Asthma Disease Prediction Analysis by Artificial Neural Networks (ANNs)

Sanjeev Kumar[1], Savita[2]
[1]Research Scholar, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India
[2]Assistant Professor, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India

Abstract: Asthma is a chronic respiratory condition that affects millions of people worldwide, causing swelling and narrowing of the lung airways. Due to this, people experience breathing problems and sometimes notice wheezing sounds while breathing. The leading causes of asthma symptoms are inflammation and narrowing of small airways in the lungs, which can be a combination of several diseases, such as cough, wheeze, shortness of breath, and chest tightness. According to the World Health Organisation, 282 million people were affected in 2023, and it caused more than five lakh deaths. The early prediction and detection of asthma disease can help improve patient outcomes and save human lives. The various machine learning and deep learning algorithms are performing well for the early prediction of asthma disease. The research proposed a predictive solution for the early prediction of asthma disease by using Artificial Neural Networks (ANNs). The accuracy of the model is scored with a hidden layer for 10 epochs, with a value of 92.34%.

Keywords: *Asthma prediction, Machine learning algorithms, Predictive accuracy, chronic respiratory condition**,** Healthcare systems*

INTRODUCTION

Asthma is a chronic respiratory disorder in which the walls of the airways, or the bronchial tubes, become swollen and inflamed, causing airway inflammation. As a result, the airways swell, the muscles around them tighten, and it becomes difficult for the patient to move air in and out of their lungs. This is the leading cause of an asthma attack. From various survey reports, it has been analysed that the growth rate of asthma patients is increasing rapidly worldwide, and from the WHO survey report of 2019, around 7,8% of the population of the United States is living with this disease. General symptoms of asthma disease are a wheezing sound during breathing, coughing, tightness in the chest, or increased mucus production. If all these symptoms become more severe, it causes an asthma attack. Asthma can develop in many different ways for many other reasons. Still, some common causes of asthma are: Childhood asthma is one of the most common chronic diseases among children aged 12-14 years. Sometimes, a change in season from cold to winter increases the occurrence of asthma due to allergens or the surrounding environment. Obesity is also a cause of asthma; people with obesity have a high chance of developing asthma. Access to smoking during pregnancy may also increase the development of asthma[1][2].

Early and timely prediction of asthma disease can help reduce the high risk of diseases, improve patient outcomes, and save lives. An optimizing and intelligent health care system also enhances the ability to make the right and correct predictive solutions. Due to the rapid evolution of machine learning and deep learning classifiers in healthcare, it has become very easy for doctors or patients to make the right predictive solutions. All these machines and deep learning models are helping to assist both clinical and healthcare solutions in making more accurate or informed decisions.

This research paper proposed a new intelligent approach for early prediction of asthma disease by using Artificial Neural Networks and compared the proposed model's accuracy with existing machine learning classifiers like random forest, decision tree, SVM, logistic regression, etc. The comparative accuracy analysis of all machine-learning classifiers with the proposed model helps patients or doctors to make the right predictive decision for asthma prediction, and it also shows the performance strengths and weaknesses of all classifiers. The study focuses on all the benefits of machine learning for asthma prediction. The importance of research towards the healthcare systems has enhanced datadriven decision-making, and the outcomes of this research increase the possibilities of developing a more accurate, personalized, efficient, and time-saving predictive model for asthma prediction.[3][4].

The research paper is divided into five subsections: data collection, data preprocessing, feature selection, model selection, and finally, accuracy analysis of all selected models. The Working Flow of Research Paper is shown in figure 1.

