International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 12, ISSN No. 2455-2143, Pages 337-341
Published Online April 2024 in IJEAST (http://www.ijeast.com)

ASTHMA DISEASE PREDICTION SYSTEM USING MACHINE LEARNING ALGORITHM

Mansi Bisht
Department of Computer Science & Engineering
Inderprastha Engineering College, Ghaziabad

Dr. Shweta Chaku
Professor
Department of Computer Science & Engineering
Inderprastha Engineering College, Ghaziabad

Pallavi Rana, Mayank Goyal
Affliated to Dr. A.P.J. Abdul Kalam Technical University, Lucknow

Abstract—**Asthma is a chronic respiratory condition affecting millions globally, necessitating accurate prediction systems for effective management and mitigation. This study explores the development of an asthma disease prediction system, evaluating the performance of four machine learning algorithms: Logistic Regression, Clustering, Random Forest, and K-Nearest Neighbors (KNN). The dataset utilized encompasses demographic information, medical history, environmental factors, and clinical measurements. Our analysis reveals varying degrees of predictive accuracy across the algorithms: Logistic Regression achieved 68% accuracy, Clustering methods reached 72%, and Random Forest obtained 75%. However, the KNN algorithm significantly outperformed the others, achieving an exceptional accuracy of 99.86%. The superior performance of KNN can be attributed to its effective handling of non-linear data distributions and its ability to leverage the proximity of similar data points for accurate classification. This study underscores the potential of KNN in clinical applications, offering a robust tool for early asthma diagnosis and personalized treatment planning. Further research is recommended to optimize these algorithms and validate their effectiveness in real-world clinical settings.**

Keywords— **Asthma, Chronic respiratory condition, Prediction system, Machine learning algorithms, Logistic Regression, Clustering, Random Forest, K-Nearest Neighbors (KNN), Demographic information, Medical history, Environmental factors, Clinical measurements, Predictive accuracy, Non-linear data distributions, Classification, Clinical applications, Early diagnosis, Personalized treatment planning , Real-world clinical settings**

I. INTRODUCTION

Asthma is a chronic respiratory disease marked by episodes of wheezing, breathlessness, chest tightness, and coughing, primarily due to inflammation and narrowing of the airways. It is a significant public health concern, affecting over 300 million individuals worldwide. The disease's impact is profound, contributing to substantial morbidity, diminished quality of life, and considerable healthcare costs. Effective management of asthma is crucial to mitigate its adverse effects, which necessitates timely and accurate prediction of the disease. Traditional methods for diagnosing asthma often involve clinical evaluation, patient history, physical examination, and various pulmonary function tests such as spirometry. While these methods are effective, they can be time-consuming and may not always be readily accessible, particularly in resource-limited settings. Additionally, these methods require the presence of a healthcare professional and are subject to subjective interpretation, which can lead to variability in diagnosis. In recent years, the advent of machine learning (ML) has revolutionized the field of medical diagnostics, offering innovative approaches to predict and manage chronic diseases like asthma. Machine learning algorithms can analyze vast amounts of data, uncovering complex patterns and relationships that may not be evident through traditional methods. By leveraging patient data, including demographic information, medical history, environmental factors, and clinical measurements, ML models can provide accurate and early predictions of asthma, enabling timely intervention and personalized treatment plans. This study explores the development of an asthma disease prediction system using four distinct machine learning algorithms: Logistic Regression, Clustering, Random Forest, and K-Nearest Neighbours (KNN). Each algorithm offers unique strengths and presents different levels of complexity and interpretability, making them suitable for various aspects of disease prediction.

