İSTANBUL UNIVERSITY PRESS

Acta Infologica

Acta Infologica,
ACIN 2025, 9 (1): 223–252

https://doi.org/10.26650/acin.1676351
Submitted: 16.04.2025
Revision Requested: 20.05.2025
Last Revision Received: 31.05.2025
Accepted: 03.06.2025
Published Online 05.06.2025

Research Article

Open Access

Voices from the Lungs: An Innovative Approach to Asthma Diagnosis using Machine Learning

Murat Gezer [1], Mehmet Atilla Uysal [2], Neval Alagöz [2], Can Ortak [3] & Elif Yelda Niksarlıoğlu [2]

[1] İstanbul University, Faculty of Science, Department of Computer Science, İstanbul, Türkiye
[2] University of Health Sciences Türkiye, Yedikule Hospital for Chest Diseases and Thoracic Surgery, Department of Chest Diseases, İstanbul, Türkiye
[3] Medcase Corp, İstanbul, Türkiye

Abstract Asthma is one of the most common chronic respiratory diseases worldwide, and early and accurate diagnosis is critical for effective clinical management. In this study, we evaluated the diagnostic potential of machine learning models based on voice analysis as a non-invasive approach for asthma diagnosis. Using audio samples containing seven different phonetic units, the performances of 13 different machine-learning algorithms were comprehensively analyzed. The StandardScaler and SMOTE techniques were applied in the data preprocessing stage, and a 5-fold cross-validation methodology was adopted to evaluate the models. Accuracy, F1-score, sensitivity, precision, specificity, and area under the curve (AUC) metrics were used for performance evaluation. The results demonstrate that ensemble learning approaches, particularly the stacking ensemble model, exhibit superior discriminative capacity for all phonetic units. Individual models, such as neural networks and support vector machines, also produced remarkable results, whereas simpler models were limited in terms of capturing complex patterns in audio data. This study demonstrated the promising diagnostic potential of voice analysis-based ensemble learning approaches for asthma diagnosis; however, it emphasizes the need for an optimal balance between sensitivity and specificity in clinical applications.

Keywords Asthma diagnosis · Sound analysis · Machine learning · Ensemble models · Acoustic feature extraction

Citation: Gezer, M., Uysal, M. A., Alagöz, N., Ortak, C. & Niksarlıoğlu, E. Y. (2025). Voices from the Lungs: An innovative approach to asthma diagnosis using machine learning. *Acta Infologica*, *9*(1), 223-252. https://doi.org/10.26650/acin.1676351

Corresponding author: Murat Gezer murat.gezer@istanbul.edu.tr

Introduction

Asthma is a prevalent chronic respiratory condition that affects millions of people globally, leading to severe morbidity and mortality if not effectively managed. As recognized as a priority area for intervention by the World Health Organization's Global Action Plan for the Prevention and Control of Noncommunicable Diseases and the United Nations 2030 Agenda for Sustainable Development, asthma is a significant global health concern (World Health Organization, 2024). This heterogeneous respiratory disease is characterized by chronic inflammation and airway narrowing (bronchoconstriction) (GINA 2023). This pathological process manifests as clinical symptoms, such as recurrent wheezing, dyspnea, chest tightness, and coughing episodes (Barnes, 2008). Structural changes, including increased mucus production, smooth muscle hypertrophy, and epithelial damage, also contribute to asthma pathophysiology (Holgate et al., 2015). These symptoms are typically exacerbated at night or in the early morning and can be triggered by exposure to various factors, such as allergens, infections, exercise, and environmental stimuli (Lambrecht & Hammad, 2015).

The burden of asthma on the global health economy is considerable. The WHO estimates that 262 million individuals were living with asthma in 2019 (World Health Organization, 2024), and 455,000 deaths were due to the disease in the same year (Ministry of Health, General Directorate of Public Health, 2024). Most deaths occur in low- and lower-middle-income countries, emphasizing the critical need for improved diagnosis and management strategies in these regions (World Health Organization, 2024).

The data obtained from Türkiye also show a similar picture. In a 2017 study, 6.9% of adults aged 15 years and older were diagnosed with asthma, and this rate was higher in women (8.7%) than in men (5.0%) (Republic of Türkiye Ministry of Health, General Directorate of Public Health, 2024). Furthermore, data from the Turkish Statistical Institute (TUIK) for 2022 show that respiratory system diseases are the third leading cause of death, with asthma accounting for 0.3% of these deaths (Ministry of Health, General Directorate of Public Health, 2024).

In addition to statistics, asthma profoundly impacts an individual's quality of life, affecting their ability to participate in daily activities and significantly altering their general wellbeing. Speech is an area considerably affected by asthma. Speech production is a complex physiological process involving coordinated interactions among multiple systems. The air expelled from the lungs vibrates the vocal cords in the larynx to produce sounds. This sound is shaped and modulated by articulators in the oral cavity, including the tongue, teeth, and palate, and the resonance spaces of the sinuses and nasal cavity (Sezer & Akıl, 2020). Airway narrowing and inflammation caused by asthma disrupt this complex process (Fesci & Görgülü, 2005). Consequently, individuals with asthma experience difficulty breathing, resulting in changes in voice quality, decreased speech intelligibility, and voice fatigue.

Pulmonary function tests, such as spirometry and peak flowmetry, have long been used as standard methods in clinical practice for the diagnosis and follow-up of asthma. These methods are applied in clinical settings under the supervision of specialized healthcare professionals and are insufficient for continuously monitoring changes in patients' daily lives. These limitations in the follow-up of the disease highlight the need for more comprehensive and dynamic approaches to asthma management.

Technological developments in recent years have paved the way for innovative asthma prediction and monitoring approaches in the healthcare field (Saygılı, 2019). These new methods enable real-time or out-of-clinic monitoring of patients' clinical observations. Research has shown that the respiratory acoustic characteristics of patients with asthma differ from those of healthy individuals (Schreur et al., 1994). These differences are manifested especially in the form of high-frequency wheezing and low-frequency crackles (Reichert et al., 2008; Schreur et al., 1994). In a study conducted by Malmberg et al. in 1994, spectral analyses showed an increase in power in certain frequency ranges of breath sounds in patients with asthma (Malmberg et al., 1994).

The unique contribution of this study is that it provides an innovative approach for asthma diagnosis by analyzing speech signals. While several studies have explored asthma diagnosis and risk analysis using speech signals (e.g., Alam et al., 2022; Chen et al., 2025; Yadav et al., 2018; Yadav et al., 2020), a comprehensive investigation comparable in scale and methodology to the present research, specifically, one utilizing data collected from Turkish-speaking individuals diagnosed in a hospital setting according to GINA criteria, and evaluating the performance of 13 different machine learning algorithms across seven distinct phonetic units, had not been reported in the literature at the time this study was conducted. This research aims to go beyond current diagnostic methods that are only applied in standardized clinical settings and develop a methodology that offers the potential for non-invasive and continuous monitoring of patients in their daily lives. The aim of our study was to determine whether a patient had asthma using acoustic analysis of voice recordings and to identify machine learning models that can achieve the highest accuracy.