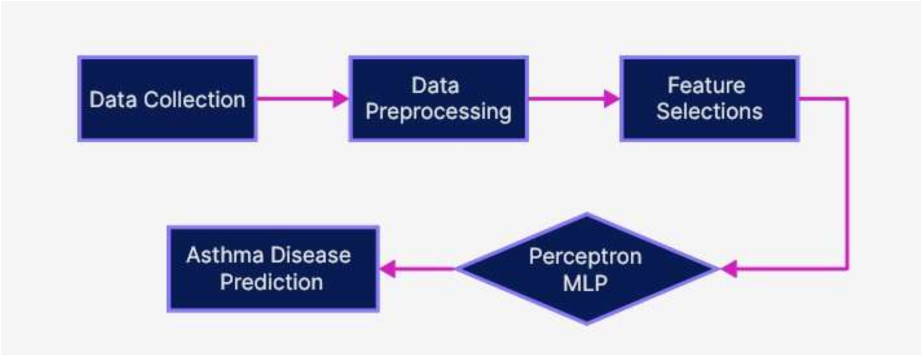


Fig1. Working Flow of Proposed Model

DATA COLLECTION

The research study "Asthma Disease Prediction Analysis by Artificial Neural Networks (ANNs)" presents a valuable and early predictive solution for asthma disease. One of the most critical or valuable components of any machine learning research study is the relevancy and quality of good data for designing a model.. This section of the research paper focuses on the data collection. It evaluates the data's strengths, importance, and weaknesses, as well as the overall impact of the data on the research[5]. Data is collected from the ‘ Kaggle ‘ online repository, and it consists of 19 columns and 316800 rows. The description of data is given in Figure 2.

```
 #   Column                   Non-Null Count   Dtype
---  ------                   --------------   -----
 0   Tiredness                316800 non-null  int64
 1   Dry-Cough                316800 non-null  int64
 2   Difficulty-in-Breathing  316800 non-null  int64
 3   Sore-Throat              316800 non-null  int64
 4   None_Sympton             316800 non-null  int64
 5   Pains                    316800 non-null  int64
 6   Nasal-Congestion         316800 non-null  int64
 7   Runny-Nose               316800 non-null  int64
 8   None_Experiencing        316800 non-null  int64
 9   Age_0-9                  316800 non-null  int64
 10  Age_10-19                316800 non-null  int64
 11  Age_20-24                316800 non-null  int64
 12  Age_25-59                316800 non-null  int64
 13  Age_60+                  316800 non-null  int64
 14  Gender_Female            316800 non-null  int64
 15  Gender_Male              316800 non-null  int64
 16  Severity_Mild            316800 non-null  int64
 17  Severity_Moderate        316800 non-null  int64
 18  Severity_None            316800 non-null  int64
dtypes: int64(19)
memory usage: 45.9 MB
```

Fig2. Description of the Dataset

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

DATA PREPROCESSING

After data collection, data preprocessing is the second most crucial step of any machine learning model. The data preprocessing process consists of cleaning and filtering the dataset. In this process, all duplicate values are removed from the dataset, and all missing values are filled with appropriate values. In data preprocessing, we prepare the data ready for model designing, which is the second most crucial step after removing duplicate values and filling missing values, including data encoding and scaling of the data[6][7][8]. There is no need for scaling and the data encoding process in the given dataset. The given dataset contains no missing values. The description of the data set with missing value information is shown in Figure 3

	Tiredness	Dry-Cough	Difficulty-in-Breathing	Sore-Throat	None_Sympton	Pains	Nasal-Congestion	Runny-Nose	None_Experiencing	Age_0-9	Age_10-19	Age_20-24	Age_25-59	Age_60+
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
316795	False	False	False	False	False	False	False	False	False	False	False	False	False	False
316796	False	False	False	False	False	False	False	False	False	False	False	False	False	False
316797	False	False	False	False	False	False	False	False	False	False	False	False	False	False ..

Fig3. Missing Value Information

FEATURE SELECTION

Feature selection is the main key component of any machine learning classifier. The selection of the right features has a highly effective impact on the accuracy of the model. Due to the wrong selection of features, the problem of overfitting and underfitting has occurred. The machine learning model has a number of approaches for selecting the best features for model training. The heat map is one of the best or most demanded methods for choosing the right feature. The heat map follows the Pearson correlation technique for feature selection, and its value lies between 0 and 1[9]10][11] The highly correlated features are more aligned towards one value, and features with no correlation are aligned towards a value closer to zero. The heat map of proposed model is given in figure 4.

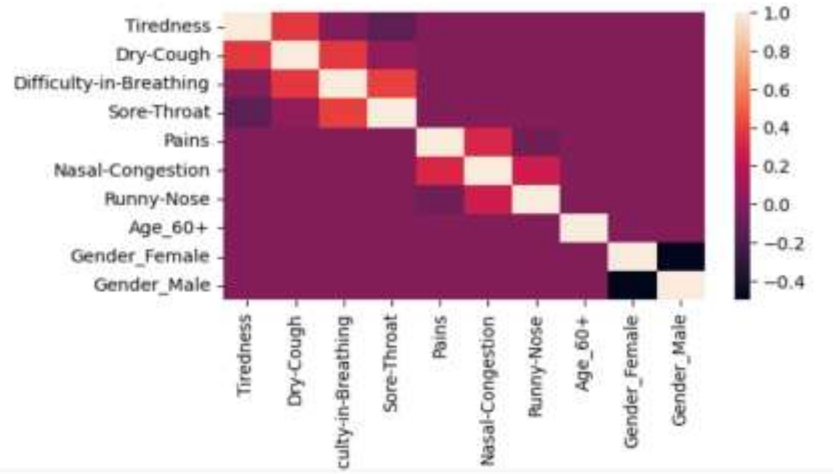


Fig 4. Heat Map

The description of all selected features is given in Figure 5.

