JACKA
Journal of Advanced Computer Knowledge and Algorithms
Journal of Advanced Computer Knowledge and Algorithms
Vol. 2, No. 2, April, 2025, pp. 30-34
Journal homepage: http://ojs.unimal.ac.id/jacka

Classification of Asthma Diseases Using Machine Learning Models at Arun Hospital

Khalis Al Muqarrabin[1*], Fadlisyah[2], T. Mirzal Safari[3]

[1,2] Universitas Malikussaleh, Indonesia
[3] Rumah Sakit Arun Lhokseumawe, Indonesia

[*]Corresponding Author Email: khalis.210170290@mhs.unimal.ac.id

Received: 9 March 2025
Revised: 20 March 2025
Accepted: 31 March 2025
Available online: 1 April 2025

Keywords:
Classification, K-Nearest Neighbor, Model Evaluation, Information System

ABSTRACT

Asthma is one of the chronic diseases that significantly affects the quality of life of patients. This study aims to classify asthma disease based on patient data from Arun Hospital using the K-Nearest Neighbor (KNN) algorithm. The dataset consists of 330 patient data with attributes such as allergy, itchy throat, and shortness of breath. The data went through preprocessing, transformation, and normalization stages. The KNN model was tested with a value of k = 3, resulting in three main classifications: Mild Asthma, Moderate Asthma, and Severe Asthma. The evaluation results showed a high accuracy rate, with an average of more than 90%. In addition, the model was implemented in the form of a system that visualizes the dataset, KNN analysis, and model evaluation. These findings demonstrate the potential of the KNN algorithm to provide accurate predictions and support the diagnosis of asthma disease effectively.

1. INTRODUCTION

Asthma is a chronic disease that affects the respiratory tract and often affects the quality of life of sufferers. It is characterized by chronic inflammation of the respiratory tract that causes hyperresponsiveness of the bronchi to various triggers, such as allergens, air pollution, and physical activity [1]. Symptoms such as shortness of breath, coughing, and wheezing are common, especially at night or early in the morning, and can limit daily activities if not properly managed [2]. The disease not only impacts individuals but has also become a significant public health problem, with an increasing global prevalence, especially in developing countries such as Indonesia [3].

Arun Hospital, as one of the major healthcare centers in Aceh, treats hundreds of asthma patients annually. Patients' clinical data includes symptoms, examination results, and treatment history, which are part of electronic medical records (EMR). EMRs have the potential to support data analysis and evidence-based decision-making, but often pose challenges in their management due to the volume and complexity of the data [4]. This requires the application of advanced analytics technologies to unearth clinically relevant patterns and insights.

The K-Nearest Neighbor (KNN) algorithm is one of the effective machine learning approaches for analyzing medical data. KNN is an instance-based algorithm that classifies data based on proximity between data in a multidimensional feature space, using Euclidean distance or other metrics as measurements [5]. The advantage of KNN lies in its simplicity and its ability to work with datasets that are not too large or that have been well normalized [6]. In the context of health, KNN has been widely used for classification and prediction of diseases, such as diabetes, cancer, and asthma.

This study uses the KNN algorithm to classify asthma patients into three categories: Mild Asthma, Moderate Asthma, and Severe Asthma. This classification process is carried out based on symptom attributes such as allergies, itchy throat, and shortness of breath, which are important indicators in asthma diagnosis according to the Global Initiative for Asthma (GINA) guidelines [7]. The patient dataset was preprocessed, including missing data handling and normalization, to ensure consistency of data format. This process is important to improve the accuracy of the classification model.

The classification results are expected to provide medical personnel with better insights in understanding patient conditions, speeding up diagnosis and improving treatment effectiveness. Previous research has shown that machine learning-based algorithms, including KNN, can improve medical decision-making, especially in handling complex clinical data [8].

By implementing the KNN algorithm, this study aims to demonstrate the advantages of this approach in supporting medical decision-making for asthma. This approach also has the potential to become a foundation for further development in the application of machine learning technology for complex clinical data management, so that it can provide wider benefits in the health sector.

2. RESEARCH METHODS

This study uses the K-Nearest Neighbor (KNN) algorithm to classify asthma patient data from Arun Hospital. The dataset used consists of 330 patient data with symptom attributes such as allergies, itchy throat, shortness of breath, and wheezing sounds. Each data was classified into one of the categories: Mild Asthma, Moderate Asthma, or Severe Asthma.

