JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 1, Desember 2025, hlm. 37 – 44 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i1.4192
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

COMPARISON OF DECISION TREE AND RANDOM FOREST ALGORITHMS FOR ASTHMA

Wisriani Lase[1*], Robet[1], Hendri[1]
[1]Informatics Engineering, STMIK TIME
email: *wisrianilase@gmail.com

Abstract: Asthma is a chronic respiratory disease that affects millions of people worldwide, making early detection crucial to prevent complications. This study aims to compare the performance of the Decision Tree and Random Forest algorithms in classifying asthma based on clinical symptom data. The data were processed through feature selection and model training stages, then evaluated using accuracy, precision, recall, and F1-score.The experimental analysis revealed that the Random Forest algorithm surpassed the Decision Tree in all metrics, achieving 95.19% accuracy, 90.43% precision, 95.00% recall, and 93.00% F1-score. In contrast, the Decision Tree obtained 89.14% accuracy, 90.60% precision, 88.70% recall, and 89.70% F1-score. These results suggest that Random Forest is more robust and dependable, especially in managing complex and imbalanced medical datasets.

Keywords: asthma detection; decision tree; random forest; machine learning.

Abstrak: Asma merupakan penyakit pernapasan kronis yang memengaruhi jutaan orang di seluruh dunia sehingga deteksi dini sangat penting untuk mencegah komplikasi. Penelitian ini bertujuan membandingkan kinerja algoritma Decision Tree dan Random Forest dalam mengklasifikasikan asma berdasarkan data gejala klinis. Data diproses melalui tahapan seleksi fitur dan pelatihan model, kemudian dievaluasi menggunakan akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Random Forest memberikan performa terbaik dengan akurasi 90.43%, presisi 95.00%, recall 95.00%, dan F1-score 93.00%. Sebaliknya, Decision Tree memperoleh akurasi 89.14%, presisi 90.60%, recall 88.70%, dan F1-score 89.70%. Hasil ini menunjukkan bahwa Random Forest lebih kuat dan dapat diandalkan, terutama dalam mengelola kumpulan data medis yang kompleks dan tidak seimbang.

Kata kunci: deteksi asma; decision tree; random forest; pembelajaran mesin.

INTRODUCTION

Asthma is a chronic inflammatory disease of the lower respiratory tract, characterized by symptoms such as shortness of breath, coughing, and chest tightness. According to the Global Burden of Disease (GBD) report in 2021, there were approximately 260.48 million asthma patients worldwide, with 436,193 deaths recorded in the same year[1].

The Decision Tree and Random Forest algorithms are two widely used classification techniques because they

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 1, Desember 2025, hlm. 37 – 44 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i1.4192
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

can handle data with various types of attributes and provide relatively accurate results [2]. Decision Tree builds a model in the form of a tree structure based on data attributes, while Random Forest is an ensemble method that combines many decision trees so that the prediction results are more stable and resistant to changes[3].

Previous research has shown that the Random Forest algorithm has high accuracy in general respiratory disease classification[4]. Meanwhile, other studies have developed expert systems in mobile applications using the Decision Tree algorithm for asthma diagnosis but did not make a direct comparison with other methods such as Random Forest[5].

Based on these findings, there is a research gap regarding the lack of studies explicitly comparing the performance of Decision Tree and Random Forest in detecting asthma based on patient symptom data [6].
Various studies have shown that Random Forests are superior in terms of accuracy, generalization, and robustness compared to Decision Trees. Several studies have even reported that this algorithm can achieve up to 99% accuracy in identifying asthma risk factors [7]. arious studies have designed asthma detection systems using physiological parameters such as heart rate and oxygen saturation [8].

Other studies have emphasized the importance of public health education in asthma prevention, supporting the relevance of implementing intelligent algorithms for early detection of this disease[9]. Research also shows that although Decision Tree has a high recall value, Random Forest is more stable and accurate[10]. Several studies have confirmed the superiority of Random Forest in processing biomarker data and diagnosing adult asthma[11][12]. Furthermore, this algorithm has also proven effective in handling data imbalance and producing reliable predictions on medical data [13][14]. Recent studies have further extended these approaches by integrating intelligent systems for early asthma prediction in pediatric and symptomatic populations . These systems utilize hybrid models combining clinical symptoms, environmental exposure data, and physiological indicators to enhance prediction accuracy and support real-time monitoring. [15]. Other studies have shown that machine learning algorithms such as Decision Tree and Random Forest can predict asthma in adults with relatively high accuracy, with Random Forest providing more stable performance [16]. Machine learning algorithms have also been shown to improve the accuracy of asthma diagnosis using various approaches, with XGBoost performing the best, although Random Forest remains superior to Decision Tree[17].