The developed acoustic analysis method detects possible asthma symptoms by analyzing various parameters, such as frequency, amplitude, and spectral features (graphical representation of the power distribution of sound waves) in sound signals. The proposed approach is less invasive than conventional methods and offers continuous monitoring. For example, analyzing respiratory sounds recorded during the night may be useful for detecting asthma symptoms that occur during sleep.

These innovative sound analysis methods can potentially lead to a paradigm shift in asthma management. Continuous and non-invasive monitoring may enable early detection and prevention of asthma attacks. It can also improve patients' quality of life and reduce healthcare costs by helping create personalized treatment plans.

However, further research and validation studies are required to fully integrate these technologies into clinical practice. Issues such as data security, algorithm reliability, and cost-effectiveness should be carefully addressed before their widespread use. In addition, comprehensive training programs should be developed to enable healthcare professionals and patients to effectively use these new technologies.

Machine learning techniques have been widely used in the field of bioinformatics to exploit disease hallmarks in the gene, proteomic, and metabolic expression of samples to build accurate prediction models. In conclusion, the asthma prediction model developed in this study can help individuals who cannot immediately visit a diagnostic doctor due to a lack of infrastructure and other health resources. Furthermore, this model can advance the prediction of other diseases affecting the pulmonary system.

Literature Review

The use of machine learning algorithms for asthma diagnosis has recently attracted considerable attention recently. In particular, the effectiveness of artificial intelligence has been observed in the analysis of lung imaging and the interpretation of pulmonary function tests (Kaplan et al., 2021). A comprehensive literature review revealed that various researchers have examined different machine-learning models to improve the clinical diagnosis of asthma (Bolat, 2021). These studies have generally been conducted on large-scale patient datasets and have yielded promising results for the automatic diagnosis of asthma.

Yadav et al. (2018) investigated the classification of asthmatic patients and healthy subjects using sustained phonations (/A:/, /i:/, /u:/, /eI/, /oU/, /s/, /z/) and nonspeech sounds (cough, wheeze) from 47 asthmatic and 48 healthy participants. Using MFCC statistics and Support Vector Machines (SVM), wheeze achieved the highest classification accuracy of 90.25%, whereas among sustained vowels, /i:/ achieved 80.79% accuracy (Yadav et al., 2018).

In a subsequent study, Yadav et al. (2020) employed INTERSPEECH 2013 Computational Paralinguistics Challenge baseline (ISCB) acoustic features from a similar dataset for more detailed analysis. Using ISCB features, /oU/ phonemes achieved 75.4% accuracy (18.28% improvement over baseline), while exhale achieved the highest classification accuracy of 77.8%. The loudness and MFCC feature groups have been identified as the most important contributors to asthma classification (Yadav et al. 2020).

Alam et al. (2022) developed machine-learning approaches to predict lung function from voice recordings of patients with asthma. In total, 323 voice recordings were collected from 26 patients with asthma who underwent bronchoprovocation tests. A threshold-based mechanism was designed to separate speech and breathing segments, and 23 features were extracted. The random forest regression model showed the best performance, with the lowest root mean square error (RMSE = 10.86) for predicting FEV1%. The Random Forest algorithm achieved 85% accuracy in binary classification to predict abnormal lung function (Alam et al., 2022).

Sterling et al. (2014) developed an Automated System for Asthma Monitoring (ADAM) that operates on a mobile platform. The system uses an iOS application to detect cough sounds using an external microphone. Using the hidden Markov model (HMM)-based Viterbi decoder and MFCC features, the system achieved 63% sensitivity with three false positives per hour (Sterling et al., 2014). This study demonstrated the potential of mobile technology for the objective monitoring of asthma symptoms, particularly among adolescents.

Deep learning methods, particularly convolutional neural networks (CNNs) and long short-term memory (LSTM) architectures, are prominent approaches for the classification of lung sounds. Aykanat et al. (2020) obtained an 86% accuracy rate by combining Mel-Frequency Log Energy (MFCC) features with the (GB) algorithm (Aykanat et al., 2020). Similarly, Zhang et al. (2024) developed a hybrid CNN-LSTM model using spectrogram images as inputs and achieved 99.01% accuracy and 99.13% sensitivity (Zhang et al., 2024). These findings confirm that time-frequency representations (e.g., spectrograms) provide an effective framework for the automatic analysis of pathological respiratory sounds (wheezes and crackles).

Beyond the success of spectrogram-based analyses, Petmezas et al. (2022) achieved an accuracy of 76.39% using a hybrid CNN-LSTM architecture (Petmezas et al., 2022). Research emphasizes that visual representations of spectrograms in the frequency-time domain are critical sources of features for both traditional and deep learning models.

To provide practical solutions for asthma screening in resource-constrained areas, Gunawardana et al. (2024) evaluated 13 different machine-learning algorithms using data collected in the Sri Lanka Health and Aging Study (SLHAS). This study found that the hybrid model of Logistic Regression and LightGBM performed best with a sensitivity of 79.85% and AUC value of 0.9062. These results indicate that a low-cost model can be developed for community-based health screening and clinical settings can be developed (Gunawardana et al., 2024).

Topaz et al. (2022) conducted a study using audio-recorded patient-primary care provider encounters to evaluate the level of shared decision-making (SDM) and predict inhaled corticosteroid adherence in primary care settings. Speech-to-text algorithms were used to automatically transcribe 80 audio-recorded encounters (ROUGE F-score = 0.9), and machine learning algorithms (Naive Bayes, Support Vector Machines, Decision Tree) were applied. The highest F-score achieved was 0.88 for SDM evaluation (Naive Bayes) and medication adherence prediction (Support Vector Machines) (Topaz et al., 2022). This study represents pioneering work demonstrating that speech data can be used to predict patient adherence to asthma treatment.

Rivas-Navarrete et al. (2025) developed an edge-computing-based system for detecting chronic respiratory diseases (COPD and asthma) by analyzing cough and breath sounds. The system operates on Raspberry Pi devices and smartphones by using MFCC and Chromagram features. Trained on a dataset collected from 86 participants (53 with respiratory conditions and 33 healthy), the system achieved 90% sensitivity, 93.55% specificity, and 91.75% balanced accuracy in detecting chronic respiratory diseases (Rivas-Navarrete et al., 2025). This study represents a significant step forward in the development of low-cost and portable diagnostic tools suitable for resource-limited areas.

Chen et al. (2025) developed an AI-based system that uses voice analysis to predict the risk of asthma. Using 1500 speech samples from the Saarbrucken Voice Database (high-pitch, normal-pitch, and low-pitch recitations of phonemes [i, a, u]), the Long-Term Average Spectrum (LTAS) and Single-Frequency Filtering Cepstral Coefficients (SFCCs) were extracted as features. Seven machine-learning algorithms (Decision Tree, Random Forest, Gradient Boosting, SVM, ANN, CNN, LSTM) were employed in this study. The decision Tree, CNN, and LSTM models achieved average accuracy values greater than 80%, and the decision tree model demonstrated the best accuracy for high pitch phonemes (accuracy: 98.66%) (Chen et al., 2025).

