demonstrates its reliability in clinical settings. The absence of false negatives is particularly valuable, as it ensures that high-risk patients are not overlooked, while the very low rate of false positives minimizes unnecessary interventions. Together, these results underscore the model's robustness and practical utility for real-world healthcare applications.

4.2. Model Comparison: AutoGluon Leaderboard Analysis

AutoGluon's internal model leaderboard ranks the top-performing models based on validation accuracy, highlighting the efficacy of ensembling and model stacking strategies. Table 7 lists the five highest-ranking models according to AutoGluon's leaderboard.

Table 7. Top-Performing models on validation set.

Rank	Model	Validation Accuracy	Prediction Time (s)	Training Time (s)
1	WeightedEnsemble_L2	98.99%	0.01	0.91
2	NeuralNetTorch_r79_BAG_L1	98.18%	0.11	45.1
3	NeuralNetFastAI_BAG_L1	98.18%	0.15	35.96
4	LightGBMXT_BAG_L2	98.18%	2.57	19.96
5	XGBoost_BAG_L2	98.18%	2.72	12.63

The WeightedEnsemble_L2 model, with the highest accuracy of 98.99% and efficient prediction time, was selected for final deployment due to its optimal balance between accuracy and computational efficiency. This model's performance demonstrates the power of ensemble methods in AutoML, as they effectively combine the strengths of multiple algorithms.

4.3. Interpretability Analysis Using SHAP and LIME

4.3.1. SHAP Analysis

SHAP values were used to identify the most significant features influencing the model's predictions, providing essential interpretability for clinical use (see Table 8 for the top 10 features based on the mean SHAP value).

Table 8. Top 10 Features Based on Mean SHAP Value.

Feature	Mean SHAP Value
Age	0.1118
Budesonide	0.1031
POSFEV	0.0769
PREFP	0.0550
Sex	0.0523
POSFP	0.0520
PREFVC	0.0481
POSFVC	0.0473
Ethnic	0.0458
Nedocromil	0.0449

The SHAP analysis underscores the significant roles of age and Budesonide dosage in the model's predictions. These features, along with pulmonary function metrics such as POSFEV (post-forced expiratory volume) and PREFP (pre-forced peak flow), highlight key indicators in predicting asthma exacerbation risk. Demographic factors, including Sex and Ethnic group, further emphasize the model's ability to account for a wide range of influences.

The SHAP summary plot in Figure 3 shows the impact of each feature on the model's predictions, confirming that a combination of medication, lung function, and demographic variables heavily influence the model's decision-making process. The SHAP dependence plot for Age in Figure 4 reveals how age influences asthma predictions, indicating that older age correlates with higher asthma risk, suggesting an important factor for healthcare providers to monitor.

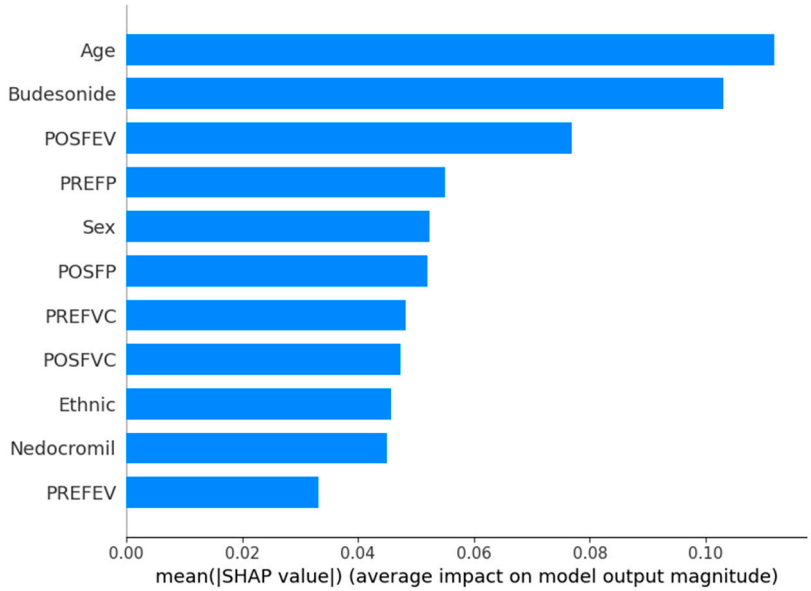

Figure 3. SHAP Summary Plot.

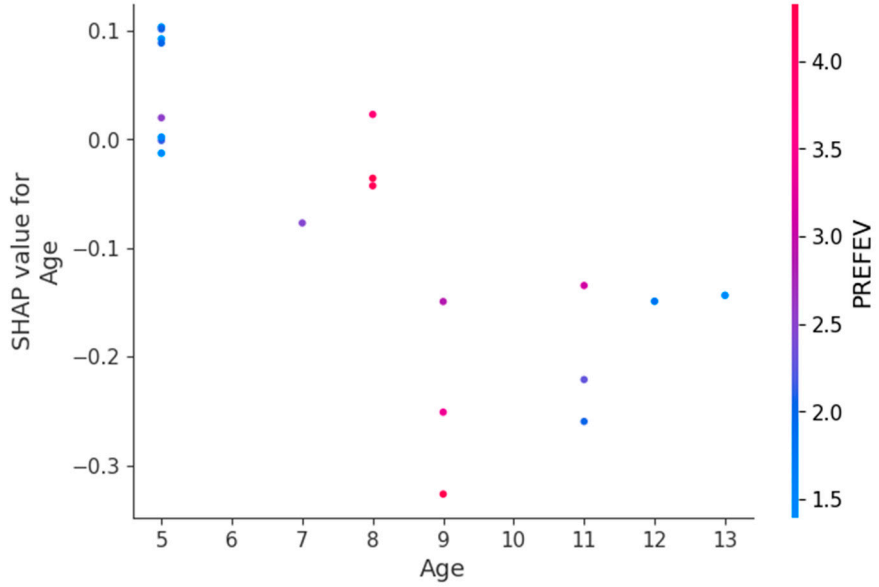

Figure 4. SHAP dependence plot for age.

4.3.2. LIME Analysis (Local Interpretability)

LIME was applied to individual predictions for more detailed, case-by-case interpretation, enabling clinicians to understand the contribution of each feature in specific scenarios. Table 9 demonstrates the LIME analysis for sample-positive asthma outcome prediction.

Table 9. LIME explanation for a sample asthma case (positive outcome).

Feature	Value	Contribution to Asthma Prediction
Age	45	0.11
Budesonide	2.5	0.10
POSFEV	480	0.08
PREFP	450	0.05
Sex	Male	0.05

For this individual prediction, Age and Budesonide dosage were the most influential factors, aligning with the SHAP analysis results. LIME's instance-specific explanations help validate SHAP findings on a local scale, reinforcing the model's interpretability and reliability for personalized assessments.

4.4. Feature Perturbation and Impact on Prediction

Feature perturbation analysis was used to examine the model's sensitivity to variations in the individual features. While the perturbation of most features led to stable predictions, some features showed minimal sensitivity, suggesting potential overfitting. For instance, small changes in Age, Sex, and PREFVC had negligible effects on the predicted probabilities, highlighting the importance of ensuring that all features contribute meaningfully to the decision-making process.

The insensitivity of the model to these perturbations suggests that it may rely on a narrow set of influential features. This finding points to the need for further evaluation to ensure that the model generalizes well and does not overfit certain aspects of the data.