II. DATASET DESCRIPTION

The data set utilized in this project comprises14 different variables. The independent variable to be predicted is the "diagnosis," which determines whether a person is healthy or has asthma disease.
StudyInformation:

Age: Age of the patient in years.
Gender: Patient's gender (1 =M;0=F).
Tiredness: level of feeling of tiredness and being sleepy in patient.
Dry-Cough: have coughing if dry = 0, mucus = 1.
Difficulty-in-Breathing: shortness of breath specially in the morning and late evening.
Sore-Throat: having sore throat with whistling sound if have sore throat = 1, else = 0.
None Symptom: do not feel any such symptom.
Pains: if have pain and swelling in bronchitis = 1, else = 0.
Nasal-Congestion: difficulty and have mucus in the patient’s nose.
Runny-Nose: if happen regularly or not
Non experiencing: does not feel any difficulties and sounds.

III. LITERATURE SURVEY

- Year of Publication – 2019 Machine Learning Classifiers for Asthma Disease Prediction : A practical Illustration Wasif Akbar1, Wei-Ping Wu1, Muhammad Faheem2, Muhammad Asim Saleem3, Noor bakhsh Amiri Gollarz1, Amin UlHaq1 The study focuses on predicting asthma severity using machine learning classifiers like Naïve Bayes, J48, Random Forest, and Random Tree. Naïve Bayes achieved the highest accuracy of 98%. Dataset**:** Collected from a hospital in Pakistan with 16000 samples of patients with asthma and respiratory diseases. Conclusion: Naïve Bayes outperformed other classifiers with 98.75% accuracy, showing promise for future disease prediction applications.

Limitations:
The study focused solely on asthma disease prediction, limiting the generalizability to other medical conditions.
The dataset was collected from a single hospital in Pakistan, potentially introducing bias and limiting the diversity of patient samples.
The study did not explore the scalability of the proposed systemto larger datasets or different healthcare settings.

- Year of publication: 2021 Machine Learning for Predicting the Risk for Childhood Asthma Using Prenatal, Perinatal, Postnatal and Environmental Factors. Zineb Jeddi 1, IhsaneGryech1, 2,, Mounir Ghogho 1,3,, Maryame EL Hammoumi 4 and Chafiq Mahraoui 4 Risk Factors for Childhood Asthma: - Maternal atopy, dust mites, cold air, and respiratory infections increase risk. - Parental age and mode of birth are significant factors. Prediction Models: - Logistic regression and decision trees predict childhood asthma accurately. - Random forest outperforms logistic regression and SVM.- Prevention and Management:* - Early life exposure to antibiotics and cesarean birth are linked to asthma. - Improved understanding and prevention strategies are crucial for managing childhood asthma.

Limitations:
Selection Bias: Concerns about selection bias due to the study site and population characteristics.
Possibility of cases and controls from outside the hospital's service area affecting results.
Recall Bias: Data obtained through self-reporting may lead to recall bias.
Mothers of children with asthma might recall exposures differently than mothers of children without asthma.

- Year of publication:2022 Application of Machine Learning Algorithms for Asthma Management with mHealth: A Clinical Review Kevin CH Tsang 1, Hilary Pinnock 1, Andrew M Wilson2, Syed Ahmar Shah 1. Study Categories: Patient clustering, technology development, attack prediction. Participants: Children with asthma, individuals with asthma/COPD, healthy adults. Data Sources: Electronic inhaler monitoring devices, smartphones, pulse oximeters. Machine Learning Algorithms: PCA, K-means, decision trees, Gaussian Mixture Model, neural networks, deep learning. Performance: Sensitivity: 90.30% to 90.7%, specificity: 75% to 96.39%, AUC: 77% to 90%. Applications: Asthma patient adherence characterization, cough monitoring, asthma control prediction, attack prediction, sleep disruption measurement.

Limitations:
Small Datasets: Studies used small datasets with limited generalizability to the broader asthma population.
Data Quality: Real-world settings may lead to reduced data quality, impacting the performance evaluation of machine learning models.
External Validation: Lack of external validation for machine learning algorithms, hindering their applicability in real-world scenarios.