Mayr et al. (2025) analyzed the speech characteristics of patients with COPD during and after exacerbation. Using machine learning (SVM) on speech data from 50 patients with COPD, they achieved 84% accuracy in classifying patient status, which was significantly higher than the 65% accuracy obtained using CAT and BORG scores alone. After exacerbation, patients exhibit reduced voice breathiness, more stable loudness and phonation, fewer pauses, and more regular reading rhythms (Mayr et al., 2025).

Material and Methods

This study applied a comprehensive methodology for asthma diagnosis based on the analysis of audio signals. The methodological approach comprises three main stages: data collection, feature extraction, and machine learning model development.

Data Collection

This study was conducted in a single center with a cross-sectional design between January 2024 and July 2024 at the University of Health Sciences Istanbul Yedikule Chest Diseases and Thoracic Surgery Training and Research Hospital. Thus, 344 participants, comprising 284 patients diagnosed with asthma by a pulmonologist according to the GINA 2023 asthma guidelines and 60 healthy individuals without any history of chronic disease (Figure 1), were included in the study and were volunteers aged between 18 and 64 years (Figure 2). The study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Clinical Research Ethics Committee of Istanbul Yedikule Chest Diseases and Thoracic Surgery Training and Research Hospital (Approval date: December 14, 2023, Approval number 2023-433).

Figure 1

Distribution of study subjects.

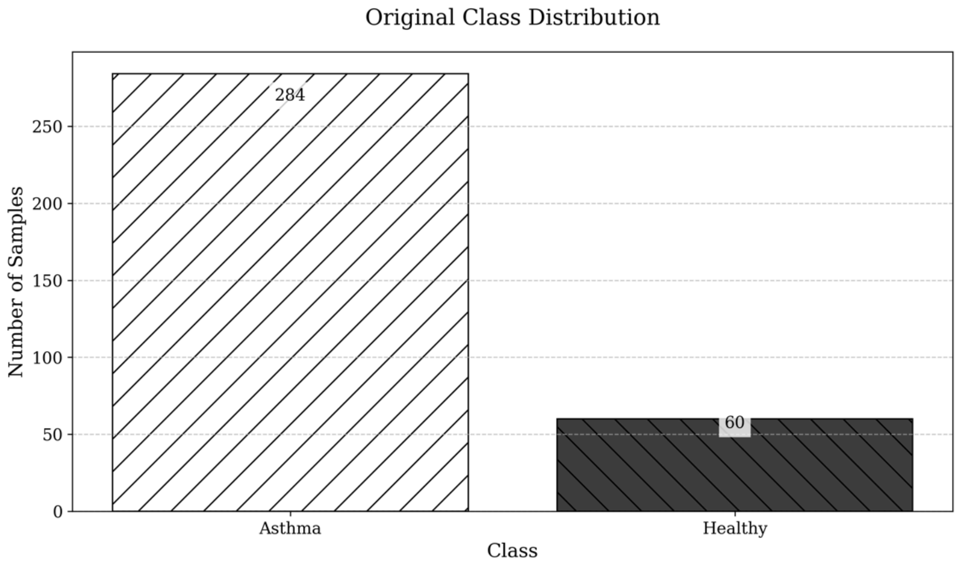
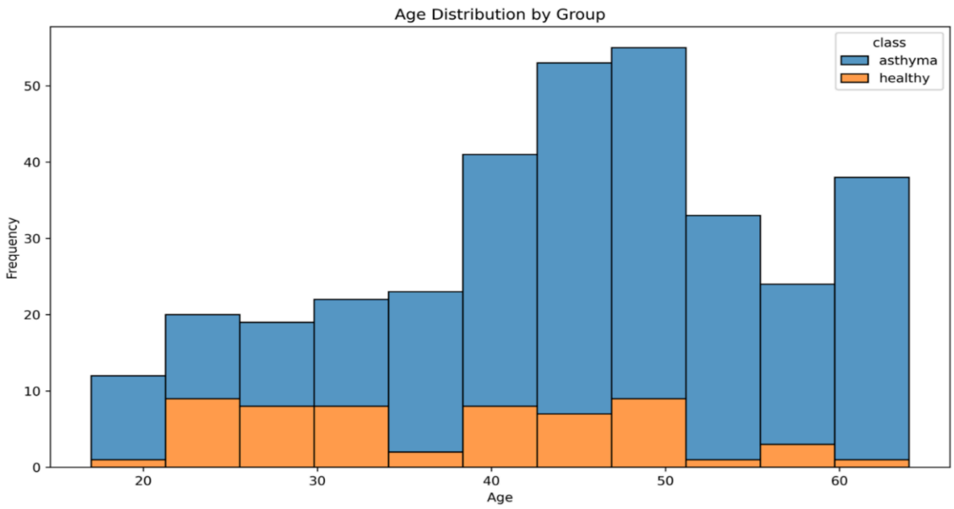


Figure 2

Distribution of Subject Age Groups



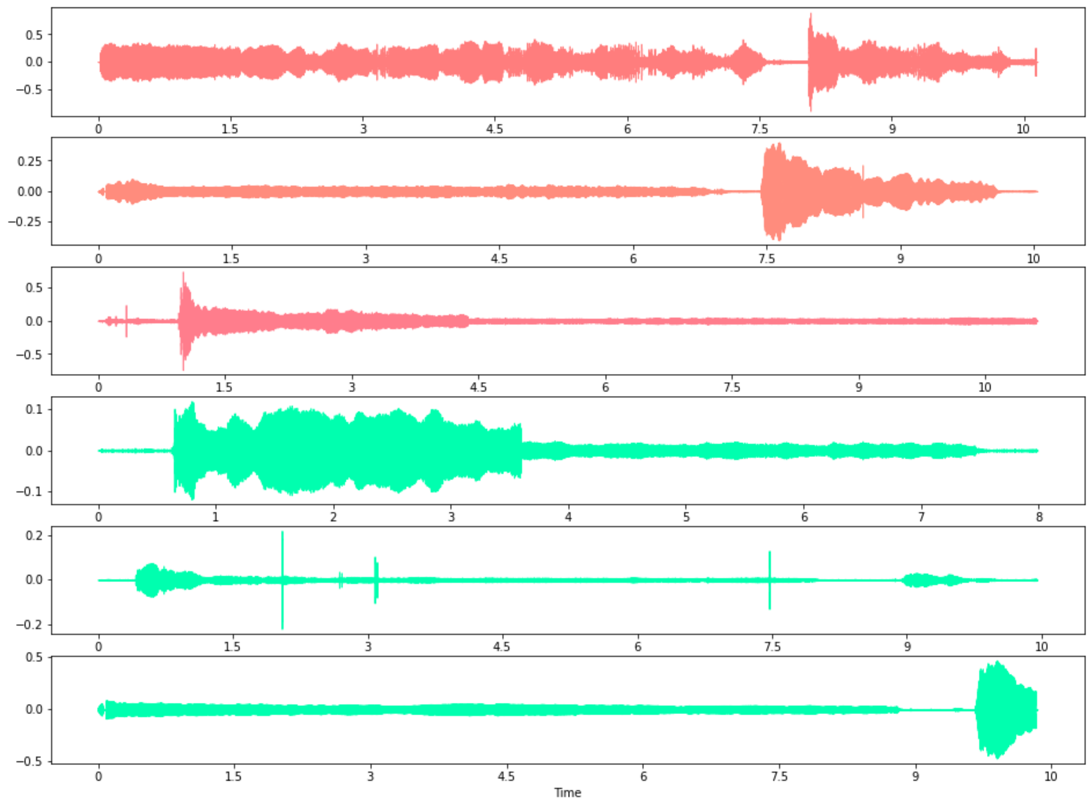
A physician detailed the purpose and processes of the study to all participants. Participants were matched by age and sex, and records of unmatched participants were excluded from the study. All individuals who agreed to participate signed a written informed consent form. During the enrollment process, the participants' age, sex, and additional diagnostic information, if any, were systematically recorded.