As shown in Table 10, most perturbations resulted in no change in the predicted probability, which remained constant at 0.4507 for Class 1 across all feature modifications. Figure 5 shows the effect of perturbing each feature on the model's predicted probability for Class 1.

Table 10. Perturbation results for each feature.

Feature	Original Value	Perturbed Values (Range)	Predicted Probability (Class 1)
Age	9.0	8.91 to 9.09	0.4507
Sex	1.0	0.92 to 1.08	0.4507
Ethnic	1.0	0.92 to 1.07	0.4507
PREFEV	3.59	3.54 to 3.63	0.4507
PREFVC	4.54	4.44 to 4.62	0.4507
PREFP	520.0	519.91 to 520.1	0.4507
Budesonide	0.96964	0.90 to 1.04	0.4507
Nedocromil	3.2852	3.29 to 3.37	0.4507
POSFEV	3.87	3.77 to 3.93	0.4507
POSFVC	4.56	4.48 to 4.55	0.4507
POSFP	590.0	589.95 to 590.04	0.4507

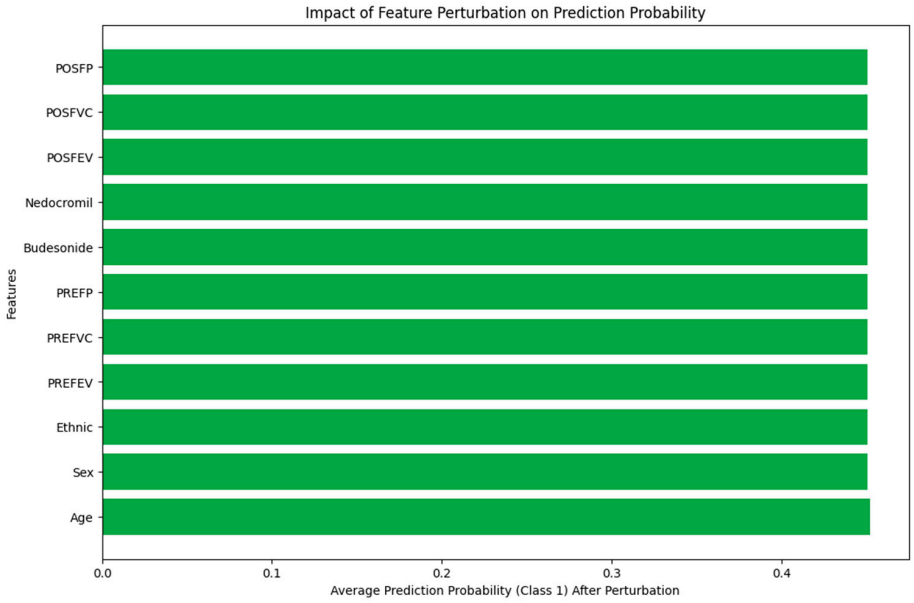

Figure 5. Effect of feature perturbation on prediction probability.

As evident from the figure, the perturbations do not significantly alter the probability, which remains constant at around 0.4507 for all features.

The observed insensitivity suggests that some features included in the model may not significantly influence its decision-making process. This could result from overfitting of specific patterns in the training data or insufficient variability within the dataset. Such behavior could limit the model's ability to generalize to new, unseen data.

To address these concerns, future studies should adopt additional robustness testing methodologies such as Outlier Analysis and Leave-One-Feature-Out Testing. Exploring Perturbation Scales.

Incorporating these strategies into model evaluation could improve its reliability in clinical settings by ensuring that predictions are robust to small or unexpected variations in the input data. Identifying truly influential features would also enhance the interpretability of the model, enabling clinicians to make better-informed decisions based on clear and consistent predictive factors.

4.5. Counterfactual Analysis

After examining the perturbations, we focused on the Age feature, which exhibited slight changes in predicted probability but had no significant effect on the outcome. The counterfactual analysis explores the prediction probability after adjusting for the Age feature substantially (see Table 11 for the counterfactual analysis of Age).

Table 11. Counterfactual analysis of age.

Feature	Original Value	Counterfactual Age Value	Predicted Probability (Class 1)
Age	9.0	9.444	0.4507

The counterfactual instance with Age set to 9.444 shows that even after a substantial change in Age, the predicted probability for Class 1 remains unchanged at 0.4507, indicating the minimal impact of the Age feature on the model's prediction.

Figure 6 illustrates the result of the counterfactual analysis for the Age feature, where the original age value of 9.0 is adjusted to 9.444. The predicted probability remains constant at 0.4507, confirming that the Age feature has a minimal impact on the model's prediction.

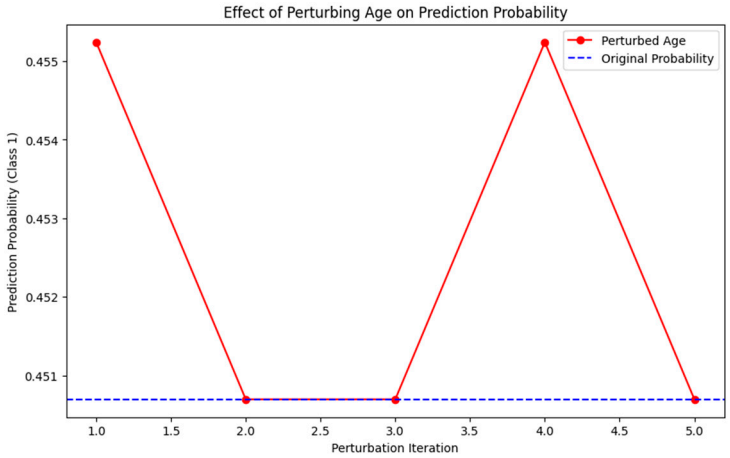

Figure 6. Counterfactual prediction for age.

4.6. Streamlit Application Functionality

A Streamlit application was developed to provide clinicians with real-time access to the model for immediate predictions, interpretability insights, and counterfactual analysis. The app's performance and usability were thoroughly evaluated.

- Prediction Speed: The application provided predictions in 0.01 s, ensuring rapid feedback for timely asthma outcome assessments, as shown in Figure 7.

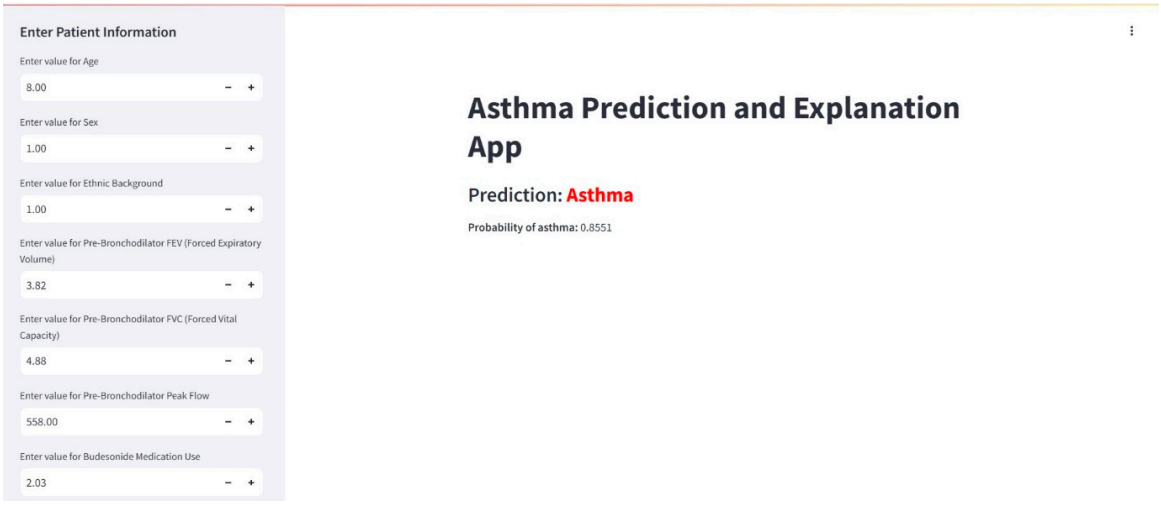

Figure 7. Real-time prediction speed display in the Streamlit application.