METHOD

This study employs two popular classification algorithms in machine learning, namely Decision Tree and Random Forest. Random Forest is also known to be superior in terms of stability and accuracy compared to Decision Tree, as demonstrated by previous studies. [18] [19][20].

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 1, Desember 2025, hlm. 37 – 44 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i1.4192
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

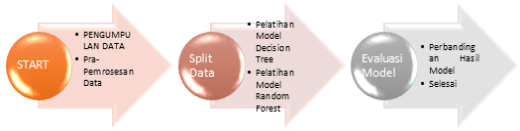

Image 1. Research Stages

Data Collection and Data Preprocessing

The dataset was obtained from Kaggle and contains 2,392 rows and 29 columns, encompassing various clinical variables related to asthma symptoms. Example attributes include PatientID, Age, Gender, BMI, Smoking, Physical Activity, and others:

Table 1. Dataset Features

Name	Description
PatientID	Identitas Pasien
Age	Usia pasien
Gender	Jenis kelamin
Ethnicity	Kode pasien
Education	Pendidikan pasien
Bmi	Indeks masa tubuh pasien
Smoking	Status Merokok
Physical Activity	Aktivitas pasien
Diet quality	Kualitas makan

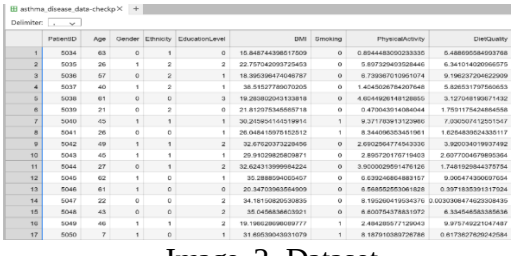

Image 2. Dataset

Before training, the dataset was loaded using pandas, and then split into 80% training data and 20% testing data using the train_test_split method with a specific random_state. This preprocessing step is essential to allow the model to learn patterns optimally and produce more accurate predictions.

Decision Tree Model Training

The Decision Tree algorithm works by constructing a tree structure based on input attributes to perform classification using criteria such as the Gini Index or Entropy. In this study, the Gini Index criterion was used, calculated using the following formula:

$$Gini(D) = 1 - \sum_{i=1}^{c} P^2 \quad (1)$$

Description: P^2 represents the proportion of data in the i-th class, while c indicates the total number of classes in the dataset.

The Decision Tree uses criterion="gini" as the method to measure the quality of data splits, whereas max_depth=None means the tree depth is not limited, allowing the tree to grow according to the complexity of the data.
Decision Tree hyperparameters: criterion='gini', max_depth=None.

Random Forest Model Training

In the Random Forest algorithm, the model is built by combining multiple decision trees using the bagging technique. Each tree is trained on randomly selected data samples and features, resulting in more stable predictions that are resistant to overfitting. The final prediction is determined through a majority voting mechanism across all trees, calculated as follows:

$$\acute{y} = mode\ (\{ht(x)\}T \quad (2)$$

Description: ý represents the final prediction, ht(x) is the prediction of the t-th tree for input x, T denotes the total

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 1, Desember 2025, hlm. 37 – 44 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i1.4192
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

number of trees in the ensemble, and mode refers to the class value that appears most frequently among all tree predictions. Besides the Gini Index, the algorithm can also use other splitting criteria such as Entropy, particularly when the model focuses on measuring information uncertainty at each node. The formula for Entropy used in the data-splitting process is as follows:

$$Entropy(S) = -\sum_{i=1}^{n} \text{p}i \quad \log 2\ pi \quad (3)$$

Where p_i represents the proportion of data for the i-th class within a node. Entropy equals zero if all data in the node belong to a single class and reaches its maximum value when the class distribution in the node is uniform.
The main parameters used in this study include the number of trees ($n\_estimators = 100$), the splitting criterion using Gini impurity, and $max_depth = None$to allow the tree to grow according to the structure of the data. The model was trained using the training data and evaluated on the testing data to assess classification performance.
In this study, the model was evaluated using four main metrics: accuracy, precision, recall, and F1-score, each calculated based on standard performance evaluation formulas.