Data collection was carried out by asking participants to read predetermined sound phrases for 10 s on an iPhone 14 Pro mobile phone, positioned 10 cm away from their mouths, in the presence of an expert physician. The phrases vocalized by the participants consisted of the phonetic sound 'aaa' and the words 'ana, ' 'araba, ' 'ordu, ' 'gelecek, ' 'titiz' and' ünlem.' Each participant provided seven distinct 10-s voice recordings, resulting in seven audio samples per user, and the phonetic features of the selected words were specifically chosen to reveal acoustic differences in the voices of asthma patients. The data collection process was standardized by applying the same recording procedure to all participants. The collected voice data were labeled by the attending physician as 'Asthma' or 'Healthy' according to the GINA criteria. Because a separate dataset was created for each of the seven phonetic units, each individual dataset comprised 344 samples (audio recordings), with 284 samples from patients with asthma and 60 from healthy controls. Machine learning models were developed on these datasets, and their performance was evaluated using the 5-fold cross-validation methodology detailed in section "2.5 Experimental Design." According to this methodology, the dataset for each phonetic unit was divided into 80% (approximately 275 samples) training data and 20% (approximately 69 samples) test data at each iteration.

To ensure the confidentiality of the participants, all data were anonymized and stored on a secure software platform developed by MedCase Yazılım Teknolojileri A.Ş. The data were grouped using unique identification numbers (ID) assigned to each participant and stored on a secure server. Only authorized administrators have access to the server. Figure 3 shows the time-domain data slice of the healthy and asthmatic data.

Figure 3

Sample data of subjects within the TIME domain: asthma patient (red) and healthy subject (green)



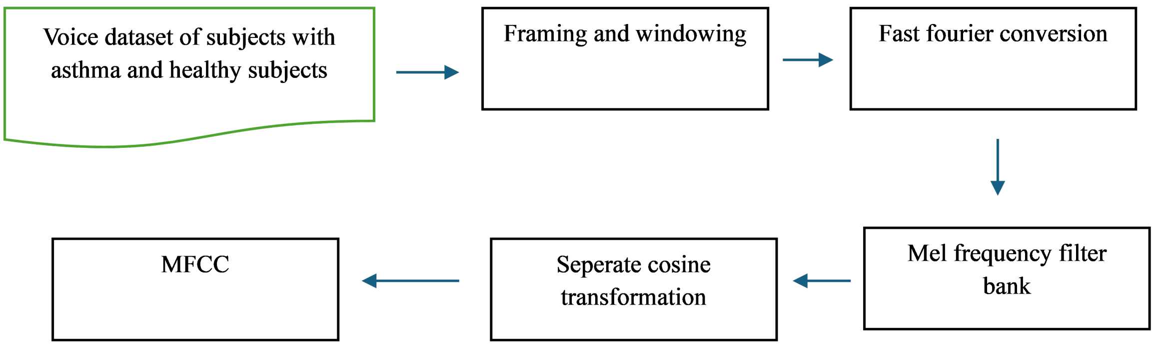
Feature Inference

Participants provided audio data for seven distinct phonetic units ('aaa,' 'ana,' 'araba,' 'ordu,' 'gelecek,' 'titiz,' and 'ünlem'), which were specifically selected due to their inclusion of diverse Turkish vowels, potentially reflecting vocal tract alterations. These recordings were captured using an iPhone 14 Pro mobile device at a 44.1-kHz sampling rate and underwent a comprehensive pre-processing and feature extraction pipeline. The Python programming language and the Librosa library were utilized for this process. To ensure computational consistency and adhere to standard feature extraction practices, all audio files were resampled to a 22.05-kHz sampling rate using the Librosa library before feature extraction. The initial pre-processing step before specific feature calculation involved steps such as silence removal from the signals and application of noise-reduction techniques to enhance signal quality. Figure 3 provides a visual example of a pre-processed audio waveform.

Subsequently, Mel-Frequency Cepstral Coefficient (MFCC) extraction was performed to convert the audio data into numerical features (Logan, 2000). MFCC is an effective technique widely used in audio processing and speech recognition; it models the frequency perception of the human ear and transforms audio signals into multidimensional feature vectors (Davis & Mermelstein, 1980; Rabiner & Schafer, 2007). The MFCC features were extracted by the following systematic steps (Figure 4):

1. Preprocessing: Silence segments were removed from the signals, and noise-reduction techniques were applied.
2. Windowing: The audio signal was divided into short time intervals suitable for analysis. For the MFCC calculation, the default parameters of the Librosa library were used: a frame size of 2048 samples (n_fft=2048), a hop length of 512 samples (hop_length=512), and a Hamming window
3. Fourier transform: The time-domain signals are transformed into frequency domains.
4. Mel Filter Application: The frequency spectrum was converted to the Mel scale to model the human hearing system.
5. Logarithmic Transform and Discrete Cosine Transform (DCT): Spectral information was processed by logarithmic transformation and DCT to obtain the final MFCC coefficients.

Figure 4
Algorithm for creating MFCC



Mel-Frequency Cepstral Coefficients (MFCCs) and Dynamic Features

For each audio signal, 12 basic MFCC coefficients were extracted and optimized to model the spectral characteristics of the human voice. These coefficients represent the amplitude information in the frequency spectrum of the audio signal (Young et al., 2006). To capture temporal dynamics within the audio signal, which is crucial as vocal characteristics may change over time due to disease, dynamic features, such as delta (first-order temporal derivative) and delta-delta (second-order temporal derivative) coefficients, were calculated from these 12 basic MFCCs and included in the analysis. This approach provided 36 MFCC-based features (12 basic coefficients, 12 delta coefficients, and 12 delta-delta coefficients).

Additional Acoustic Features and Statistics

To complement the MFCC features and enable more sensitive detection of potential disease markers, the following additional acoustic features were extracted:

- **Spectral Centroid:** Indicates the "brightness" center of the sound.
- **Spectral bandwidth: shows the energy spread** in the frequency spectrum.
- **Spectral Rolloff:** This parameter specifies the frequency below which a large portion (default 85%) of the spectral energy is concentrated, providing information about the high-frequency content.
- **Zero-Crossing Rate:** Indicates the frequency at which the signal crosses the zero axis, which is often related to the degree of noisiness or periodicity.