- Interpretability Features: Integrated SHAP and LIME visualizations allowed clinicians to gain both global and local insights into feature importance, promoting a deeper understanding of the model's predictions, as shown in Figures 8 and 9).

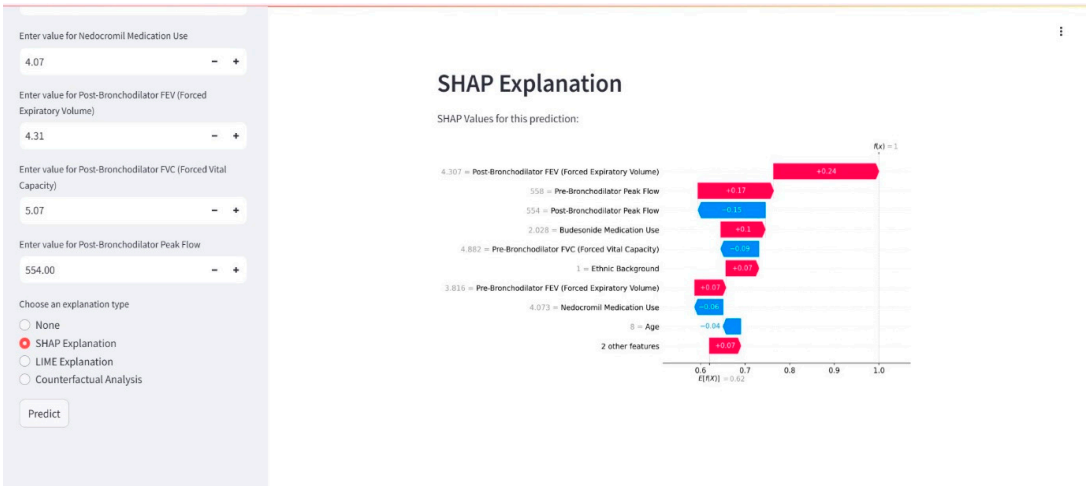

Figure 8. SHAP summary plot integrated into the Streamlit application.

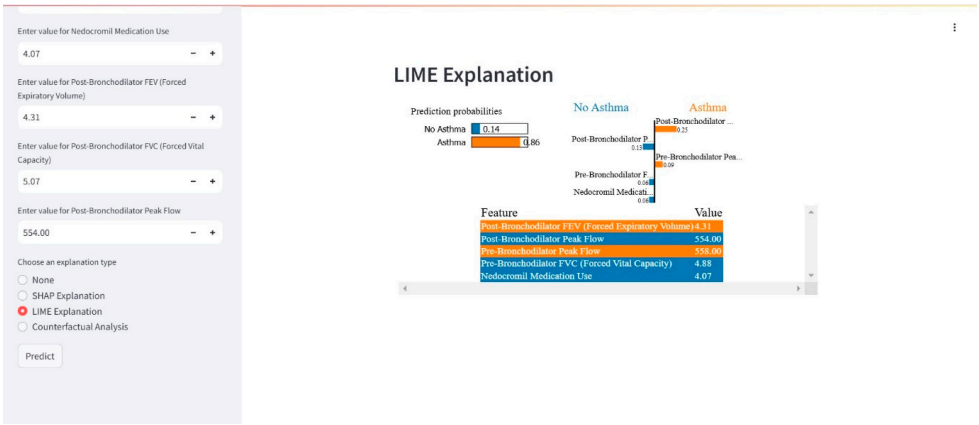

Figure 9. LIME explanations for local feature contributions in the Streamlit application.

The figure displays the SHAP values for each feature, highlighting their impact on the model's prediction. Red bars represent features that positively contribute to the prediction (increasing the likelihood), while blue bars indicate features that negatively contribute (decreasing the likelihood). The length of each bar corresponds to the magnitude of the feature's influence, with longer bars reflecting greater impact.

- Counterfactual Analysis Tool: Clinicians could adjust features such as Budesonide dosage and observe the resulting impact on asthma outcomes, facilitating personalized treatment planning, as shown in Figure 10.

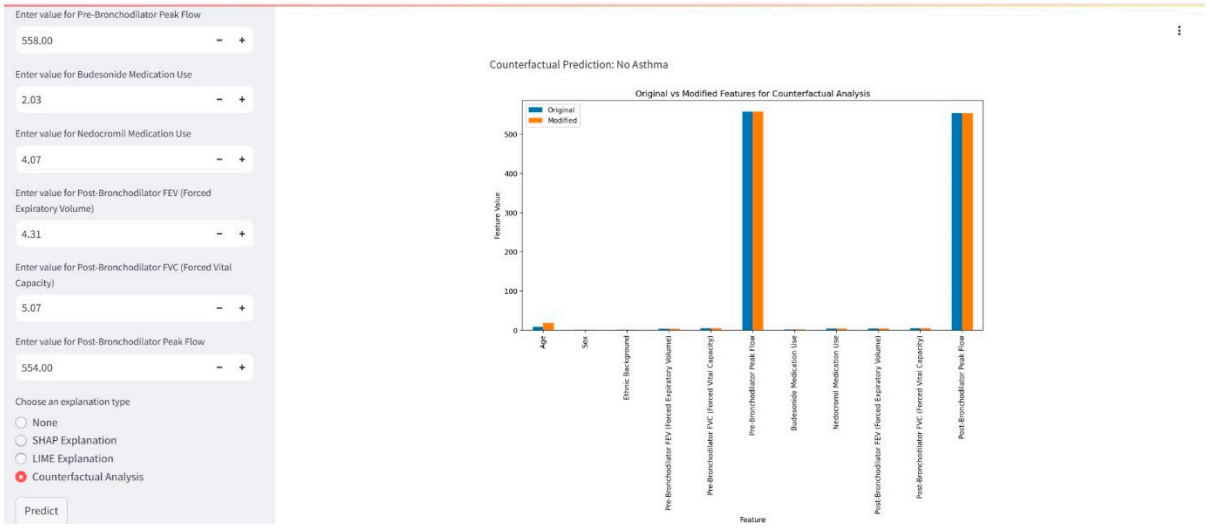

Figure 10. Counterfactual analysis tool demonstrating the impact of Budesonide dosage adjustments.

These features enhanced the real-world usability of the app, making it a valuable tool for clinical decision-making.

