INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 4 - V7I4-1680)

Available online at: https://www.ijariit.com

Asthma prediction using Machine Learning

Manasa G. V. Kumar
manasagvkumar3@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Yogesh R
Yogeshv1112@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Soumya ranjan nayak
nayaksoumya719@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Vinod R
vinodvinu567@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Shreyas M S
shreyasiyengar619@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

ABSTRACT

Patient telemonitoring brings about a conglomeration of huge measures of data about quiet illness direction. Notwithstanding, the possible utilization of this data for the early expectation of asthma in grown-ups has not been methodically assessed. The point of this examination was to investigate the information for building AI calculations that anticipate asthma before they happen. The investigation dataset involved 278847 records presented by grown-up asthma patients. Prescient displaying included readiness of preparing informational indexes, prescient component choice, and assessment of coming about classifiers. AI classifiers are utilized to foster these prescient models; including Random Forest, Logistic Regression, Decision Tree, and Naïve Bayes strategy. Of the multitude of classifiers carried out, strategic relapse classifier brought about the most elevated expectation precision. Our investigation showed that AI methods have huge potential in creating customized choice help for ongoing illness telemonitoring frameworks. Future examinations may profit with a far-reaching prescient system that consolidates information with different elements influencing the probability of creating asthma. Approaches carried out for cutting edge asthma expectations might be stretched out to early mediation of persistent ailments in patients.

Keywords: *Decision Tree, Random Forest, Naïve Bayes, Logistic Regression*

1. INTRODUCTION

Asthma, a chronic inflammatory disorder that affects the airways, is characterized by obstruction of airflow. It can be treated with or without therapy. Interactions between cells, cellular elements and cytokines are the cause of airway inflammation. Recurrent or persistent bronchospasm can occur in those who are susceptible to the disease. This includes symptoms such as wheezing and breathlessness. Asthma can be characterized by a polymorphic phenotype that is affected by many environmental and genetic factors. These factors play an important role in the development of the disease and its persistence. These factors include a family history of asthma, allergic rhinitis or atopic dermatitis, wheezing episodes in childhood, maternal smoking during pregnancy and several other prenatal and environmental factors.

Most asthma sufferers develop symptoms by the age of five years. Because the symptoms of asthma are so similar, it can be difficult to distinguish between them and other wheezing disorders. As a result, asthma can often be mistakenly diagnosed as bronchiolitis, common cold, or pneumonia. A detailed medical history and physical exam are required to diagnose asthma. However, a lung function test can be difficult to perform in children under five years of age. Preventive medicine is all about identifying those who are most at risk and who need intervention. Particularly important in the case asthma is the accuracy of the risk classification. Patients at high risk of developing asthma disease may be identified early, which can lead to better treatment options and hopefully better outcomes for patients in adulthood.

Asthma is a common disease that affects millions of people around the globe. It can be described clinically as a combination of variable symptoms and significant changes in the function of the lungs. This disease is diagnosed using international guidelines. It includes symptoms like wheezing, shortness or coughing, and shortness of breath. These symptoms are not associated with asthma and 30% of asthma patients are misdiagnosed when a doctor relies on only the symptoms. Misdiagnosis can lead to incorrect treatment, and possibly financial and physical complications.

2. PROPOSED METHODOLOGY

The motivation behind this study is to make a relative model for recognizing highlights that are the most characteristic of the advancement of asthma in youngsters. The proposed model is not difficult to use for clinical experts and over comes the restriction of existing framework, for example, the way that

current framework couldn't adjust to the outside conditions and the precision of recognizable proof is less just as tedious.

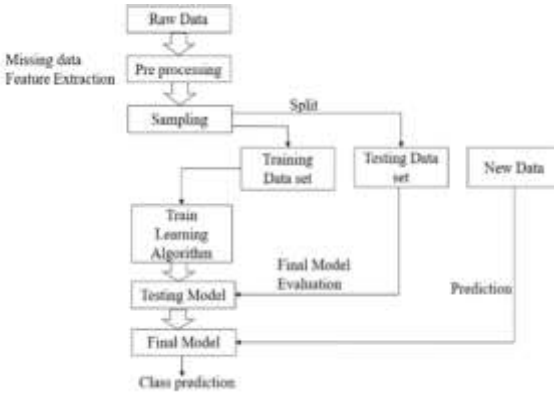

Fig-1 Architecture of asthma predictor

There are restricted investigations that address the utilization of an assortment of AI prescient models in foreseeing asthma with high precision on huge datasets. A significant part of the examination is obsolete and strategies for expectation of the improvement of asthma didn't utilize AI.