The mean and standard deviation values of the extracted additional acoustic features were calculated to enrich the feature vectors. For instance, instability or tonal shifts in the voice caused by respiratory illness might manifest as increases in the standard deviation of these features, whereas shifts in mean values could indicate overall changes in vocal characteristics. This resulted in 4*2 = 8 additional features from these four acoustic properties and their statistics. By combining the 36 MFCC-based features and the 8 features derived from additional acoustic properties, a total feature vector of 44 dimensions was created for each audio sample. This comprehensive and multidimensional feature extraction strategy aims to capture in more detail the spectral variations (diversity and changes) in the voices of patients with asthma, which might reflect the subtle and complex effects of diseases like asthma, thereby aiming to establish a more robust and discriminative basis for disease diagnosis. For each participant, recordings with seven different words (‘aaa,’ ‘ana,’ ‘araba,’ ‘ordu,’ ‘gelecek,’ ‘titiz,’ ‘ünlem’) were systematically pre-processed, and MFCC and other acoustic features were extracted and saved in CSV files with their labels (asthma or healthy). As a result, a separate dataset was created for each word. In addition, exploratory data analyses were performed, and class distributions were examined in detail to understand the statistical properties of the generated datasets and provide a basis for the classification models.

Performance Metrics

The precision, Recall, F1 Score, and accuracy metrics were used to evaluate the performance of the proposed system.

The inputs used in these equations reflect the relationship between the model predictions and actual situations.

- True Positive (TP - True Positive): Cases in which the model predicts asthma and is actually asthma.
- True Negative (TN - True Negative): Cases in which the model predicts as healthy and are actually healthy.
- False Positive (FP - False Positive): Cases in which the model predicts asthma but is actually healthy.
- False Negative (FN - False Negative): Cases that the model predicts as healthy but actually have asthma.

These classifications form the basis for calculating the metrics used to evaluate the performance of the model and allow us to measure its ability to detect asthma cases and correctly classify healthy individuals.

The metrics used to evaluate the performance of an asthma-detection system are crucial. Precision, Recall, Specificity, F1 Score, and Accuracy help us comprehensively evaluate the system’s success.

- **Precision**: Indicates the proportion of cases predicted by the model as asthma, which is actually asthma. Precision is important for minimizing false positives (FPs), and high-precision values are desirable.
- **Recall:** Refers to the proportion of actual asthma cases correctly detected by the model. A high sensitivity is important for reducing false negatives (FN).
- **Specificity:** The proportion of healthy individuals correctly identified by the model. Specificity is important for reducing false positives (FP). High specificity helps prevents healthy individuals from receiving unnecessary treatment.
- **F1 Score (F1 Score):** this is the harmonic mean of the sensitivity and specificity. This metric provides balanced assessment of sensitivity and specificity.
- **Accuracy:** Indicates the proportion of correct predictions among all predictions. A high-accuracy value represents the overall accuracy of the model. However, it can be misleading on unbalanced datasets; therefore, it should be evaluated together with other metrics.

In medical applications such as asthma diagnosis, there is often a trade-off between sensitivity and specificity. High sensitivity increased the ability to identify patients with asthma, whereas high specificity increased the ability to correctly identify healthy individuals. In a clinical context, the most important metric may vary depending on the potential risks of false-positive or false-negative diagnoses.

Models Used

In this study, various machine-learning models for asthma diagnosis via voice analysis were developed and evaluated. The classification algorithms used in this study and their specific parameters are described below.

1. **Naive Bayes (NB)**: The NB is a computationally efficient classifier based on probability theory that works under the assumption of feature independence.
2. **K-Nearest Neighbor (K-NN)**: This is a classification algorithm based on sample similarity. Predictions were made according to the classes of the five nearest neighbors of each sample using the k=5 parameter values.
3. **Decision Tree (Decision Tree)**: Based on hierarchical decision rules, this classifier was optimized with the parameter's maximum depth=5, minimum sampling split=5, and minimum leaf sample=2.
4. **Neural Network**: A neural network architecture consisting of four hidden layers with (256, 128, 64, 32) neuron numbers was used and trained with 'relu' activation function and adaptive learning rate.
5. **Extreme gradient boosting (XGBoost)**: An optimized version of the gradient boosting technique in terms of computational efficiency and performance.
6. **Categorical Boost (CatBoost)**: A gradient boosting algorithm designed for categorical variables was used with parameters of 100 iterations, five maximum tree depths, and a learning rate of 0.1.
7. **Adaptive Boosting (AdaBoost)**: An ensemble learning algorithm was applied with decision trees with a maximum depth of three, trained using 200 iterations, and a learning rate of 0.1.
8. **Gradient Boosting**: An ensemble learning algorithm configured with 100 trees, 0.1 learning rate, 3 maximum depth of three, and five minimum split value parameters.
9. **Random Forest**: An ensemble learning algorithm optimized with 100 trees, 10 maximum depths and 'balanced' class weights.

10. **Support Vector Machines (SVM)**: This classifier separates data points with hyperplanes, and is configured with regularization parameter C=1.0, kernel function 'rbf' and gamma parameter 'scale.'
11. **Logistic Regression**: A binary classification model configured with maximum iteration=2000, regularization intensity C=0.1, regularization l2, optimization algorithm 'saga' and class weights 'balanced' was applied.

In addition, various ensemble models were developed using the five best-performing models for each phonetic unit.

1. **Voting Ensemble**: An ensemble model was implemented in which the five best-performing models for each phoneme were combined with equal weights, and predictions were generated using a 'soft' voting strategy (averaging probability estimates).
2. **Stacking Ensemble**: This is an advanced meta-learner that uses the predictions of the top five models for each voice as input and applies Logistic Regression (maximum iteration=2000, C=0.5) as a meta-learner. It generates probability predictions with 5-fold cross-validation and the 'predict_proba' method.
3. **Weighted Voting Ensemble**: An ensemble model was developed in which the top five models for each voice were weighted according to their AUC performance metric; thus, the higher-performing models had more influence on the final classification decision.

The selection of these models was guided by the dataset size and the need for robustness in clinical applications. Deep-learning architectures (e.g., CNNs, LSTMs, Transformers) were excluded because of their requirement for large-scale datasets to achieve optimal performance (Goodfellow, Bengio, & Courville, 2016). With only 284 patients with asthma and 60 healthy controls, the sample size is significantly below the thousands or tens of thousands of examples typically required for effective deep learning training (Zhang et al., 2021). Furthermore, high-dimensional voice data combined with limited samples increases the risk of overfitting, where models memorize training data rather than generalize to new cases (Shorten, Khoshgoftaar, & Furht, 2019). Traditional machine learning and ensemble methodologies were chosen to prioritize the generalizability and reliability of medical diagnostics. These approaches have demonstrated robust performance on small- to medium-scale datasets (Dietterich, 2000; Ganaie et al., 2022). Ensemble models (Voting Ensemble, Stacking Ensemble, Weighted Voting Ensemble) were specifically developed using the top five base models for each phonetic unit, leveraging soft voting and meta-learning strategies.