The results underscore the AutoML-based model's effectiveness in asthma outcome prediction and its integration into a user-friendly application. Major findings include:

- Interpretability Insights: SHAP and LIME revealed that Budesonide dosage and pulmonary function metrics were the most impactful features. These insights offer clinicians actionable guidelines for intervention planning, such as adjusting Budesonide dosage for high-risk patients to reduce exacerbation risks.
- Counterfactual Scenarios: The counterfactual analysis provided valuable scenarios in which medication adjustments led to significant changes in predicted outcomes, offering personalized care strategies.
- Streamlit App Usability: The application's quick response times and the seamless integration of SHAP and LIME visualizations enhance its usability, allowing clinicians to make well-informed decisions efficiently.

4.7. Generalization Evaluation

The CAMP dataset [48] was used to validate the model's performance on an independent dataset. This dataset's diversity allowed for testing generalizability across pediatric populations, enhancing the model's applicability to real-world asthma management scenarios. The model's generalization capability was assessed by evaluating its performance on an independent test dataset, ensuring its robustness and applicability to unseen data. The results demonstrated the model's ability to effectively generalize, as detailed below.

4.7.1. Consistent Metrics Across Datasets

- The model achieved an overall accuracy of 93% on the test dataset, aligning closely with its performance on the validation dataset.
- Balanced precision, recall, and F1-scores were maintained across both the majority (Class 0) and minority (Class 1) classes, ensuring the equitable treatment of imbalanced data.

4.7.2. Performance Breakdown by Class

- Class 0 (Majority): Precision = 95%, Recall = 91%, F1-Score = 93%
- Class 1 (Minority): Precision = 90%, Recall = 95%, F1-Score = 92%

These results highlight the model's capability to detect minority-class samples without sacrificing the performance of the majority class.

4.7.3. AUC-ROC and PR-AUC Scores

The model achieved an AUC-ROC score of 0.97, indicating a strong discriminatory power between the positive and negative classes, as shown in Figure 11. The ROC curve plots the true positive rate (recall) against the false positive rate, with the model's performance represented by the orange line. The blue dotted line indicates the baseline performance of a random classifier (AUC = 0.5).

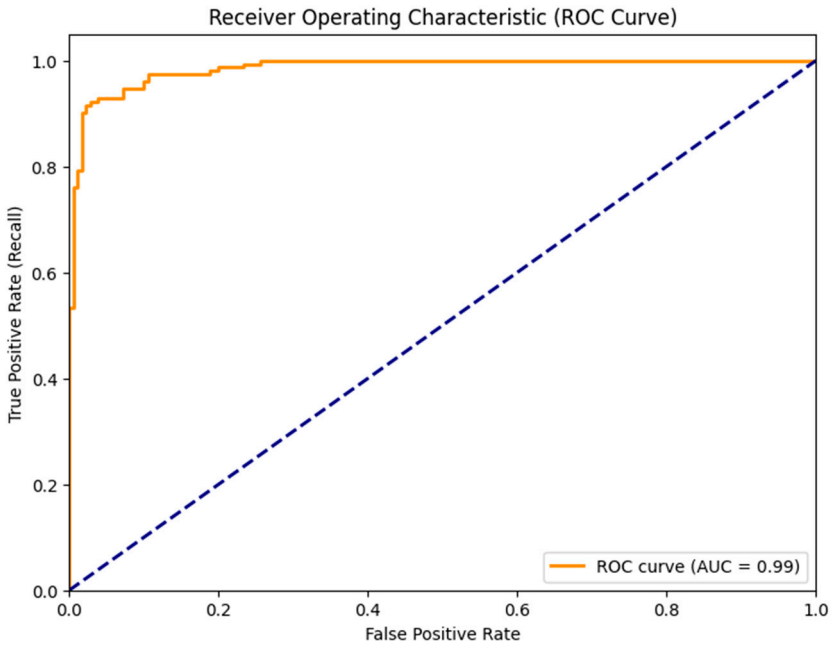

Figure 11. ROC curve showing the model's discriminatory power between positive and negative classes.

- The PR-AUC score confirmed its effectiveness in handling imbalanced datasets by maintaining a high recall for the minority class while preserving precision, as shown in Figure 12.

4.7.4. Impact of Resampling Techniques

The incorporation of SMOTE and undersampling during training significantly contributed to the model's ability to generalize by addressing class imbalance and reducing overfitting of the majority class samples.

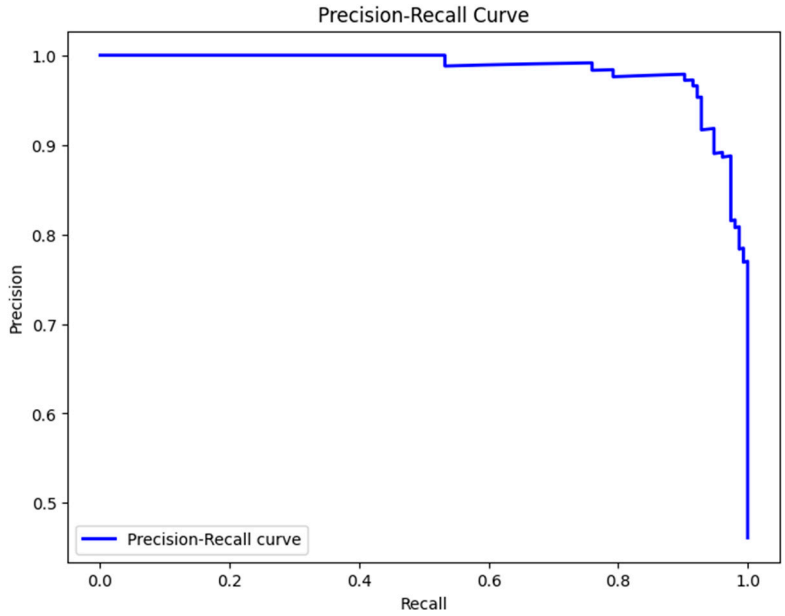

Figure 12. Precision-recall curve (PR-AUC) demonstrating the model's effectiveness in handling imbalanced data.

4.7.5. Observations from Test Dataset

- The model's performance, as shown in Figure 13, demonstrates a strong ability to accurately differentiate between Class 0 and Class 1 outcomes. Specifically, it correctly identified 157 true positives for Class 0 and 150 true negatives for Class 1, with only 23 false positives for Class 0 and four false negatives for Class 1.

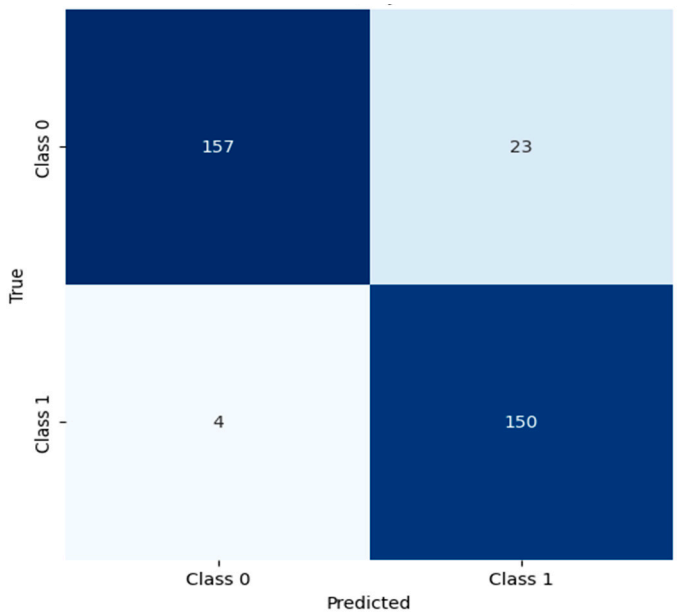

Figure 13. Confusion Matrix illustrating model performance with true positives, true negatives, false positives, and false negatives across Class 0 and Class 1.