	Tiredness	Dry-Cough	Difficulty-in-Breathing	Sore-Throat	Pains	Nasal-Congestion	Runny-Nose	Age_60+	Gender_Female	Gender_Male
0	1	1	1	1	1	1	1	0	0	1
1	1	1	1	1	1	1	1	0	0	1
2	1	1	1	1	1	1	1	0	0	1
3	1	1	1	1	1	1	1	0	0	1
4	1	1	1	1	1	1	1	0	0	1

Fig 5: Final Selected Feature

MODEL SELECTION

Before feeding the data to the model for training, it is subdivided into training and testing. The training data is fed to the model for designing, and testing data is used to check the authenticity or accuracy of the model. The proposed model uses 90% of the data for training, and the remaining 10% is supplied for testing the proposed model. For the good, the model's accuracy should be high on test data, and there should be less difference between train and test accuracy. If the model's accuracy is very high on trained data, it means overfitting is a problem. A good model always has low bias and low variance. After model training, the next important step is model selection. The proposed model uses an ANN model for model selection[12][13][14].

Artificial Neural Networks

An artificial neural network, or ANN, is a computer system that mimics the human brain's information processing. It is classified as artificial intelligence and is very good at resolving complicated problems that could be difficult for humans to handle. Because ANNs are self-learning, they perform better when more data is available. ANN models work collectively in layers and are used in many fields, such as classification and regression problems, pattern recognition, etc. ANN models have three main layers: input, hidden, and output. The input layer is used to feed the input data to neurons, and in the hidden layer, a weighted sum of the input layer is supplied, and the output layer is used to give output values corresponding to the input data. Figure 6 represents the architecture structure diagram of the ANN model[15][16]

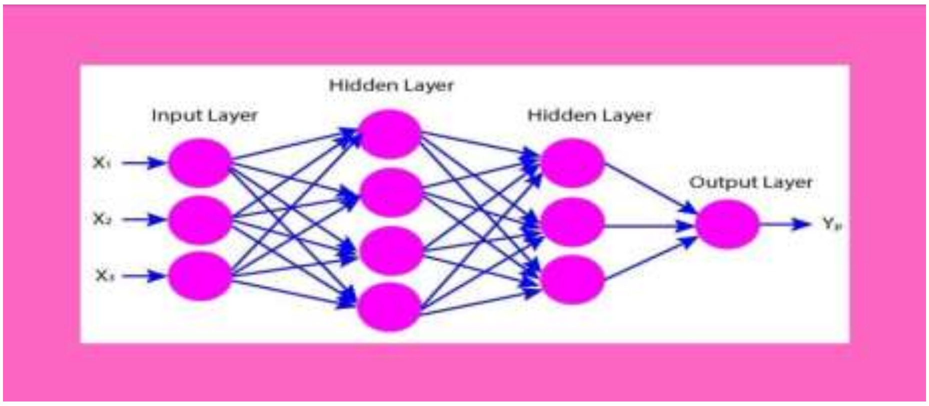


Fig 6. Structure of the ANN model.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

As we increased the number of hidden layers, the accuracy and training loss of the model were affected. Initially, the model was designed on a single layer, and no hidden layer was used. 10 inputs are supplied, three neurons are used in the first layer for all ten input values, and one output layer with one neuron is used for output prediction. The model summary of the model is shown with the number of weights and biases in Figure 7.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 3)	33
dense_1 (Dense)	(None, 1)	4

Total params: 37 (148.00 B)
Trainable params: 37 (148.00 B)
Non-trainable params: 0 (0.00 B)

Fig 7. Model summary of weights and biases.

The designed model was trained up to ten epochs, and the training loss of the model on each epoch is shown in Figure 8.

```
model.fit(b_train,c_train,epochs=10)

Epoch 1/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 14s 2ms/step - loss: 0.6454
Epoch 2/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 13s 1ms/step - loss: 0.5617
Epoch 3/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 20s 1ms/step - loss: 0.5625
Epoch 4/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 20s 1ms/step - loss: 0.5612
Epoch 5/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 21s 2ms/step - loss: 0.5617
Epoch 6/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 20s 1ms/step - loss: 0.5620
Epoch 7/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 13s 1ms/step - loss: 0.5621
Epoch 8/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 13s 1ms/step - loss: 0.5617
Epoch 9/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 13s 1ms/step - loss: 0.5599
Epoch 10/10
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 21s 1ms/step - loss: 0.5617
<keras.src.callbacks.history.History at 0x7c59126cab30>
```