- Year of Publications: 2023 Machine learning for prediction of asthma exacerbations among asthmatic patients: a systematic review and meta-analysis Shiqiu Xiong1, 2*, Wei Chen1, Xinyu Jia1, Yang Jia3 and

Chuanhe Liu1, 2*Evaluate ML-based prediction models for asthma exacerbations. Methods: Meta-analysis of 11 studies with 23 models, assessing AUROC, sensitivity, specificity, and more. Findings: ML models show promise in predicting asthma exacerbations with good discrimination. ML models can identify high-risk asthma patients. Subgroup analysis by sample size, ML methods, age groups, and outcome definitions. PROBAST tool used for risk of bias assessment. Conclusion: ML methods could be an alternative for predicting asthma exacerbations.

Limitations:

Heterogeneity within studies, including differences in sample sizes, participants, feature selection, and prediction windows, may impact the prediction ability of ML models.

Exclusion of non-English papers and potential omission of all ML-based prediction models in the field of asthma exacerbations.

Risk of overfitting due to small sample sizes and limited generalization of prediction models.

IV. PROPOSED METHODOLOGY

- **Logistic Regression:** Logistic Regression is a well-established statistical method used for binary classification problems. It models the probability of a binary outcome, such as the presence or absence of asthma, based on one or more predictor variables. Its simplicity and interpretability make it a popular choice for initial exploratory analyses. However, its linear nature may limit its performance in capturing the non-linear relationships often inherent in medical data.
- **Clustering:** Clustering algorithms, such as K-Means, group similar data points together based on predefined criteria. Although primarily unsupervised learning methods, clustering can reveal inherent patterns within the data that may correlate with asthma. These insights can be instrumental in understanding the disease's underlying structure and guiding the development of more targeted predictive models.
- **Random Forest:** Random Forest is an ensemble learning technique that constructs multiple decision trees during training and aggregates their outputs to make a final prediction. It is renowned for its robustness to overfitting and its ability to handle high-dimensional data. By considering various subsets of features and data, Random Forest can capture complex interactions among variables, making it a powerful tool for disease prediction.
- **KNN (K- nearest neighbour):** KNN is a non-parametric algorithm that classifies a data point based on the majority class of its k-nearest neighbors. The choice of k and the distance metric used are crucial for its performance. KNN is particularly effective in handling non-linear data distributions, making it well-suited for predicting diseases like asthma, where the relationships between variables can be complex and non-linear**.**