The inspiration for this examination is because of the adaptability of AI in assortment of field. This paper plans to decide a viable model for recognizing factors that can be utilized to foresee asthma advancement. The Center for Disease Control announced in 2016 that 8.3% of American youngsters right now have asthma. As it is hard to segregate asthma from other wheezing problem due to the similitude in indications so an exact model that can be utilized to anticipate whether an individual will foster asthma dependent on data gathered, can end up being progressive to diminish the danger factors.

3. MODULES AND ANALYSIS

The implementation part of the project consists of different modules. Simply, a module is a file consisting of Python code. A module can define functions, classes and variables. A module can also include runnable code.

Module 1: Data Acquisition and Pre-processing

The dataset used in this study consisted of 278847 records collected from asthma patients. The data was collected from the online repository. The severity of asthma varied from mild persistent to severe persistent. The data uploaded as "AsthmPrediction.csv" using the python libraries and further the data is pre-processed. The pre-processing step is a very crucial stage where the data is cleaned, which makes the data set more efficient that requires less storage required and removes the unwanted data.

Information pre-processing in Machine Learning alludes to the strategy of planning (cleaning and coordinating) the raw information to make it appropriate for a structure and preparing Machine Learning models. In straightforward words, information pre-processing in Machine Learning is an information mining procedure that changes raw data into a justifiable and understandable configuration. This step includes:

- Acquiring the data.
- Importing all the crucial libraries that are NumPy, pandas, matplotlib.
- Importing the data set where the data set includes 9 attributes which are age, gender, sleeping problem, chest tightness, breathing problem, cough, allergies, wheezing and asthma. Here the age is a continuous variable whereas all the other attribute is categorical so we will encode all the attributes.
- Identifying and handling the missing values.
- Encoding the categorical values to numbers before we can use it to fit and evaluate the model.
- Identifying the dependent and independent variable to find the possible effect on the dependent variable by changing the independent variable.
- The following stage in information pre-handling is dividing the informational collection. Each dataset for Machine Learning model ought to be separated into two separate sets – planning set and test set. Preparing set means the subset of a dataset that is utilized for preparing the AI model. A test set, then again, is the subset of the dataset that is utilized for testing the AI model.

Module 2: Data Modelling

in this step we are building asthma prediction models based on a number of features. We are performing three sets of experiments: train and test a classifier with the original data, train and test classifiers using a stratified sample with equal representation of each class, train a classifier with the training data from the stratified sample and test data from the rest of the dataset.

Four classification algorithms were used for building classification models: Random Forest, naive Bayesian classifier, logistic regression, decision tree. This step includes importing the classifiers from sklearn and training the model using the ". fit ()" function and saving the trained model.

Random forest is an outfit procedure equipped for performing both relapse and grouping errands with the utilization of numerous choice trees. The fundamental thought behind this is joining different choice trees in deciding the last yield as opposed to depending on singular choice trees.

Decision tree is the most remarkable and well known device for grouping and forecast. A Decision tree is a flowchart like tree structure, where each inner hub signifies a test on a characteristic, each branch addresses a result of the test, and each leaf hub (terminal hub) holds a class name.

Logistic regression is fundamentally a regulated characterization calculation. In an order issue, the objective variable (or yield) can take just discrete qualities for a given arrangement of highlights (or sources of info).

Naive Bayes is a gathering of regulated AI grouping calculations dependent on the Bayes hypothesis. It is a straightforward grouping method, however has high usefulness. Gaussian Naive Bayes is a variation of Naive Bayes that follows Gaussian ordinary dispersion.

Module 3: Prediction model evaluation and testing

The evaluation of each model was captured in terms of the number of values for true positive (TP), true negative (TN), false negative (FN), and false positive (FP), which subsequently were converted into overall accuracy, sensitivity.

A confusion matrix is a table used to describe the performance or "classifier" of a classification model on test data. The true values of the models are known.