The research steps were systematically designed to ensure the accuracy and validity of the results. The research procedure involved data collection, data preprocessing, algorithm implementation, and result evaluation. For ease of understanding, the research flowchart is organized in a clear systematic manner.

2.1 Research Flow

The research flow in this study is made so that the steps taken by the author in this design do not deviate from the subject matter and are easier to understand, so the sequence of steps will be made systematically so that it can be used as a clear and easy guideline to solve existing problems. The sequence of steps that will be made in this study can be seen as follows.

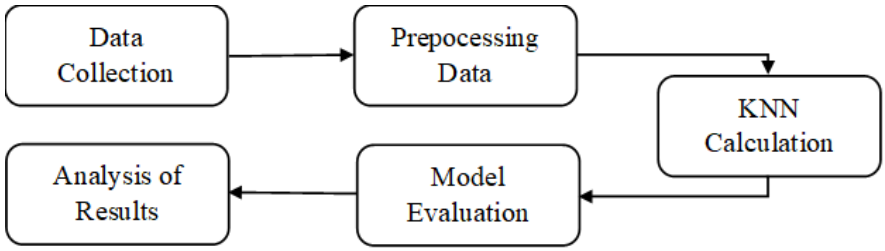

Figure 1. Research flow

Based on Figure 1, it shows the research flow with the main stages: Data Collection, Data Preprocessing, KNN Calculation, Model Evaluation, and Results Analysis [9]. Starting with the collection of relevant data for research, followed by preprocessing to ensure data quality through processes such as cleaning or normalization. After that, the processed data is used in KNN Calculation to perform classification or prediction. Next, model evaluation is performed using metrics such as accuracy, precision, recall, or F1-score to measure model performance. Finally, the evaluation results are analyzed to draw conclusions that support the research objectives. This flow reflects a systematic approach in building and evaluating KNN-based models.

The system scheme of the research flow carried out in this study, starting from data collection, preprocessing, classification with the KNN algorithm, to model evaluation, is as follows:

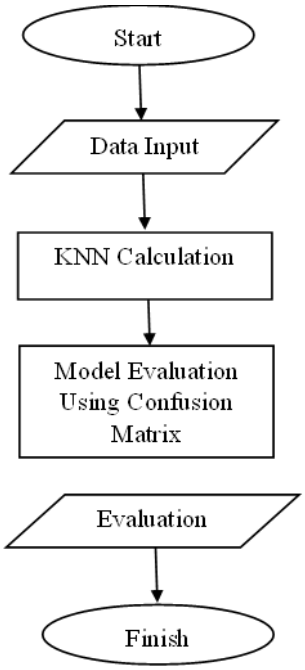

Figure 2. System scheme

Based on Figure 2, the step taken in the Data Input process is to input the dataset that will be classified using the K-Nearest Neighbors (KNN) algorithm. This dataset must be prepared in advance, such as through normalization or division into training data and test data. In the KNN Calculation section, the dataset is tested by calculating the distance between the test data and the training data using a specific distance method, such as Euclidean Distance, to determine the k nearest neighbors. Next, class predictions are made based on the majority of these neighbors. In the Model Evaluation Using Confusion Matrix process, the prediction results are compared with the original labels to calculate the evaluation metric. This process aims to measure the success rate of the model in performing classification.

3. RESULT AND DISCUSSION

At this stage, the research results obtained through the application of the K-Nearest Neighbor (KNN) algorithm will be explained in detail. This research aims to classify asthma patient data into three main categories: Mild Asthma, Moderate Asthma, and Severe Asthma. The classification process starts with processing the dataset through data normalization, followed by Euclidean distance calculation to determine the nearest neighbor of each test data.

The classification results are then compared with the actual class to evaluate the accuracy of the model. Based on the evaluation results, the KNN algorithm shows excellent performance with an accuracy rate of more than 90%. The following is a discussion of the main steps in data analysis:

Table 1. The Dataset

NUMBER	NAME_PATIENT	GENDER	AGE	ALERGY	...	DIAGNOSA
1	Z****	Woman	10	No	...	MILD ASTHMA
2	S*******	Woman	7	Food	...	MILD ASTHMA
3	A******	Men	24	No	...	MILD ASTHMA
4	L********	Woman	21	Food	...	MILD ASTHMA
5	N**********	Woman	8	Dust	...	MILD ASTHMA
...	...	...	...	...	...	...
330	D*********	Woman	26	No	...	MILD ASTHMA

Table 1 above records patient data consisting of number, patient name, gender, age, allergy type, and diagnosis. All patients in this data were diagnosed with "Mild Asthma". The age of the patients varied from children to adults, with some having allergies such as food or dust, while others had no allergies.