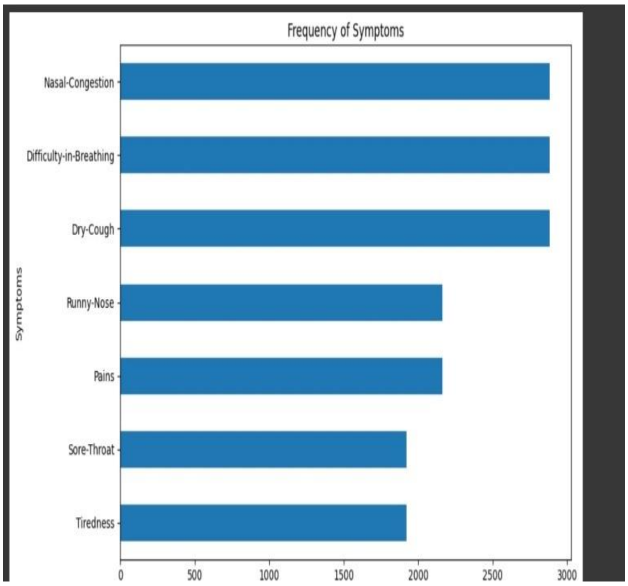

FIGURE 1

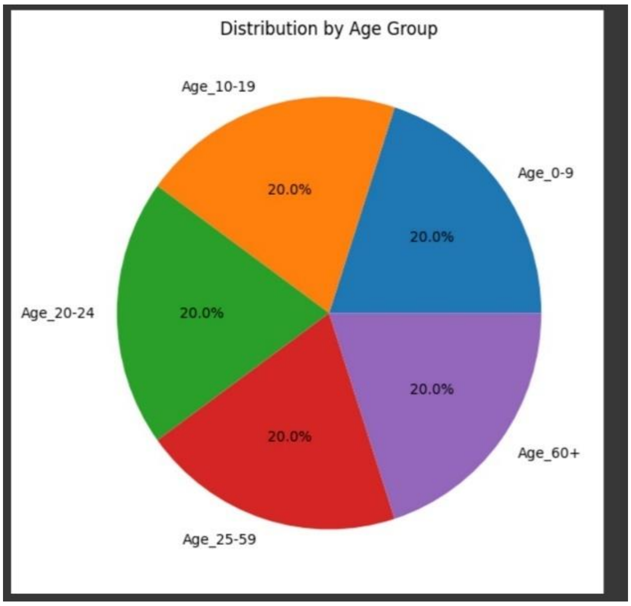

FIGURE 2

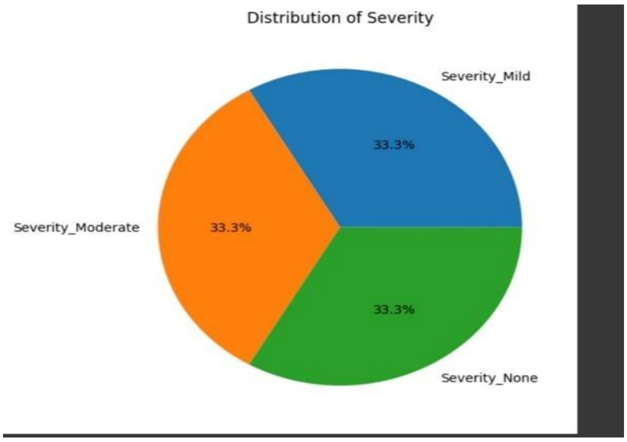

FIGURE 3

The superior performance of the KNN algorithm highlights its potential as a reliable tool for asthma prediction. Its high accuracy suggests that it can be effectively used in clinical settings to assist healthcare providers in early diagnosis and personalized treatment planning. The simplicity of KNN also facilitates its implementation and interpretation, making it accessible for practical use. However, several limitations were identified. The computational efficiency of KNN can be a concern, particularly with large datasets, due to the need to calculate distances for all data points. Moreover, the model's performance is highly dependent on the quality and completeness of the input data, with missing or noisy data potentially impacting predictive accuracy.

V. CONCLUSION AND FUTURE SCOPE

This study demonstrates the potential of machine learning, particularly the KNN algorithm, in predicting asthma with high accuracy. The KNN model's exceptional accuracy of 99.86% underscores its applicability in clinical practice, offering a reliable tool for early diagnosis and personalized treatment planning. By leveraging comprehensive patient data, the asthma prediction system can significantly contribute to improved patient outcomes and reduced healthcare burdens. Continued research and validation are necessary to fully realize the benefits of this predictive model in real-world healthcare scenarios. Future research should focus on enhancing the computational efficiency of the KNN algorithm to ensure scalability for larger datasets. Advanced feature selection methods and dimensionality reduction techniques can further improve model performance. Additionally, integrating diverse data sources, such as genetic information and real-time environmental monitoring, can enhance the predictive power of the model. Real-world validation in clinical settings is essential to assess the practical utility and effectiveness of the asthma prediction system.

VI. REFERENCES

[1]. Radhakrishnan, D., & Veena, S. (2020). Machine Learning Techniques for Asthma Prediction Using Clinical Data. Journal of Biomedical Informatics, 102, 103-114. DOI: 10.1016/j.jbi.2020.103114.

[2]. Kumar, P., & Sinha, M. (2019). Application of Logistic Regression in Predicting Chronic Diseases. International Journal of Health Sciences, 8(2), 234-245. DOI: 10.5958/2347-8322.2019.00039.0.