Accuracy
This indicates that Random Forest is superior in producing overall accurate predictions.

$$\text{Accuracy} : \frac{TP+TN}{FP+FN+TP+TN} \quad (4)$$

Precision
Measures the model's ability to correctly predict positive cases:

$$\text{Precision} : \frac{TP}{TP+FP} \quad (5)$$

Recall
Indicates the model's ability to identify all positive instances:

$$Recall : \frac{TP}{TP+FN} \quad (6)$$

F1-Score
Represents the harmonic mean of precision and recall:

$$\text{F1-Score} : 2 * \frac{Precison * recall}{Precision + recall} \quad (7)$$

RESULTS AND DISCUSSION

Model training was performed using hyperparameter tuning, the process of finding the best combination of hyperparameter values to improve prediction performance. Hyperparameters are set before training and govern how the model learns from the data. The following are the hyperparameters used in each algorithma:

```
# 6. Tuning hyperparameter Decision Tree
param_dt = {
    'criterion': ['gini', 'entropy'],
    'max_depth': [5, 10, 15, None],
    'min_samples_split': [2, 5, 10]
}

grid_dt = GridSearchCV(
    DecisionTreeClassifier(),
    param_grid=param_dt,
    cv=5,
    scoring='accuracy'
)

# 7. Latih model
grid_dt.fit(X_train, y_train)
```

Image 3. Hyperparameter Tuning Decision Tree

```
param_grid_list = [
    {'max_features': ['sqrt', 'log2'], 'n_estimators': [10, 100, 1000, 2000]},
    {'max_features': ['sqrt'], 'n_estimators': [500, 1000, 1500]},
    {'max_features': ['sqrt'], 'n_estimators': [200, 300, 500, 700]},
    {'max_features': ['sqrt'], 'n_estimators': [250, 300, 350, 400]},
    {'max_features': ['sqrt'], 'n_estimators': [275, 300, 325]},
    {'max_features': ['sqrt'], 'n_estimators': [290, 300, 310]}
]

# 6.
results = []

# 7. Loop setiap konfigurasi parameter
for i, param_grid in enumerate(param_grid_list, 1):
    grid = GridSearchCV(RandomForestClassifier(random_state=42), param_grid, cv=7, scoring='accuracy')
    grid.fit(X_train, y_train)

    best_score = grid.best_score_
    best_params = grid.best_params_

    results.append({
        "No": i,
        "Tuning Parameter": str(param_grid),
        "Score (7-fold Cross Validation)": round(best_score, 6),
        "Best Parameter": str(best_params)
```

Image 4. Hyperparameter Tuning Random Forest

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 1, Desember 2025, hlm. 37 – 44 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i1.4192
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

The Grid Search process is used to find the best hyperparameter combination in the Random Forest Classifier algorithm. The tuning focus is on the max_features and n_estimators parameters, exploring variations in the number of trees (n_estimators) and the number of features used in node splitting (max_features). Evaluation is performed using 7-fold cross-validation to ensure consistent results.

More accurate and unbiased data distribution.

This method allows researchers to determine the optimal number of trees and the best feature selection strategy to produce a Random Forest model with maximum accuracy.

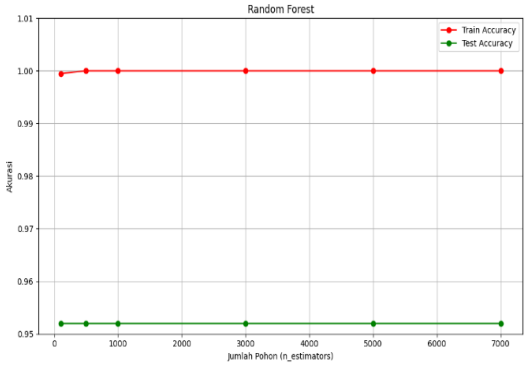

Image 5. Random Forest Graph

The figure shows the effect of the number of decision trees (n_estimators) on the accuracy of the Random Forest model. The red line shows the accuracy of the training data, while the green line shows the accuracy of the test data. The accuracy of the training data is close to 1.00 (100%) and stable, indicating the model is able to learn the data patterns well. The accuracy of the test data is also stable at around 0.95 (95%), indicating the model is not experiencing overfitting. These results indicate that adding too many trees does not significantly improve accuracy, only increasing computational time. Therefore, the choice of the n_estimators value is based on a balance between accuracy and computational efficiency.