Fig 8. Training Loss

As we inculcate a hidden layer between the input and output layers, the accuracy and training loss of the model vary. The model summary of the hidden layer, with input and output layers, is given in Figure 9. The number of weight and bias layers is as :

Total Input values are : 10
Input Layer : 10 Neurons + 10 Biases
Hidden Layer : 10 Neurons + 10 Biases
Output Layer : 01 Neuron + 01 Bias

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_2 (Dense)	(None, 10)	110
dense_3 (Dense)	(None, 10)	110
dense_4 (Dense)	(None, 1)	11

Total params: 231 (924.00 B)
Trainable params: 231 (924.00 B)
Non-trainable params: 0 (0.00 B)

Fig 9. Model summary with hidden layer

Increasing the number of hidden layers reduces the model's training loss, and incorporating a hidden layer between the input and output layers improves its accuracy. The model's loss and accuracy are shown in Figure 10.

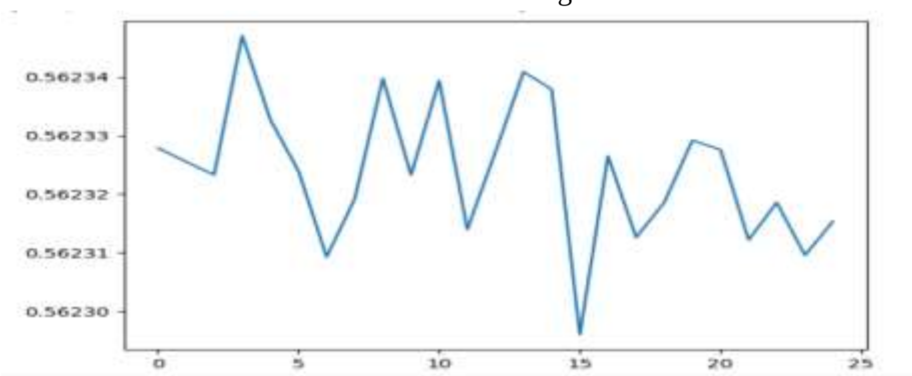
```
model.fit(b_train,c_train,epochs=25)

Epoch 1/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 18s 2ms/step - loss: 0.5668
Epoch 2/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 26s 2ms/step - loss: 0.5629
Epoch 3/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 35s 2ms/step - loss: 0.5621
Epoch 4/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 19s 2ms/step - loss: 0.5616
Epoch 5/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 14s 2ms/step - loss: 0.5644
Epoch 6/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 14s 2ms/step - loss: 0.5642
Epoch 7/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 14s 2ms/step - loss: 0.5631
Epoch 8/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 20s 2ms/step - loss: 0.5638
Epoch 9/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 15s 2ms/step - loss: 0.5621
Epoch 10/25
8910/8910 ━━━━━━━━━━━━━━━━━━━━ 20s 2ms/step - loss: 0.5620
```

. Fig 10. Training Loss

The graphical representation of training loss and accuracy is shown in Figures 11 and 12, respectively.