Future studies should systematically assess deep learning architectures for asthma diagnosis using expanded and heterogeneous datasets.

Experimental Design

In this study, a 5-fold cross-validation methodology was adopted to evaluate the performance of the developed machine learning models. The proposed approach is based on the principle of dividing the dataset into five equal parts, training the model with four parts, and testing the remaining parts. In each iteration, a different piece of data was used as a test set; thus, all data points were evaluated in both the training and testing phases (Kohavi, 1995). By averaging the results of five iterations, the generalization performance of the models was measured in a more reliable and objective manner. The standard deviation of the results across these folds was also calculated to assess the stability of model performance.

In the data preprocessing stage, StandardScaler was applied to normalize the scale differences between the features. This standardization process optimized the performance of the algorithms by removing the imbalance between the attributes of different sizes. In addition, the SMOTE (synthetic minority oversampling technique) algorithm was used to address the sample distribution imbalance between asthma and healthy classes (Chawla et al., 2002). By generating synthetic samples of the minority class, SMOTE ensures that classification models have equal learning opportunities for both categories. Importantly, in terms of methodological rigor, the SMOTE technique was applied only to the training data, while the test data were retained in their original form, preserving their true distribution. This strategic approach enables a more realistic assessment of model performance under real-world conditions in clinical settings.

The data preprocessing steps (standardization and SMOTE) and model training procedures were integrated into the cross-validation cycle. This integration was realized through the ImbPipeline structure of the Scikit-learn library, ensuring that the data preprocessing and model training phases were independently and methodologically consistent.

The model training and evaluation processes were performed on a MacBook Pro (2023) computer with an Apple M3 Pro processor (36 GB RAM) running on a macOS Sonoma operating system in a high-performance computing environment. The following specific libraries were used to implement the algorithms developed in the Python 3.11.9 programming language using Spyder Integrated.

Development Environment:

1. Scikit-learn 1.5: Used to implement machine learning models and cross-validation procedures.
2. Librosa 0.10, provides specialized audio data processing and acoustic feature extraction functions.
3. Matplotlib: Used for visualizing research findings and model performance.

This comprehensive technological and methodological framework maximized the scientific reproducibility of the study's experimental design and its potential for integration into clinical practice.

Results

This study comprehensively evaluated the diagnostic performance of voice analysis-based machine-learning models for asthma diagnosis. The classification efficiency of the developed algorithms was systematically tested on audio samples containing various phonetic units ('aaa, ' 'ana,' 'araba,' 'gelecek, ' 'ordu, ' 'titiz,' 'ünlem'), and the quantitative results obtained are presented in detail in Tables 1–7. The results are reported as mean ± standard deviation from the 5-fold cross-validation to reflect the consistency of model performance across different data folds. The analyses demonstrate that ensemble models perform better than individual classifiers, with the Stacking Ensemble and Voting Ensemble approaches in particular showing superior performance in terms of diagnostic accuracy.

Analyzing the sound "aaa"

When the performance metrics of the machine learning models applied on the 'aaa' sound presented in Table 1 and Figure 5 are analyzed, it can be seen that the Stacking Ensemble model exhibits superior classification capacity. This meta-learning model achieved the highest overall performance, with an accuracy of 80.2% ± 2.6% and an AUC of 0.627 ± 0.107. Similarly, the Neural Network model showed equivalent discriminative ability with an AUC value of 0.628 ± 0.141, followed by the Gradient Boosting algorithm with an AUC value of 0.607 ± 0.065.

When we evaluated model performance from the perspective of classification algorithms, it is observed that ensemble learning methods produced more consistent results than individual classifiers. The Voting

Ensemble and Weighted Voting Ensemble models exhibited almost identical performance profiles (AUC values of 0.649 ± 0.112 and 0.649 ± 0.113, respectively), suggesting that the model weighting strategy did not provide a clear advantage for the 'aaa' sound.

The simpler classification approaches, Naive Bayes (57.0% ± 4.6% accuracy, 0.564 ± 0.083 AUC) and K-NN (57.2% ± 7.8% accuracy, 0.620 ± 0.088 AUC) algorithms, demonstrated limited discrimination capacity compared with the complex models. This suggests that voice analysis for asthma diagnosis requires nonlinear models and capable of capturing complex relationships.

In terms of performance metrics, although the Random Forest model achieved high accuracy (75.8% ±5.5%) and an F1 score (0.856 ± 0.039), it showed potential limitations in clinical applications due to its low specificity (0.183 ± 0.062). This suggests that the model has difficulty correctly categorizing healthy individuals and, therefore, may produce a high rate of false positive results under real conditions.

In terms of the balance of sensitivity and specificity, the Stacking Ensemble model (sensitivity: 0.874 ± 0.035, specificity: 0.443 ± 0.096) was highly successful in detecting patients with asthma, but it performed moderately in discriminating healthy individuals. This profile has potential value for screening purposes by detecting high-risk individuals with minimal false negatives.

The results of the 'Aha' sound analysis demonstrated the feasibility of voice-based asthma diagnosis and revealed the limiting factors of a single phonetic unit. The AUC values of the models were generally in the range of 0.60-0.65, suggesting that the 'aaa' sound alone is not sufficient for a definitive diagnosis; however, in combination with other phonetic units, its diagnostic value may increase.