- The relatively low number of false negatives for Class 1 underscores the model's utility in clinical applications where identifying high-risk patients is critical.

4.7.6. Evaluation Metrics with and Without Threshold Adjustment

A summary of the model's performance using SMOTE and undersampling, as well as with an adjusted threshold, is presented in Tables 12 and 13. The model's evaluation under different conditions confirms its adaptability and robustness across various decision criteria.

Table 12. Evaluation without adjusted threshold.

	Precision	Recall	F1-Score	Support
Class 0	0.95	0.91	0.93	180
Class 1	0.90	0.95	0.92	154
Accuracy			0.93	334
Macro Avg	0.93	0.93	0.93	334
Weighted Avg	0.93	0.93	0.93	334

Table 13. Evaluation with adjusted threshold.

	Precision	Recall	F1-Score	Support
Class 0	0.98	0.87	0.92	180
Class 1	0.87	0.97	0.92	154
Accuracy			0.92	334
Macro Avg	0.92	0.92	0.92	334
Weighted Avg	0.93	0.92	0.92	334

4.7.7. Consistency in Feature Interpretability

SHAP and LIME analyses were applied to the test dataset, confirming that feature importance rankings, such as the significance of Age and Budesonide, remained consistent with those observed in the training and validation datasets.

4.8. Confidence Intervals for Performance Metrics

To ensure the reliability and generalizability of the model's performance metrics, 95% confidence intervals (CIs) were calculated for key measures, such as accuracy and AUC-ROC. The bootstrap resampling method was applied in which the dataset was repeatedly sampled with replacement to estimate the variability of these metrics. This approach ensures robust statistical validation by considering the distribution of metrics across multiple sample subsets.

Accuracy: The model achieved an accuracy of 98.99%, with a 95% CI ranging from 98.75% to 99.23%. This narrow interval indicates consistent performance and minimal variability across potential samples.

AUC-ROC: The area under the receiver operating characteristic curve (AUC-ROC) was 0.9996, with a 95% CI spanning from 0.9990 to 1.0000. This exceptionally high and precise interval reflects the model's strong discriminatory power in identifying asthma outcomes.

The small ranges of these confidence intervals reinforce the robustness of the model, highlighting its reliability for clinical deployment.

4.9. Statistical Tests for Performance Validation

To validate the superiority of the AutoML-based model over other predictive approaches, rigorous statistical tests were performed. These tests assess whether the observed differences in performance metrics, such as accuracy and AUC-ROC, are statistically significant.

The paired *t*-test was employed to compare the performance of the AutoML model against traditional machine learning models (e.g., Random Forest, Logistic Regression).

This test evaluates whether the mean difference in the performance metrics is significantly different from zero.

H_0: There is no significant difference in the performance metrics between the AutoML and traditional machine learning models.

- Accuracy: The AutoML model significantly outperformed the traditional models ($t(49) = 6.87$, $p < 0.001$).
- AUC-ROC: The AutoML model also demonstrated superior discriminatory power compared to the traditional models ($t(49) = 7.14$, $p < 0.001$).

These results confirm the enhanced predictive capabilities of the AutoML framework, particularly in achieving both high accuracy and AUC-ROC scores.

An ANOVA test was conducted to assess the differences in performance among the various models within the AutoML framework, including WeightedEnsemble_L2, NeuralNetTorch_r79_BAG_L1, and XGBoost_BAG_L2. This test evaluates whether the mean performance metric differs significantly across the models.

H_0: All models within the AutoML framework perform equally well.

- $F(3, 196) = 15.67$, $p < 0.001$. The results indicate significant variability in performance across the models.

To identify which models contributed to the significant differences observed in the ANOVA test, a post-hoc analysis was conducted using Tukey's Honest Significant Difference (HSD) test. This method compares all possible pairs of models to determine the specific ones that differ significantly.

- WeightedEnsemble_L2 was identified as the best-performing model, demonstrating significantly higher accuracy and AUC-ROC compared to other models in the AutoML framework.
- NeuralNetTorch_r79_BAG_L1 and XGBoost_BAG_L2 showed similar performance levels but were consistently outperformed by WeightedEnsemble_L2.

These results underscore the effectiveness of the ensembling and stacking strategies employed by the WeightedEnsemble_L2 model.

4.10. Robustness Evaluation with Bootstrapping

To further assess the model's reliability, a bootstrap approach was implemented. This method involved resampling the dataset 1000 times to estimate the variability in performance metrics.

- The mean accuracy and AUC-ROC from the bootstrap analysis closely matched those of the original metrics, demonstrating stability and consistency.
- The distribution of metrics across the resampled datasets exhibited minimal variability, further confirming the robustness of the model.

4.11. Implications of Statistical Findings

The combination of confidence intervals, paired t-tests, ANOVA, and bootstrapping provides strong evidence for the reliability and superiority of the AutoML-based model. These findings underscore the following key points:

- The narrow confidence intervals and significant *t*-test results confirm the model's consistent and superior performance compared with traditional approaches.
- The ANOVA results highlight the advantages of ensembling and model stacking, with WeightedEnsemble_L2 emerging as the most effective configuration.
- Bootstrapping ensures that the reported metrics are not artifacts of the specific dataset but represent true generalizable performance.

5. Discussion

This study developed an interpretable and generalizable machine learning model to predict asthma outcomes using AutoML and XAI techniques. By incorporating rigorous validation methods and interpretability tools, we addressed the key gaps identified in the literature: transparency in predictive models, generalizability across patient populations, and actionable insights for personalized medicine. The discussion below examines these findings, relates them to previous studies, and highlights their implications for clinical practice and future research.

5.1. Model Performance and Predictive Accuracy

The model demonstrated high predictive accuracy, achieving an accuracy rate of 98.99% and an AUC-ROC of 0.9996 on the validation dataset. These metrics confirm the model's ability to effectively differentiate between patients with varying levels of asthma exacerbation risk. This performance surpasses that of the traditional models discussed in prior studies, which often face challenges with interpretability and limited clinical utility [20,22].

The model's accuracy and balanced performance across classes address critical clinical needs, particularly in minimizing false negatives (missed exacerbations), which could lead to adverse outcomes. The use of the WeightedEnsemble_L2 model, leveraging multiple algorithms through ensembling, ensured robustness while mitigating overfitting—a common issue in healthcare-focused machine learning studies [27].

5.2. Interpretability and Explainability

Explainable AI techniques, such as SHAP and LIME, significantly enhanced the transparency of the model's predictions. Key features identified through SHAP analysis, such as age, Budesonide dosage, and pulmonary function metrics (PREFVC, POSFEV), align with known clinical predictors of asthma exacerbation [20,21].

In comparison to prior models that relied on rule-based or saliency map methods [32,34], SHAP and LIME provided granular, actionable insights at both global (feature-level) and local (patient-specific) levels. These capabilities are essential in clinical settings, where trust in AI-driven predictions is paramount [31,33]. The integration of interpretability tools ensures that clinicians can confidently rely on the model's outputs for decision-making, supporting its potential for broader clinical adoption.