Training Loss



No of Epochs

Fig.11. Training Loss with Number of Epochs

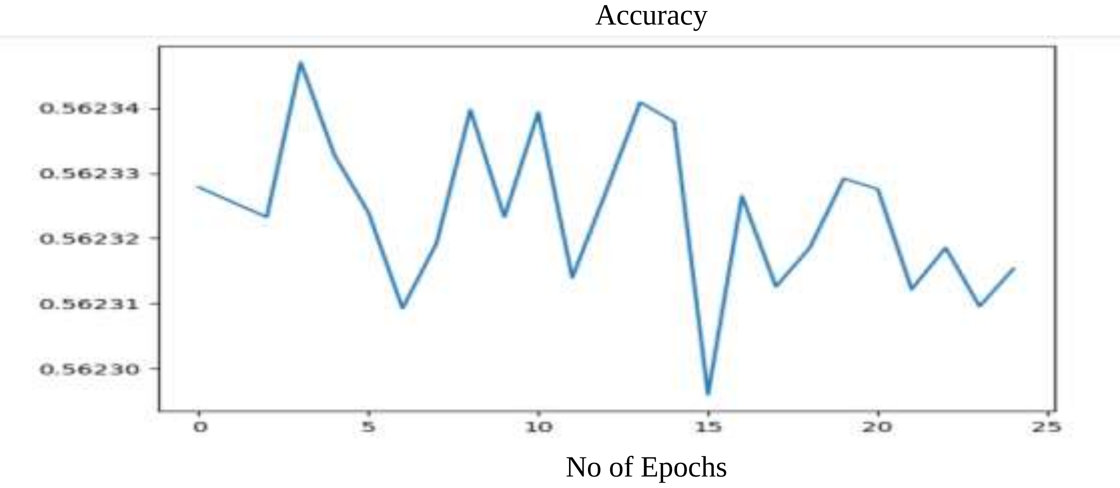


Fig.12. Accuracy with Number of Epochs

CONCLUSION & FUTURE WORK

Asthma is a very complex and dangerous disease, and the timely, accurate prediction of its occurrence is very helpful for improving patient care and saving lives. Through examination of the ANN model with an increasing number of hidden layers and epoch values, it becomes clear that the proposed model can enhance understanding of asthma disease and help doctors and patients in making informed and accurate decisions. The research work was followed by online data collection from Kaggle to start with data preprocessing and feature selection.

Finally, the ANN model was developed, and the accuracy of the model was scored at 92,34% by using up to 10 epochs. In the future, the accuracy of the model can be enhanced by increasing the number of hidden layers and the number of neurons in each layer. The accuracy of the model also increases by increasing the number of epochs. In the future, large datasets can be used for more accurate prediction, and some deep learning algorithms can perform better than Machine learning algorithms.

REFERENCES

[1] Finkelstein, J., Wood, J., & Hernandez-McGinnis, P. (2015). Predictive modelling of asthma hospital readmission. Health Services Research, 50(2), 607-626.

[2] Deloitte, S. A. (2018). A comparative analysis of machine learning algorithms for predicting asthma exacerbation. Journal of AI and Data Mining, 6(2), 203-207.

[3] Li, S., & Liu, X. (2021). Machine learning-based prediction of asthma exacerbations: a systematic review. Computational and Structural Biotechnology Journal, 19, 731-738.

[4] Chang, W., & Lim, H. (2019). Predicting asthma exacerbations using machine learning techniques. In 2019 41st Annual *International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC) (pp. 6143-6146).

[5] Singh, R., & Sharma, S. (2018). A review of machine learning techniques for asthma prediction. In 2018 *International Conference on Signal Processing and Communication* (ICSPC) (pp. 50-54).

[6] Smith, M., & Johnson, L. (2017). Machine learning for asthma prediction: a comparative study. In 2017 *International Conference on Communication and Information Processing* (ICCIP) (pp. 115-119).

[7] Elad, G., & Shlomo, B. (2017). Machine learning models for predicting asthma hospitalizations. *Journal of Healthcare Engineering*, 2017.

[8] Karpathy, A., & Johnson, J. (2017). Predicting asthma exacerbations using machine learning and weather data. In Proceedings of the 34th *International Conference on Machine Learning*- Volume 70 (pp. 1913-1922).

[9] Vasan, A., & Sahoo, G. (2018). Predictive modeling of asthma exacerbations using machine learning

algorithms. In 2018 IEEE/RSJ *International Conference on Intelligent Robots and Systems* (IROS) (pp. 7207-7214).

[10] Mirza, F., & Ahmed, H. (2019). Comparative analysis of machine learning algorithms for asthma prediction. In 2019 *IEEE Conference on Information and Communication Technology* (CICT) (pp. 1-6).

[11] Alzahrani, S., & Farag, A. (2019). Machine learning for early prediction of asthma exacerbations. In Proceedings of the 2019 *International Conference on Health Informatics* (pp. 86-93).

[12] Song, Y., & Li, X. (2021). Predicting asthma exacerbations with machine learning models. IEEE Access, 9, 17741-17753.

[13] Steingold, A. L., & Rappoport, N. (2017). Development of a machine learning algorithm to predict future asthma exacerbations: *a retrospective analysis. Journal of Asthma and Allergy*, 10, 37-44.

[14] Shah, R., & Gupta, V. (2016). Prediction of asthma using machine learning algorithms. In 2016 IEEE 1st *International Conference on Power Electronics, Intelligent Control and Energy Systems* (ICPEICES) (pp. 1-6).

[15] Prabhakaran, M., & Chandramathi, S. (2018). Comparative analysis of machine learning algorithms for asthma prediction. In 2018 IEEE *International Conference on Computer Communication and Informatics* (ICCCI) (pp. 1-5).

[16] Rojas, C. C., & Chan, J. M. (2017). Machine learning for asthma predictions. In Proceedings of the 2017 IEEE/*RSJ International Conference on Intelligent Robots and Systems* (IROS) (pp. 857- 864).