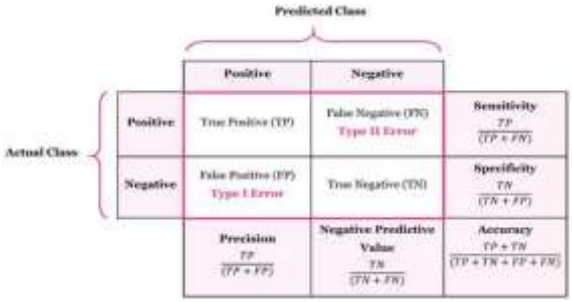

Chart-1 Confusion Matrix

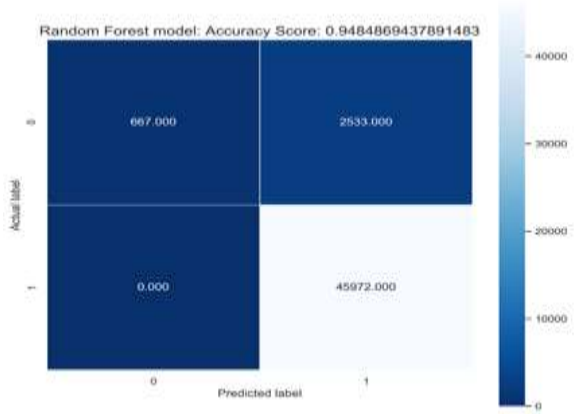

Chart-2 Random Forest Confusion Matrix

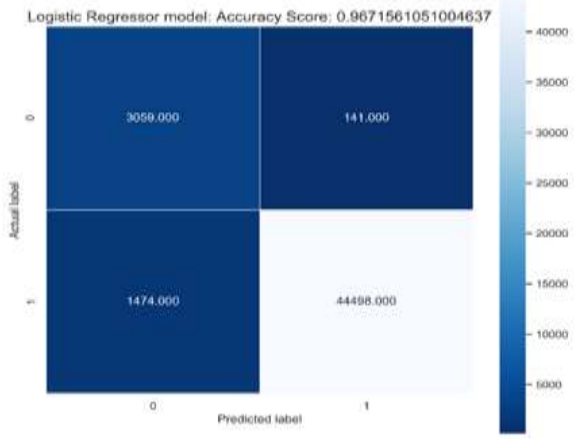

Chart-3 Logistic regression

Chart-4 Naïve Bayes Confusion matrix

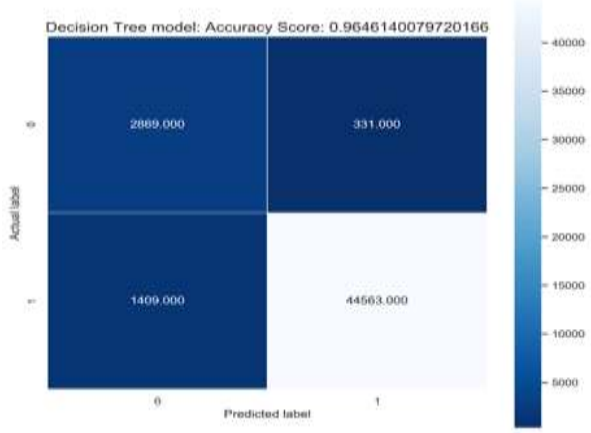

Chart-5 Decision tree confusion matrix

Module 4: UI model

The UI model provides a user interface for the designed classifier model to the user. Python provides the various features for creating an interface. Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit.

4. RESULTS AND DISCUSSION

The final output of then project is the UI module, this module provides a window that collects the require information of the patients for the prediction. The details include age, gender, sleeping problem, breathing problem, coughing, allergies, wheezing problem. The instruction to provide the details is also include in the UI according to the instruction, individual must put the values to avoid any error message.

The select algorithm option gives the option to select an algorithm for prediction and compare between the results.

Fig-2 welcome and instruction window

Fig-3 output with a graph showing the accuracy of prediction

5. CONCLUSION

The use of machine learning in asthma detection (i,e healthcare field) is still limited. It is important to detect asthma early so that treatment can be started. Our method is efficient in predicting the condition of patients with chronic asthma. Different classifiers are used to train all patient data. All experiments were conducted using an open-source machine-learning toolkit. Different parameters were used to evaluate the effectiveness of these classifiers. The Logistic Regression is the most accurate classifier. These results show that data analysis can be used to identify highlights that affect the advancement of asthma. The proposed framework with AI arrangement calculations could be utilized for future illnesses forecast. The work can likewise be stretched out by the computerization of asthma illness examination including some other ML algorithm calculation.