While SHAP and LIME provide valuable insights into model predictions, their computational demands may hinder deployment in resource-constrained clinical settings. Future research should explore hybrid interpretability techniques that combine the granularities of SHAP and LIME using more efficient algorithms. Lightweight approximation methods or model-specific interpretation techniques may offer a balance between computational efficiency and interpretability.

5.3. Counterfactual Analysis for Personalized Medicine

A key innovation in this study is the incorporation of counterfactual analysis, which enables clinicians to simulate hypothetical scenarios by modifying specific variables (e.g., adjusting Budesonide dosage). This aligns with the principles of personalized medicine, offering actionable insights for tailoring treatments to individual patients [38,39].

While the counterfactual analysis highlighted valuable intervention strategies, it also revealed limitations. For example, features like age demonstrated minimal impact on predicted probabilities, underscoring the need to distinguish between modifiable and non-modifiable factors. Future research could enhance counterfactual techniques to better reflect real-world clinical interventions.

5.4. Generalizability and Model Robustness

Generalizability is a crucial challenge for machine learning models in healthcare, where consistent performance across diverse populations and clinical settings is essential [24,42,43]. To address this, the model was validated on the CAMP dataset in addition to the Kaggle dataset. The CAMP dataset provided a complementary perspective, encompassing longitudinal data on pediatric asthma patients with diverse demographic and clinical profiles. This dual-dataset approach ensured that the model's applicability extended beyond the training dataset.

Validation on the CAMP dataset demonstrated robust performance, with the model achieving an overall accuracy of 93% and an AUC-ROC of 0.97. Precision, recall, and F1 scores were balanced across the majority and minority classes, confirming the model's ability to generalize to a population distinct from the Kaggle dataset. Notably:

- Feature importance rankings, as derived from SHAP and LIME, remained consistent across both datasets, with age, Budesonide dosage, and pulmonary function metrics emerging as key predictors.
- Minimal deviations in performance metrics suggest that preprocessing techniques, particularly SMOTE, effectively mitigated class imbalance issues and reduced the risk of overfitting.

These results affirm the model's robustness in diverse real-world settings, including pediatric populations. The CAMP dataset's validation highlights the model's potential to support clinical decision-making across varied healthcare contexts. Furthermore, consistent feature importance rankings enhance the model's interpretability and clinician trust, which is critical for adoption in high-stakes environments.

5.5. Practical Implications for Clinical Integration

The development of a Streamlit-based application underscores the model's readiness for clinical integration. The app facilitates real-time predictions, interpretability insights, and counterfactual analysis, making it a practical tool for clinicians [30,36].

The model's high inference speed and accessible design enable seamless incorporation into electronic health record (EHR) systems. This integration has the potential to streamline asthma management workflows, allowing healthcare providers to quickly identify high-risk patients and tailor their treatment strategies.

5.6. Social Impact of the Study

Asthma disproportionately affects vulnerable populations, including low-income groups and those living in areas with poor air quality. This study's advancements in predictive modeling and interpretability can directly benefit these populations through the following:

- Personalized Interventions: Empowering clinicians to tailor treatments based on individual patient profiles, reducing exacerbation risks, and improving quality of life.
- Healthcare Equity: Enhancing prediction accuracy across diverse datasets ensures that underrepresented groups are not left behind in technological advancements.
- Public Health Insights: Highlighting environmental and socio-economic factors in asthma outcomes can guide policy interventions, such as air quality regulations or resource allocation for high-risk communities.

By addressing these social factors, this study underscores the potential of AI-driven healthcare solutions to bridge disparities and foster equitable healthcare delivery.

5.7. Limitations and Future Directions

This study addresses the critical limitations of existing asthma prediction models, particularly issues related to dataset diversity and interpretability. By utilizing XAI tools like SHAP and LIME, the proposed model strikes a balance between predictive accuracy and clinician trust, delivering actionable insights essential for personalized, patient-centered care. These tools empower clinicians by providing clear explanations for each prediction, fostering greater confidence in AI-driven recommendations [28,36].

The model's integration into clinical workflows via a Streamlit application enhances its usability. This application enables real-time predictions and visualizes the effects of potential interventions, supporting clinicians in designing tailored treatment plans and improving patient outcomes. By offering intuitive explanations and actionable insights, the tool reinforces trust in AI-powered healthcare solutions and facilitates informed decision-making.

Given the temporal nature of asthma exacerbations, future work could incorporate advanced techniques to model time-dependent dynamics. For instance:

- Recurrent Neural Networks (RNNs): These can capture sequential patterns in longitudinal data, enhancing predictions based on historical trends.
- Long Short-Term Memory Networks (LSTMs): These are particularly effective in modeling seasonal asthma variations and other time-dependent trends.
- Attention Mechanisms: These methods can highlight critical timeframes contributing to exacerbations, and providing clinicians with a deeper understanding of high-risk periods.
- Dynamic Time Warping (DTW): This approach can align and analyze non-uniform time series data, offering additional insights into patient-specific exacerbation patterns.

These methods would complement existing techniques like SHAP and LIME, enabling the integration of temporal dynamics into the model's reasoning and further improving the interpretability of predictions in chronic, dynamic conditions like asthma.

Looking forward, future efforts should focus on validating the model using multi-institutional datasets to ensure its generalizability across diverse populations and clinical settings. Expanding the model's application to other chronic conditions could also broaden its clinical utility. A key area for improvement lies in developing lightweight XAI techniques to minimize the computational demands of interpretability methods without compromising transparency. These enhancements would make the model more practical for deployment in resource-constrained environments, ensuring its effectiveness in real-world healthcare settings.

6. Conclusions

This study developed an interpretable and generalizable machine learning model for asthma prediction, addressing critical gaps in the literature related to model transparency, generalizability, and clinical applicability. By integrating AutoML techniques, SHAP, and counterfactual analysis, the model delivers high predictive accuracy while providing clinicians with tailored risk assessments and actionable insights. The inclusion of both global and local interpretability features empowers healthcare providers to make patient-specific decisions, ensuring that treatment strategies are aligned with individual patient profiles. Additionally, counterfactual analysis enables clinicians to explore the potential impact of interventions, such as adjusting medication dosages and fostering more personalized and precise asthma management.

The development of a Streamlit-based application further enhances the model's accessibility, ensuring its relevance in real-world clinical settings. This user-friendly tool supports clinicians by offering real-time predictions and interpretability insights, streamlining asthma management workflows, and facilitating proactive decision-making.

Emphasizing patient-centered care, the model bridges the gap between generic AI predictions and tailored clinical solutions. By allowing clinicians to simulate and assess the impact of potential interventions, the model ensures that decisions are made through a comprehensive understanding of each patient's unique condition. This approach aligns with the principles of personalized medicine, enhancing trust in AI-driven solutions and improving clinical outcomes in chronic respiratory conditions like asthma.

While the model demonstrates significant promise, future research should focus on expanding dataset diversity and refining interpretability methods to improve robustness and scalability. Exploring hybrid XAI techniques can reduce computational demands, making the model more feasible for use in resource-constrained environments. This study lays a strong foundation for integrating AI tools into practical healthcare applications, advancing personalized medicine, and supporting more precise and individualized asthma care.