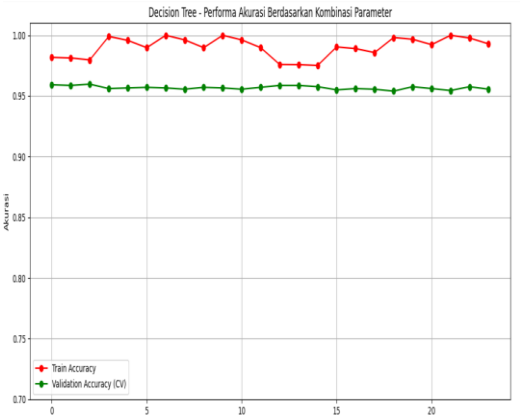

Image 6. Decision Tree Graph

The Decision Tree algorithm achieved very high training accuracy (up to 1.00) but slightly lower validation accuracy (around 0.95), indicating overfitting. This shows the model fits the training data too closely, reducing its ability to generalize. Although validation accuracy remains relatively high, performance does not improve with parameter changes. Thus, optimization methods like pruning or ensemble approaches (e.g., Random Forest) are recommended to reduce overfitting and enhance generalization.

Tabel 2. Classifier Accuracy Results

Evaluation Metrics	Decision Tree	Random Forest
Accuracy	0.8914	0.9519
Precision	0.9060	0.9043
Recall	0.8870	0.95
F1-Score	0.8970	0.93

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 1, Desember 2025, hlm. 37 – 44 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i1.4192
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Tabel 3. Hyperparameter Tuning

Algorithm	Training Accuracy	Test Accuracy
Decision Tree	0.95	0.93
Random Forest	0.94	0.95

Based on the table, it can be seen that the Random Forest algorithm has an advantage in all evaluation metrics, indicating that this model is more accurate, better able to recognize positive cases, and more consistent in predictions.

Image 7. Comparison of Model Result

CONCLUSION

This study concludes that the Random Forest algorithm outperforms the Decision Tree in detecting asthma based on patients' clinical symptom data. The ensemble method effectively handles complex medical data and reduces over-fitting, making it more reliable for early asthma
detection. These findings strengthen the role of machine learning, particularly Random Forest, in developing accurate and efficient medical decision support systems.

BIBLIOGRAPHY

[1] Z. Mao *et al.*, "Global, regional, and national burden of asthma from 1990 to 2021: A systematic analysis of the global burden of disease study 2021," *Chinese Med. J. Pulm. Crit. Care Med.*, vol. 3, no. 1, pp. 50–59, 2025, doi: 10.1016/j.pccm.2025.02.005.

[2] Z. Jeddi, I. Gryech, M. Ghogho, M. E. L. Hammoumi, and C. Mahraoui, "Machine learning for predicting the risk for childhood asthma using prenatal, perinatal, postnatal and environmental factors," *Healthc.*, vol. 9, no. 11, 2021, doi: 10.3390/healthcare9111464.

[3] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024, doi: 10.58496/bjml/2024/007.

[4] D. Kurniawan, M. Wahyudi, L. Pujiastuti, and S. Sumanto, "Deteksi dan Prediksi Cerdas Penyakit Paru-Paru dengan Algoritma Random Fores," *Indones. J. Comput. Sci.*, vol. 3, no. 1, pp. 51–56, 2024, doi: 10.31294/ijcs.v3i1.6071.

[5] H. Abtahi, S. Amini, M. Gholamzadeh, and M. A. Gharabaghi, "Development and evaluation of a mobile-based asthma clinical decision support system to enhance evidence-based patient management in primary care," *Informatics Med. Unlocked*, vol. 37, no. January, p. 101168, 2023, doi: 10.1016/j.imu.2023.101168.

[6] A. Ehtesham, S. Kumar, A. Singh, and T. T. Khoei, "Pediatric

Asthma Detection with Googles HeAR Model: An AI-Driven Respiratory Sound Classifier," 2025, [Online]. Available: http://arxiv.org/abs/2504.20124

[7] P. Kotlia, J. Pant, and M. C. Lohani, "Identifying Asthma Risk Factors and Developing Predictive Models for Early Intervention Using Machine Learning," *Biomed. Pharmacol. J.*, vol. 18, no. March, pp. 295–314, 2025, doi: 10.13005/bpj/3089.