Table 2. Data Transformation

NAME_PATIENT	ALERGY	THROAT_ITCHY	SNEEZE	NASAL_CONGESTION	...	DIAGNOSA
Z****	0	0	0	1	...	MILD ASTHMA
S*******	1	1	0	0	...	MILD ASTHMA
A******	0	1	0	0	...	MILD ASTHMA
L********	1	0	0	1	...	MILD ASTHMA
N**********	2	1	1	0	...	MILD ASTHMA
...	...	...	...	...	...	...
D*********	0	1	0	1	...	MILD ASTHMA

Table 2 contains patient data with attributes such as allergies, symptoms of itchy throat, sneezing, and nasal congestion. Each symptom is assigned a binary value (0 or 1), where 0 indicates the symptom is absent and 1 indicates the symptom is present. All these data are associated with a uniform diagnosis of "Mild Asthma." This kind of data transformation is often used for statistical analysis or machine learning algorithms.

Table 3. Data Normalization

NAME_PATIENT	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	DIAGNOSA
Z****	0	0	0	0	0	0	0	0	0	0	MILD ASTHMA
S*******	0	0	0	0	0	0	0	0	0	0	MILD ASTHMA
A******	0	0	0	0	0	0	0	0	0	0	MILD ASTHMA
L********	0	0	0	0	0	0	0	0	0	0	MILD ASTHMA
N**********	0	0	0	0	0	0	0	0	0	0	MILD ASTHMA
...	...	...	...	...	...	...	...	...	...	...	...
D*********	0	1	0	1	0	0	0	0	0	0	MILD ASTHMA

Table 3 above explains that after the normalization process using the Min-Max Scaler method, the values of variables X1 to X10 have been adjusted to the range [0,1]. This normalization ensures that each variable has a uniform scale and does not dominate other variables in the model calculation. These variables reflect patient-perceived symptoms, such as X1 to X10, which indicate the intensity or presence of certain symptoms. Diagnosis variables are retained as category indicators and do not undergo normalization as they are labels or targets to be predicted.

3.1 Distance Calculation

Calculation of the Euclidean distance between test data (testing) and training data (training) using the K-Nearest Neighbor (KNN) algorithm. Each row in the table shows the distance value between one test data and all training data in the dataset. Calculations are based on normalized attributes (X1 to X10), using the Euclidean Distance formula.

Table 4. Distance calculation and sorted from the smallest

Number	Dinstance	Class	Classification
1	0	MILD ASTHMA	MILD ASTHMA
	0	MILD ASTHMA	
	0	MILD ASTHMA	
2	0	MILD ASTHMA	MILD ASTHMA
	0.333333	MILD ASTHMA	
	0.333333	MILD ASTHMA	
3	0	MILD ASTHMA	MILD ASTHMA
	0	MILD ASTHMA	
	0	MILD ASTHMA	
4	0.333333	MILD ASTHMA	MILD ASTHMA
	0.333333	MILD ASTHMA	
	0.333333	MILD ASTHMA	
5	0	MILD ASTHMA	MILD ASTHMA
	0.333333	MILD ASTHMA	
	0.333333	MILD ASTHMA	
...	...	...	...
66	0	MILD ASTHMA	MILD ASTHMA
	0	MILD ASTHMA	
	0.333333	MILD ASTHMA	

This distance result is used to identify the k nearest neighbors (k=3 in this case). After the distance is calculated, the values are sorted from smallest to largest, and the three training data with the smallest distance (nearest neighbors) are used to determine the class (diagnosis) of the test data. The class is determined based on the majority of the nearest neighbors.

Table 5. Actual and Predicted Results

NAME_PATIENT	ACTUAL	PREDICT
Z***	MILD ASTHMA	MILD ASTHMA
K****	MILD ASTHMA	MILD ASTHMA
M******	MILD ASTHMA	MILD ASTHMA
E****	MILD ASTHMA	MILD ASTHMA
L*********	MILD ASTHMA	MILD ASTHMA
...	...	...
D*********	MILD ASTHMA	MILD ASTHMA

The actual results show the diagnosis of the disease based on the original data, while the predicted results show the diagnosis generated from the model based on the Euclidean distance calculation and the majority of the nearest neighbors (k=3). This table is used to evaluate the performance of the model, such as calculating accuracy, precision, and recall, by comparing the predicted results with the actual results.