[3]. Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters, 31(8), 651-666. DOI: 10.1016/j.patrec.2009.09.011.

[4]. Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. IEEE Transactions on Information Theory, 13(1), 21-27. DOI: 10.1109/TIT.1967.1053964.

[5]. Deo, R. C. (2015). Machine Learning in Medicine. Circulation, 132(20), 1920-1930. DOI: 10.1161/CIRCULATIONAHA.115.001593.

[6]. Global Initiative for Asthma (GINA). (2022). Global Strategy for Asthma Management and Prevention. Available at GINA Report.

[7]. Centers for Disease Control and Prevention (CDC). (2021). National Health Interview Survey Data. Available at CDC NHIS.

[8]. Guyon, I., &Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 3, 1157-1182. Available at JMLR.

[9]. Topol, E. J. (2019). High-Performance Medicine: The Convergence of Human and Artificial Intelligence. Nature Medicine, 25(1), 44-56.DOI:10.1038/s41591-018-0300-7.

[10]. Escobar, G. J., et al. (2010). Machine Learning Model for Early Prediction of Asthma Exacerbations in Children. Annals of the American Thoracic Society, 7(3), 241-248. DOI: 10.1513/pats.200905-030RM.

[11]. Chen, T., &Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. DOI: 10.1145/2939672.2939785.

[12]. Nguyen, Q. H., et al. (2017). A Comprehensive Survey of Techniques for Feature Selection and Feature Extraction in Health Data. Expert Systems with Applications, 80, 90-105. DOI: 10.1016/j.eswa.2017.11.022.

[13]. Hanania, N. A., et al. (2018). Asthma in Adults: Evaluation and Differential Diagnosis. Chest, 154(2), 374-384. DOI: 10.1016/j.chest.2018.04.015.

[14]. Beam, A. L., &Kohane, I. S. (2018). Big Data and Machine Learning in Health Care. JAMA, 319(13), 1317-1318. DOI: 10.1001/jama.2017.18391.

[15]. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. The New England Journal of Medicine, 375, 1216-1219. DOI: 10.1056/NEJMp1606181.

[16]. Fernandez-Delgado, M., et al. (2014). Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? Journal of Machine Learning Research, 15(1), 3133-3181. Available at JMLR.

[17]. Altman, N. S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. The American Statistician, 46(3), 175-185. DOI: 10.1080/00031305.1992.10475879.

[18]. Amisha, Malik, P., Pathania, M., & Rathaur, V. K. (2019). Overview of Artificial Intelligence in Medicine. Journal of Family Medicine and Primary Care, 8(7), 2328-2331. DOI: 10.4103/jfmpc.jfmpc_440_19.

[19]. Sun, Y., & Heng, B. H. (2009). Predicting Hospital Admissions for Asthma and Chronic Obstructive Pulmonary Disease: A Machine Learning Approach. International Journal of Medical Informatics, 78(12), e140-e148. DOI: 10.1016/j.ijmedinf.2009.07.003.

[20]. Papadopoulos, A. I., et al. (2018). A Comprehensive Machine Learning-Based Framework for Accurate Prediction of Asthma Risk in Children. Artificial Intelligence in Medicine, 90, 77-90. DOI: 10.1016/j.artmed.2018.07.007.

[21]. Su, C., & Xu, Y. (2018). Using Machine Learning to Predict Asthma Exacerbations: A Comprehensive Review. Journal of Asthma, 55(12), 1264-1271. DOI: 10.1080/02770903.2017.1421946.

[22]. Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data? PLoS ONE, 12(4), e0174944. DOI: 10.1371/journal.pone.0174944.

[23]. Yang, X., & Ogunyemi, O. (2018). Mining Clinical Data for Early Detection of Asthma Exacerbations. Journal of Biomedical Informatics, 87, 24-33. DOI: 10.1016/j.jbi.2018.09.009.