6. FUTURE ENHANCEMENTS

This paper's arrangement models can be utilized to foresee other wellbeing estimates utilizing similar informational index or diverse datasets. The models can be utilized by medical services experts to choose highlights to decide whether they have any relationships and decide how those variables could be utilized to foresee asthma advancement in kids. These models can be

utilized to assist kids with getting precaution care, finding and treatment. It is fundamental to analyze asthma right off the bat to further develop the patient's personal satisfaction. It would be intriguing to analyze the late beginning asthma in more seasoned grown-ups and the old to discover authentic variables that could be utilized to anticipate when asthma will foster further down the road.

7. REFERENCES

[1] O. Kocsis and G. Arvanitis, "Multi-model Short-term Prediction Schema for mHealth Empowering Asthma," Electron. Notes Theor. Comput. Sci., vol. 343, pp. 3–17, 2019.

[2] J. S. Moldenhauer et al., "Influence of maternal asthma and asthma severity on newborn morphometry," J. Asthma, vol. 47, no. 2, pp. 145–149, 2010.

[3] J. Milagro et al., "Nocturnal Heart Rate Variability Spectrum Characterization in Preschool Children with Asthmatic Symptoms," vol. XX, no. X, pp. 1–9, 2017.

[4] J. Finkelstein and I. C. Jeong, "Using CART for advanced prediction of asthma attacks based on telemonitoring data," in 2016 IEEE 7th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2016, 2016.

[5] M. Deliu, M. Sperrin, D. Belgrave, and A. Custovic, "Identification of Asthma Subtypes Using Clustering Methodologies," Pulm. Ther., vol. 2, no. 1, pp. 19–41, Jun. 2016.

[6] N. H. A. Suaini et al., "Children with East Asian-Born Parents Have an Increased Risk of Allergy but May Not Have More Asthma in Early Childhood," J. Allergy Clin. Immunol. Pract., vol. 7, no. 2, pp. 539-547.e3, Feb. 2019.

[7] E. M. Vermeulen et al., "Food Allergy Is an Important Risk Factor for Childhood Asthma, Irrespective of Whether It Resolves," J. Allergy Clin. Immunol. Pract., vol. 6, no. 4, pp. 1336-1341.e3, Jul. 2018.

[8] H. A. Smit et al., "Childhood asthma prediction models: A systematic review," The Lancet Respiratory Medicine, vol. 3, no. 12. Lancet Publishing Group, pp. 973–984, 2015.

[9] E. Chatzimichail, E. Paraskakis, M. Sitzimi, and A. Rigas, "An Intelligent System Approach for Asthma Prediction in Symptomatic Preschool Children," Comput. Math. Methods Med., vol. 2013, pp. 1–6, Mar. 2013.

[10] A. Q. Ansari, N. K. Gupta, and Ekata, "Automatic diagnosis of asthma using neurofuzzy system," Proc. - 4th Int. Conf. Comput. Intell. Commun. Networks, CICN 2012, pp. 819–823, 2012.

[11] B. Prasadl, P. E. S. N. Krishna Prasad, and y Sagar, "An Approach to Develop Expert Systems in Medical Diagnosis Using Machine Learning Algorithms (Asthma) and A Performance Study," Int. J. Soft Comput., vol. 2, no. 1, pp. 26–33, Mar. 2011.

[12] A. K. Patnaik and S. Prasad, "Data Mining and Its Current Research Directions," Inf. Sci. (Ny)., vol. 181, no. 7, pp. 1264–1284, 2011.

[13] A. Kusiak, B. Dixon, and S. Shah, "Predicting survival time for kidney dialysis patients: A data mining approach," Comput. Biol. Med., vol. 35, no. 4, pp. 311–327, 2005.

[14] A. AhmedK, S. Aljahdali, and S. Naimatullah Hussain, "Comparative Prediction Performance with Support Vector Machine and Random Forest Classification Techniques," Int. J. Comput. Appl., vol. 69, no. 11, pp. 12–16, Dec. 2013.

[15] Amin Ul Haq et al.,"A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, Volume 2018, Article ID 3860146, 21 pages.

[16] Amin Ul Haq et al. Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings, IEEE access, April 2019.