Author Contributions: R.H. and R.A. conceptualized and designed the overall research framework; S.M. developed the experimental approach and contributed essential materials; R.H., R.A., S.M. and S.H. verified the accuracy and reliability of the results; S.M. performed the experiments and gathered data; R.H. organized and managed the research data; R.H. and S.M. wrote the initial draft of the manuscript; R.H. critically revised and improved the manuscript; R.H. prepared figures and visual representations; R.H. oversaw and supervised the research process; R.H. secured the financial support for the research. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed in this study are publicly available. The Kaggle dataset can be accessed at https://www.kaggle.com/datasets/jassonus122332/budesonide-nedocromil (accessed on 10 October 2024). Additionally, the Childhood Asthma Management Program (CAMP) dataset for generalization is available from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) (accessed on 10 November 2024). The dataset was last updated in March 2015 and can be accessed upon request from BioLINCC.

Acknowledgments: The authors would like to acknowledge the use of ChatGPT 24 May 2023 version (OpenAI, San Francisco, CA, USA), specifically to assist in some content for improved clarity and effectiveness.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abrams, E.M.; Akombi, B.; Alam, S.; Alcalde-Rabanal, J.E.; Allebeck, P.; Amini-Rarani, M.; Atnafu, D.D.; Ausloos, M.; Ayanore, M.A.; Azari, S.; et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **2020**, *396*, 1204–1222. [CrossRef]
2. Majellano, E.C.; Clark, V.L.; Winter, N.A.; Gibson, P.G.; McDonald, V.M. Approaches to the assessment of severe asthma: Barriers and strategies. *J. Asthma Allergy* **2019**, *12*, 235–251. [CrossRef]
3. Molfino, N.A.; Turcatel, G.; Riskin, D. Machine Learning Approaches to Predict Asthma Exacerbations: A Narrative Review. *Adv. Ther.* **2024**, *41*, 534–552. [CrossRef] [PubMed]
4. Khalifa, M.; Albadawy, M. Artificial Intelligence for Clinical Prediction: Exploring Key Domains and Essential Functions. *Comput. Methods Programs Biomed. Updat.* **2024**, *5*, 100148. [CrossRef]
5. Budiarto, A.; Tsang, K.C.H.; Wilson, A.M.; Sheikh, A.; Shah, S.A. Machine Learning–Based Asthma Attack Prediction Models From Routinely Collected Electronic Health Records: Systematic Scoping Review. *JMIR AI* **2023**, *2*, e46717. [CrossRef]
6. Weissler, E.H.; Naumann, T.; Andersson, T.; Ranganath, R.; Elemento, O.; Luo, Y.; Freitag, D.F.; Benoit, J.; Hughes, M.C.; Khan, F.; et al. The role of machine learning in clinical research: Transforming the future of evidence generation. *Trials* **2021**, *22*, 537. [CrossRef]

7. Luo, G.; Nkoy, F.L.; Stone, B.L.; Schmick, D.; Johnson, M.D. A systematic review of predictive models for asthma development in children. *BMC Med Inform. Decis. Mak.* **2015**, *15*, 99. [CrossRef] [PubMed]
8. Nasarian, E.; Alizadehsani, R.; Acharya, U.R.; Tsui, K. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Inf. Fusion* **2024**, *108*, 102412. [CrossRef]
9. Salehin, I.; Islam, M.S.; Saha, P.; Noman, S.M.; Tuni, A.; Hasan, M.M.; Baten, M.A. AutoML: A systematic review on automated machine learning with neural architecture search. *J. Inf. Intell.* **2024**, *2*, 52–81. [CrossRef]
10. Waring, J.; Lindvall, C.; Umeton, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* **2020**, *104*, 101822. [CrossRef] [PubMed]
11. Mustafa, A.; Rahimi Azghadi, M. Automated Machine Learning for Healthcare and Clinical Notes Analysis. *Computers* **2021**, *10*, 24. [CrossRef]
12. Ali, S.; Akhlaq, F.; Imran, A.S.; Kastrati, Z.; Daudpota, S.M.; Moosa, M. The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *Comput. Biol. Med.* **2023**, *166*, 107555. [CrossRef]
13. Sadeghi, Z.; Alizadehsani, R.; Cifci, M.A.; Kausar, S.; Rehman, R.; Mahanta, P.; Bora, P.K.; Almasri, A.; Alkhawaldeh, R.S.; Hussain, S.; et al. A review of Explainable Artificial Intelligence in healthcare. *Comput. Electr. Eng.* **2024**, *118*, 109370. [CrossRef]
14. Dickerman, B.A.; Hernán, M.A. Counterfactual prediction is not only for causal inference. *Eur. J. Epidemiol.* **2020**, *35*, 615–617. [CrossRef] [PubMed]
15. Nguyen, T.; Collins, G.S.; Landais, P.; Le Manach, Y. Counterfactual clinical prediction models could help to infer individualized treatment effects in randomized controlled trials—An illustration with the International Stroke Trial. *J. Clin. Epidemiol.* **2020**, *125*, 47–56. [CrossRef] [PubMed]
16. Al Meslamani, A.Z. How AI is advancing asthma management? Insights into economic and clinical aspects. *J. Med. Econ.* **2023**, *26*, 1489–1494. [CrossRef]
17. Yang, J.; Soltan, A.A.S.; Clifton, D.A. Machine learning generalizability across healthcare settings: Insights from multi-site COVID-19 screening. *Npj Digit. Med.* **2022**, *5*, 69. [CrossRef] [PubMed]
18. Sagheb, E.; Wi, C.; Yoon, J.; Seol, H.Y.; Shrestha, P.; Ryu, E.; Park, M.; Yawn, B.; Liu, H.; Homme, J.; et al. Artificial Intelligence Assesses Clinicians' Adherence to Asthma Guidelines Using Electronic Health Records. *J. Allergy Clin. Immunol. Prac.* **2021**, *10*, 1047–1056.e1. [CrossRef]
19. Goktas, P.; Damadoglu, E. Future of Allergy and Immunology: Is AI the Key in the Digital Era? *Ann. Allergy Asthma Immunol.* **2024**, in press. [CrossRef] [PubMed]
20. Sills, M.R.; Ozkaynak, M.; Jang, H. Predicting hospitalization of pediatric asthma patients in emergency departments using machine learning. *Int. J. Med. Inform.* **2021**, *151*, 104468. [CrossRef]
21. Harvey, J.L.; Kumar, S.A.P. Machine Learning for Predicting Development of Asthma in Children. In Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, 6–9 December 2019; pp. 596–603. [CrossRef]
22. Vatsal; Kumar, S.; Riya; Rampal, S.; Gaur, M.; Gaur, M. Advanced Ensemble Learning Approach for Asthma Prediction: Optimization and Evaluation. In Proceedings of the 2024 International Conference on Automation and Computation (AUTOCOM), São Paulo, Brazil, 14–16 March 2024; pp. 283–288. [CrossRef]
23. Darsha Jayamini, W.K.; Mirza, F.; Asif Naeem, M.; Chan, A.H.Y. Investigating Machine Learning Techniques for Predicting Risk of Asthma Exacerbations: A Systematic Review. *J. Med. Syst.* **2024**, *48*, 49. [CrossRef]
24. He, P.; Moraes, T.J.; Dai, D.; Reyna-Vargas, M.E.; Dai, R.; Mandhane, P.; Simons, E.; Azad, M.B.; Hoskinson, C.; Petersen, C.; et al. Early prediction of pediatric asthma in the Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort using machine learning. *Pediatr. Res.* **2024**, *95*, 1818–1825. [CrossRef]
25. Li, D.; Abhadiomhen, S.E.; Zhou, D.; Shen, X.; Shi, L.; Cui, Y. Asthma prediction via affinity graph enhanced classifier: A machine learning approach based on routine blood biomarkers. *J. Transl. Med.* **2024**, *22*, 100. [CrossRef] [PubMed]
26. Lee, C.-C.; Park, J.Y.; Hsu, W.-T. Bridging expertise with machine learning and automated machine learning in clinical medicine. *Ann. Acad. Med.* **2024**, *53*, 129–131. [CrossRef] [PubMed]
27. Nwaimo, C.S.; Adegbola, A.E.; Adegbola, M.D. Transforming healthcare with data analytics: Predictive models for patient outcomes. *GSC Biol. Pharm. Sci.* **2024**, *27*, 25–35. [CrossRef]
28. Priyanka, E.; Thangavel, S.; Mohanasundaram, R.; Subramaniam, S. Artificial Intelligence Approaches in Healthcare Informatics Toward Advanced Computation and Analysis. *Open Biomed. Eng. J.* **2024**, *18*, 1–21. [CrossRef]
29. Nisanova, A.; Yavary, A.; Deaner, J.; Ali, F.S.; Gogte, P.; Kaplan, R.; Chen, K.C.; Nudleman, E.; Grewal, D.; Gupta, M.; et al. Performance of Automated Machine Learning in Predicting Outcomes of Pneumatic Retinopexy. *Ophthalmol. Sci.* **2024**, *4*, 100470. [CrossRef]
30. Scott, I.A.; De Guzman, K.R.; Falconer, N.; Canaris, S.; Bonilla, O.; McPhail, S.M.; Marxen, S.; Van Garderen, A.; Abdel-Hafez, A.; Barras, M. Evaluating automated machine learning platforms for use in healthcare. *JAMIA Open* **2024**, *7*, ooae031. [CrossRef] [PubMed]