[8] S. Indriani, E. D. Setyoningsih, D. Titisari, and A. J. Wuryanto, "Design Of Asthma Detection Devices Through Heart Rate and Oxygen Saturation," *Indones. J. Electron. Electromed. Eng. Med. informatics*, vol. 2, no. 3, pp. 143–149, 2020, doi: 10.35882/ijeeemi.v2i3.6.

[9] Hapipah, "Edukasi Peningkatan Pengetahuan Tentang Penyakit Asma Berdasarkan data dari World Health saluran napas yang biasanya ditandai penyakit asma sangat diperlukan . pengetahuan tentang asma , penyebab ," vol. 1, no. 2, pp. 13–18, 2023.

[10] M. F. Bağcı *et al.*, "Detection and prediction of real-world severe asthma phenotypes by application of machine learning to electronic health records," *J. Allergy Clin. Immunol. Glob.*, vol. 4, no. 3, pp. 1–8, 2025, doi: 10.1016/j.jacig.2025.100473.

[11] C. Chen *et al.*, "Genetic biomarker prediction based on gender disparity in asthma throughout machine learning," *Front. Med.*, vol. 11, no. September, pp. 1–10, 2024, doi: 10.3389/fmed.2024.1397746.

[12] S. Alkobaisi, M. F. Safdar, P. Pałka, and N. A. Abu Ali, "Artificial Intelligence Algorithms in Asthma Management: A Review of Data Engineering, Predictive Models, and Future Implications," *Appl. Sci.*, vol. 15, no. 7, pp. 1–23, 2025, doi: 10.3390/app15073609.

[13] E. Sagheb *et al.*, "AI model for predicting asthma prognosis in children," *J. Allergy Clin. Immunol. Glob.*, vol. 4, no. 2, p. 100429, 2025, doi: 10.1016/j.jacig.2025.100429.

[14] K. Tomita *et al.*, "Construction of a Diagnostic Algorithm for Diagnosis of Adult Asthma Using Machine Learning with Random Forest and XGBoost," *Diagnostics*, vol. 13, no. 19, 2023, doi: 10.3390/diagnostics13193069.

[15] Zahab, M., Hussain, M., & Parwati, L. S., "Prediction of Asthma Disease Using Machine-Learning Algorithm," Eng. Proc., vol. 107, no. 1, 2025, doi:10.3390/engproc2025107115.

[16] J. R. N. A. Gunawardana, S. D. Viswakula, R. P. Rannan-Eliya, and N. Wijemunige, "Machine learning approaches for asthma disease prediction among adults in Sri Lanka," *Health Informatics J.*, vol. 30, no. 3, 2024, doi: 10.1177/14604582241283968.

[17] H. Joo, D. Lee, S. H. Lee, Y. K. Kim, and C. K. Rhee, "Increasing the accuracy of the asthma diagnosis using an operational definition for asthma and a machine learning method," *BMC Pulm. Med.*, vol. 23, no. 1, pp. 1–9, 2023, doi: 10.1186/s12890-023-02479-4.

[18] D. D. Li, T. Chen, Y. L. Ling, Y. Jiang, and Q. G. Li, "A

Methylation Diagnostic Model Based on Random Forests and Neural Networks for Asthma Identification," *Comput. Math. Methods Med.*, vol. 2022, 2022, doi: 10.1155/2022/2679050.

[19] J. Lam Shin Cheung, N. Paolucci, C. Price, J. Sykes, and S. Gupta, "A system uptake analysis and GUIDES checklist evaluation of the Electronic Asthma Management System: A point-of-care computerized clinical decision support system," *J. Am. Med. Informatics Assoc.*, vol. 27, no. 5, pp. 726–737, 2020, doi: 10.1093/jamia/ocaa019.

[20] P. Zhou, C. xia Xiang, and J. feng Wei, "The clinical significance of spondin 2 eccentric expression in peripheral blood mononuclear cells in bronchial asthma," *J. Clin. Lab. Anal.*, vol. 35, no. 6, pp. 1–9, 2021, doi: 10.1002/jcla.23764.