After obtaining classification results that illustrate the relationship between actual and predicted values generated by the K-Nearest Neighbor (KNN) model, the next step is to conduct an in-depth analysis of the model's performance through data visualization. This visualization aims to identify the level of accuracy, precision, and sensitivity of the model in classifying data in each category. The following is a discussion of visualizations used to comprehensively evaluate model performance.

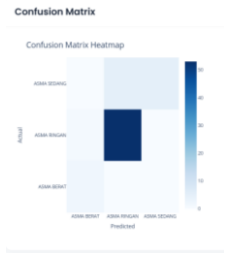

Figure 3. *Confusion Matrix*

This matrix is visualized in the form of a heatmap, with the vertical axis representing the actual values (ground truth) and the horizontal axis representing the model predicted values. Darker colors indicate a higher frequency of correct or incorrect predictions. Overall, the KNN model showed good ability to identify certain categories, such as mild asthma, with sufficient accuracy. However, the model showed misclassification in other categories, particularly in moderate and severe asthma severity, indicating the need for improvement in distinguishing between these categories. Next, the analysis was carried out with reference to the results obtained from the various metrics, which will be elaborated further

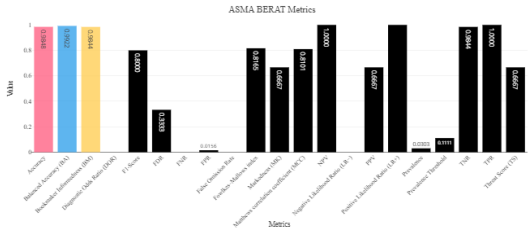

Figure 4. Severe Asthma Accuracy Rate Chart

The graph shows various performance metrics to evaluate the severe asthma diagnosis model. Key metrics such as accuracy, precision, recall, and F1-score have high values, indicating that the model is quite reliable in identifying patients with severe asthma. Low False Positive Rate and False Negative Rate values indicate little misclassification, while high specificity reflects the model's ability to recognize patients who do not have the condition. Overall, this graph shows that the model performs well to effectively detect and classify severe asthma.

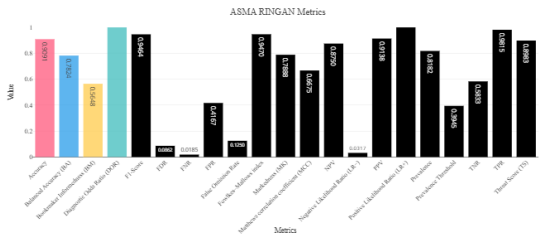

Figure 5. Mild Asthma Accuracy Rate Chart

The model performance graph for mild asthma classification shows that the accuracy of the model is at a high level, reflecting the overall ability to provide correct predictions. High precision and recall values indicate that the model can detect positive cases well while minimizing false positive errors. A balanced F1-score indicates harmonization between precision and sensitivity. In addition, the low False Positive Rate (FPR) and False Negative Rate (FNR) confirmed the low misclassification rate, while the high specificity strengthened the model's ability to recognize mild non-asthmatic patients. These results show that the model is effective and reliable in detecting mild asthma with minimal risk of error.

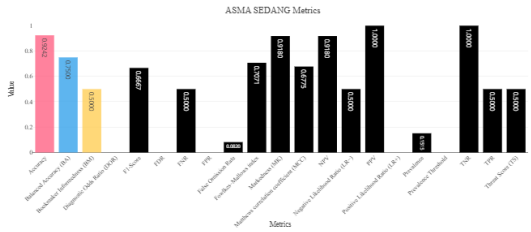

Figure 6. Accuracy Rate Chart of Moderate Asthma

The graph shows a high level of accuracy, indicating the reliability of the model in providing correct predictions. High precision values indicate the model's ability to minimize false positive errors, while adequate recall indicates good coverage of positive case detection. A balanced F1-score confirms the balance between precision and sensitivity. The low False Positive Rate (FPR) and False Negative Rate (FNR) indicate minimal misclassification, and the high specificity value supports the model's ability to identify moderate non-asthmatic patients. These results reflect that the model has an effective and consistent performance for detecting moderate asthma with a low error rate.