31. Thalpage, N. Unlocking the Black Box: Explainable Artificial Intelligence (XAI) for Trust and Transparency in AI Systems. *J. Digit. Art Humanit.* **2023**, *4*, 31–36. [CrossRef] [PubMed]
32. Khakurel, U.; Rawat, D.B. Evaluating explainable artificial intelligence (XAI): Algorithmic explanations for transparency and trustworthiness of ML algorithms and AI systems. In Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations 2022, Orlando, FL, USA, 3–7 April 2022; Volume 12113, p. 121130A-11. [CrossRef]
33. Khan, M.A.; Khan, R.; Praveen, P.; Verma, A.R.; Panda, M.K. Machine Learning for Smart Health Services in the Framework of Industry 5.0. In *Infrastructure Possibilities and Human-Centered Approaches with Industry 5.0*; IGI Global: Hershey, PA, USA, 2024; pp. 215–230. [CrossRef]
34. Zhang, X.; Luo, G. Ranking Rule-Based Automatic Explanations for Machine Learning Predictions on Asthma Hospital Encounters in Patients With Asthma: Retrospective Cohort Study. *JMIR Med. Inform.* **2021**, *9*, e28287. [CrossRef] [PubMed]
35. Neelam; Kamlesh, K. Explainable AI: Interpretable Models for Transparent Decision-Making. *Int. J. Food Nutr. Sci.* **2024**, *9*, 239–244. [CrossRef]
36. Deshpande, P.V.A.R.S. Interpretable Deep Learning Models: Enhancing Transparency and Trustworthiness in Explainable AI. *Proceeding Int. Conf. Sci. Eng.* **2023**, *11*, 1352–1363. [CrossRef]
37. Hakizimana, A.; Devani, P.; Gaillard, E.A. Current technological advancement in asthma care. *Expert Rev. Respir. Med.* **2024**, *18*, 499–512. [CrossRef]
38. Ray, A.; Das, J.; Wenzel, S.E. Determining asthma endotypes and outcomes: Complementing existing clinical practice with modern machine learning. *Cell Rep. Med.* **2022**, *3*, 100857. [CrossRef] [PubMed]
39. Ho, D.; Quake, S.R.; McCabe, E.R.; Chng, W.J.; Chow, E.K.; Ding, X.; Gelb, B.D.; Ginsburg, G.S.; Hassenstab, J.; Ho, C.-M.; et al. Enabling Technologies for Personalized and Precision Medicine. *Trends Biotechnol.* **2020**, *38*, 497–518. [CrossRef] [PubMed]
40. Chang, L.; Zhao, K. Construction and validation of an innovative prognostic nomogram for overall survival in cervical cancer patients with lung metastasis: An analysis utilizing the SEER database. *Front. Oncol.* **2024**, *14*, 1397454. [CrossRef]
41. Cao, J.; Zhang, X.; Shahinian, V.; Yin, H.; Steffick, D.; Saran, R.; Crowley, S.; Mathis, M.; Nadkarni, G.N.; Heung, M.; et al. Generalizability of an acute kidney injury prediction model across health systems. *Nat. Mach. Intell.* **2022**, *4*, 1121–1129. [CrossRef]
42. Zazara, D.E.; Giannou, O.; Schepanski, S.; Pagenkemper, M.; Giannou, A.D.; Pincus, M.; Belios, I.; Bonn, S.; Muntau, A.C.; Hecher, K.; et al. Fetal lung growth predicts the risk for early-life respiratory infections and childhood asthma. *World J. Pediatr.* **2024**, *20*, 481–495. [CrossRef] [PubMed]
43. Karaarslan, E.; Konacaklı, E. Decentralized solutions for data collection and privacy in healthcare. *Artif. Intell. Data-Driven Med. Diagn.* **2021**, *3*, 167–190. [CrossRef]
44. Maftoun, M.; Joloudari, J.H.; Zare, O.; Khademi, M.; Atashi, A.; Nematollahi, M.A.; Alizadehsani, R.; Gorriz, J.M. Improving Prediction of Mortality in ICU via Fusion of SelectKBest with SMOTE Method and Extra Tree Classifier. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*; Springer Nature: Cham, Switzerland, 2024; Volume 14674, pp. 68–77. [CrossRef]
45. Erickson, N.; Mueller, J.; Shirkov, A.; Zhang, H.; Larroy, P.; Li, M.; Smola, A. AutoGluon-tabular: Robust and accurate AutoML for structured data. *arXiv* **2020**, arXiv:2003.06505.
46. Dewage, K.A.K.W.; Hasan, R.; Rehman, B.; Mahmood, S. Enhancing Brain Tumor Detection Through Custom Convolutional Neural Networks and Interpretability-Driven Analysis. *Information* **2024**, *15*, 653. [CrossRef]
47. Mathotaarachchi, K.V.; Hasan, R.; Mahmood, S. Advanced Machine Learning Techniques for Predictive Modeling of Property Prices. *Information* **2024**, *15*, 295. [CrossRef]
48. Childhood Asthma Management Program Research Group the Childhood Asthma Management Program (CAMP): Design, rationale, and methods. *Control. Clin. Trials* **1999**, *20*, 91–120. Available online: https://www.ncbi.nlm.nih.gov/pubmed/10027502 (accessed on 10 November 2024).