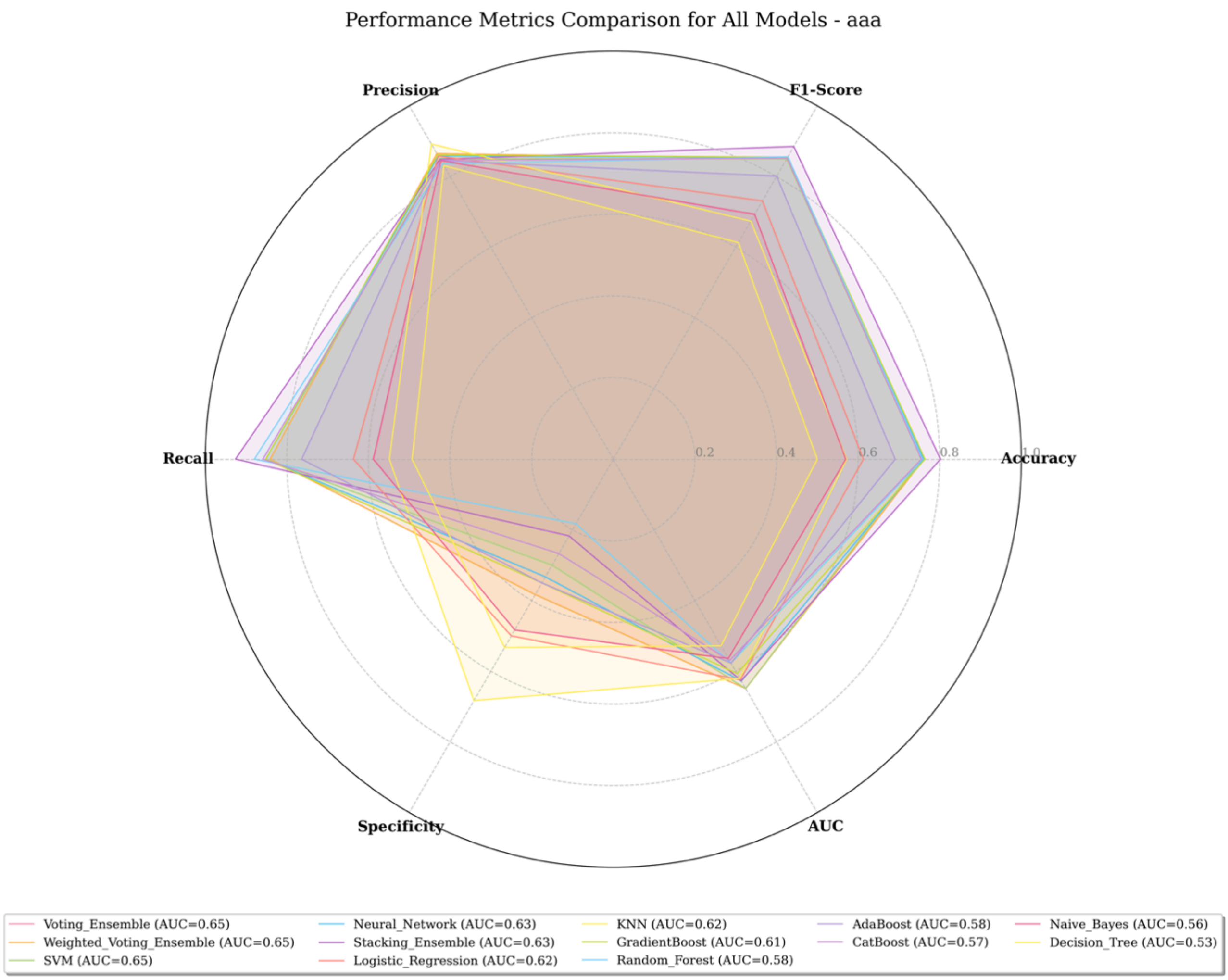
Table 1

Machine learning model performance evaluation metrics for the sound "aaa"

Model	Accuracy	F1-Score	Precision	Recall	Specificity	AUC
Naive_Bayes	0.570±0.046	0.693±0.033	0.845±0.045	0.588±0.035	0.483±0.170	0.564±0.083
KNN	0.572±0.078	0.674±0.081	0.892±0.040	0.549±0.100	**0.683±0.143**	0.620±0.088
Decision_Tree	0.500±0.080	0.613±0.091	0.831±0.029	0.493±0.103	0.533±0.100	0.528±0.067
Neural_Network	0.761±0.063	0.854±0.042	0.858±0.035	0.852±0.061	0.333±0.175	0.628±0.141
CatBoost	0.755±0.065	0.852±0.045	0.846±0.022	0.859±0.070	0.267±0.062	0.570±0.052
AdaBoost	0.691±0.087	0.801±0.066	0.845±0.033	0.764±0.094	0.350±0.097	0.577±0.086
GradientBoost	0.764±0.055	0.855±0.038	0.861±0.021	0.852±0.059	0.350±0.082	0.607±0.065
Random_Forest	0.758±0.055	0.856±0.039	0.836±0.011	0.880±0.073	0.183±0.062	0.577±0.076
SVM	0.755±0.066	0.850±0.047	0.851±0.019	0.852±0.082	0.300±0.085	**0.649±0.120**
Logistic_Regression	0.613±0.063	0.731±0.045	0.858±0.050	0.637±0.041	0.500±0.175	0.624±0.100
Voting_Ensemble	0.761 ± 0.081	0.850 ± 0.060	**0.865 ± 0.032**	0.841 ± 0.098	0.383 ± 0.163	0.649 ± 0.112
Stacking_Ensemble	**0.802 ± 0.026**	**0.885 ± 0.017**	0.849 ± 0.016	**0.926 ± 0.041**	0.217 ± 0.113	0.627 ± 0.107
Weighted_Voting_Ensemble	0.761 ± 0.081	0.850 ± 0.060	**0.865 ± 0.032**	0.841 ± 0.098	0.383 ± 0.163	0.649 ± 0.113

Figure 5

Radar graphic for the sound "aaa"



Analysis of the word 'ana'

In the performance evaluation of the classification models on the word 'ana' presented in Table 2 and Figure 6, the Stacking Ensemble algorithm showed the highest discriminative performance with an accuracy of 78.8% ± 1.6% and an AUC value of 0.639 ± 0.056. The Neural Network model ranked second, with an accuracy of 77.9 ± 3.4% and an AUC value of 0.600 ± 0.072.

In the model-based evaluation, the Random Forest algorithm showed a high capacity to correctly identify asthmatic patients, with a sensitivity of 87.0% ± 6.1%. However, this model has a significant limitation in identifying healthy individuals, with a specificity of 0.233 ± 0.082. This unstable performance profile, similar to the Random Forest results for the 'aaa' sound, suggests that the model is prone to producing a high rate of false positives in clinical applications.

The Logistic Regression algorithm showed moderate success in discriminating between asthmatic and healthy individuals, with an AUC of 0.628 ± 0.100 and specificity of 0.550 ± 0.113. This model provides a more balanced performance between sensitivity and specificity, making it a reliable option for clinical applications.

The K-NN algorithm demonstrated unbalanced performance with an accuracy of 55.8% ± 5.7%, precision of 0.887 ± 0.034, and sensitivity of 0.535 ± 0.084. These results indicate that a high proportion of cases classified as positive by the K-NN model were indeed asthmatic (high precision); however, it was only able to detect half of all asthmatic patients (moderate precision). This profile is consistent with the limited discriminative capacity of the simple algorithms observed in the 'aaa' sound.

Analyzing the ensemble learning approaches, the Voting Ensemble and Weighted Voting Ensemble algorithms showed almost identical performance values for the word 'ana' (both AUC 0.650 ± 0.080). This result suggests that for the phonetic unit 'ana,' as for the sound 'aaa,' different model-weighting strategies do not provide a clear advantage.

In terms of the F1-score metric, the Stacking Ensemble (0.879 ± 0.009) and XGBoost (0.858 ± 0.021) models yielded the highest values. These results demonstrate that these models can achieve an optimal balance between sensitivity and precision, particularly in datasets with class imbalance.