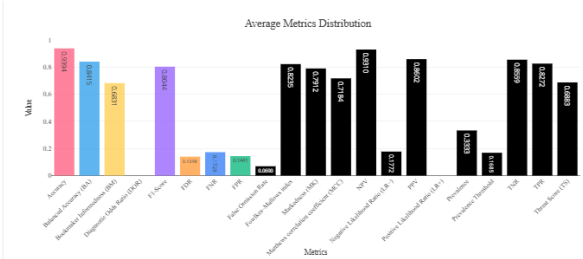

Figure 7. Graph of Average Accuracy Rate of 3 Categories

The graph above illustrates the average accuracy distribution of the three main categories of model evaluation: methods or algorithms, traditional evaluation metrics, and average performance across datasets based on specific areas. In the first category, the average accuracy of various methods such as Random Forest and Decision Tree are compared, showing the difference in performance between algorithms. The second category evaluates traditional metrics such as precision, recall, and F1-score, which are key indicators to understand the balance between positive and negative prediction errors. Meanwhile, the third category focuses on the average performance of the model in a particular dataset or specific area, providing a broader perspective of performance. These results demonstrate the importance of using diverse metrics to holistically evaluate the performance of machine learning models.

4. CONCLUSIONS

This study shows that the K-Nearest Neighbor (KNN) algorithm performs well in classifying asthma patient data into three severity categories: mild, moderate, and severe. Based on the evaluation results, the model achieved a high level of accuracy with balanced precision, recall, and F1-score values reflecting the optimal balance between precision and

sensitivity. Visualization of the confusion matrix and performance graph confirmed the effectiveness of the model in detecting mild asthma with minimal error, although there were some misclassifications in the moderate and severe asthma categories that require further improvement. These results emphasize the importance of KNN parameter optimization and exploration of other algorithms such as Random Forest and Decision Tree to improve accuracy and inter-category differentiation, to support more accurate and reliable medical decision-making.

REFERENCES

[1] R. W. S. Putra, E. Windartik, and R. Merbawani, "Pengaruh Pemberian Tindakan Pursed Lips Breathing Terhadap Perubahan Respiration Rate Pada Pasien Asma Di Rs Kamar Medika Mojokerto," PhD Thesis, Perpustakaan Bina Sehat PPNI, 2023.

[2] S. K. Syafiq Aufa, S. K. Asmaul Husna, and S. K. Syahrizal, "Penatalaksanaan Asma Bronkial pada Anak Melalui Pendekatan Kedokteran Keluarga: Sebuah Laporan Kasus," *Journal of Medical Science*, vol. 4, no. 2, pp. 130–143, 2023.

[3] H. Setiadi, S. KM, and S. K. M. Fifi Dwijayanti, "Pentingnya kesehatan masyarakat, edukasi dan pemberdayaan perempuan untuk mengurangi stunting di negara berkembang," in *Jurnal Seminar Nasional*, 2020, pp. 16–25.

[4] R. Sri Rusmini, "Pengembangan Model Dengan Proses Manajemen Risiko Pembentukan Struktur Di Keperawatan (Studi Rumah Sakit Di Semarang)," PhD Thesis, Universitas Karya Husada, 2022.

[5] R. Rina, P. M. Hasan, N. Ayu, and R. A. Saputra, "KLASIFIKASI KERINGANAN UKT MAHASISWA UHO MENGGUNAKAN K-NEAREST NEIGHBOR (KNN)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 11939–11945, 2024.

[6] E. Safitri, R. H. R. Sibarani, Y. S. Sidabutar, and D. Kiswanto, "Klasifikasi Penyakit Daun Anggur Berbasis Citra Menggunakan Metode K-Nearest Neighbors (KNN)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 12633–12642, 2024.

[7] R. Hiday, "Asthma Guideline Updates", Accessed: Jan. 08, 2025.

[8] M. A. M. Mustofa, H. N. Wahiid, B. M. Islami, A. Ristyawan, and E. Daniati, "Penggunaan Algoritma KNN dalam Deteksi Awal Kanker Paru-Paru Menggunakan Data Medis," in *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 2024, pp. 485–493.

[9] R. Nurhidayat and K. E. Dewi, "Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek," *Komputa: Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 91–100, 2023.