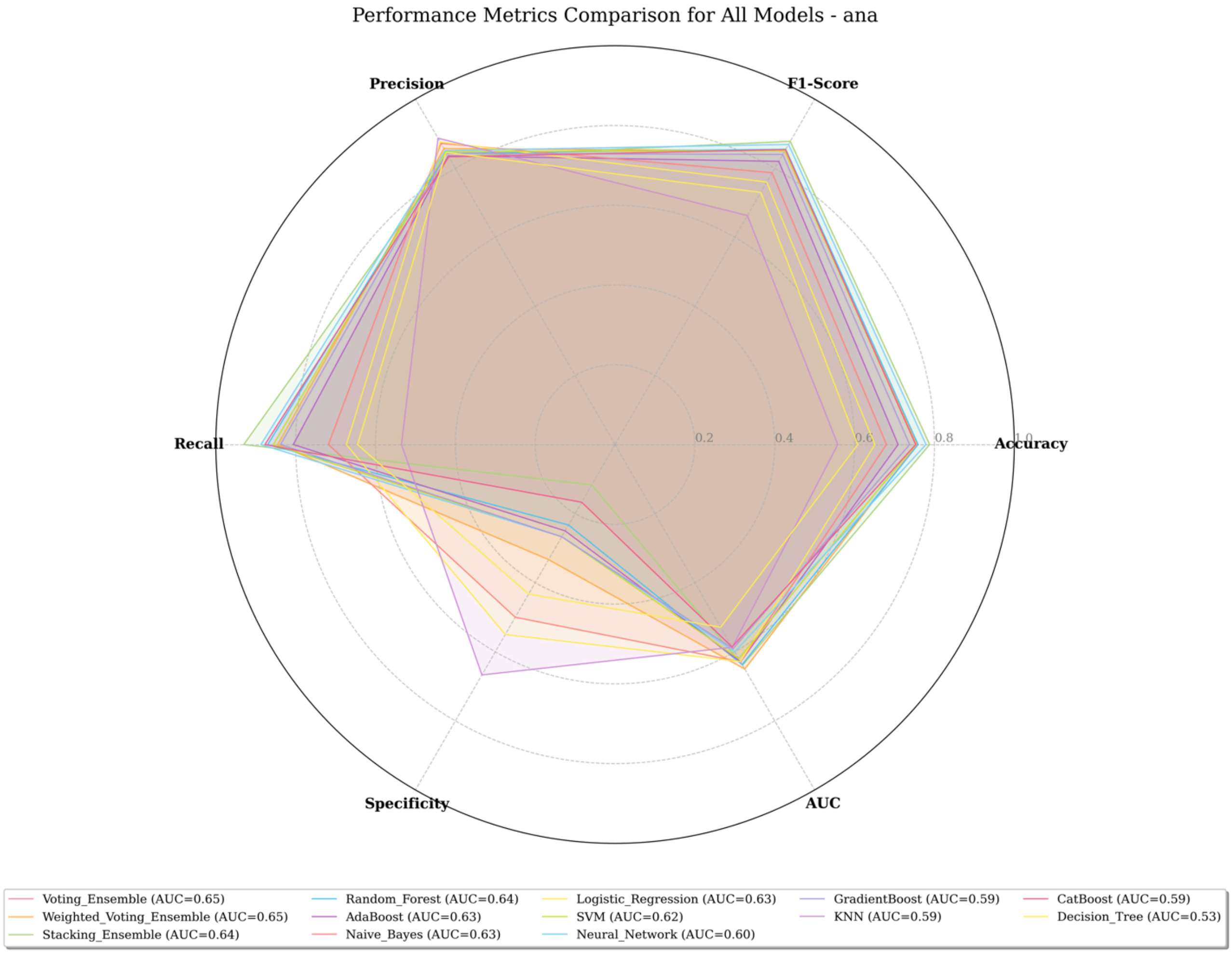
The results of the analysis of the word 'ana,' similar to the findings for the sound 'aaa,' show the limitations of the classification performance of a single phonetic unit but reveal that ensemble-based approaches in particular produce promising results. AUC values in the range of 0.600-0.650 indicate the potential value of this phonetic unit as an auxiliary biomarker for asthma diagnosis.

Table 2

Machine learning model performance evaluation metrics for the sound "ana"

Model	Accuracy	F1-Score	Precision	Recall	Specificity	AUC
Naive_Bayes	0.680 ± 0.030	0.787 ± 0.022	0.872 ± 0.019	0.718 ± 0.028	0.500 ± 0.075	0.629 ± 0.077
KNN	0.558 ± 0.057	0.663 ± 0.063	0.887 ± 0.034	0.535 ± 0.084	**0.667 ± 0.139**	0.587 ± 0.068
Decision_Tree	0.608 ± 0.029	0.730 ± 0.025	0.846 ± 0.037	0.645 ± 0.045	0.433 ± 0.162	0.529 ± 0.081
Neural_Network	0.779 ± 0.034	0.869 ± 0.023	0.851 ± 0.009	0.887 ± 0.040	0.267 ± 0.033	0.600 ± 0.072
CatBoost	0.753 ± 0.016	0.854 ± 0.011	0.833 ± 0.008	0.877 ± 0.022	0.167 ± 0.053	0.586 ± 0.052
AdaBoost	0.709 ± 0.029	0.820 ± 0.020	0.836 ± 0.015	0.806 ± 0.035	0.250 ± 0.091	0.629 ± 0.051
GradientBoost	0.738 ± 0.056	0.840 ± 0.038	0.843 ± 0.019	0.838 ± 0.059	0.267 ± 0.062	0.591 ± 0.052
Random_Forest	0.759 ± 0.050	0.855 ± 0.033	0.843 ± 0.017	0.870 ± 0.061	0.233 ± 0.082	0.637 ± 0.068
SVM	0.753 ± 0.025	0.851 ± 0.019	0.848 ± 0.021	0.856 ± 0.048	0.267 ± 0.143	0.620 ± 0.093
Logistic_Regression	0.651 ± 0.051	0.760 ± 0.040	**0.876 ± 0.030**	0.673 ± 0.051	0.550 ± 0.113	0.628 ± 0.100
Voting_Ensemble	0.756 ± 0.047	0.850 ± 0.033	0.857 ± 0.010	0.845 ± 0.066	0.333 ± 0.075	**0.650 ± 0.080**
Stacking_Ensemble	**0.788 ± 0.016**	**0.878 ± 0.010**	0.833 ± 0.011	**0.930 ± 0.027**	0.117 ± 0.085	0.639 ± 0.056
Weighted_Voting_Ensemble	0.756 ± 0.047	0.850 ± 0.033	0.857 ± 0.010	0.845 ± 0.066	0.333 ± 0.075	**0.650 ± 0.080**

Figure 6
Radar graphic for sound "ana"



Analysis of the word 'araba'

In the classification analyses of the word 'araba' presented in Table 3 and Figure 7, the Stacking Ensemble model exhibited the highest discriminative performance among all models, with an accuracy of 82.3 ± 1.5% and an AUC of 0.684 ± 0.076. This result confirms the superiority of ensemble learning approaches in generalizing complex acoustic patterns, as in previous phonetic units ('aaa' and 'ana'). In the model-based evaluation, the Random Forest algorithm showed superior sensitivity in detecting asthma cases (sensitivity: 95.1% ± 2.8%. However, the low specificity value of 0.167 ± 0.075% maintains the similarity limitation observed for the words 'aaa' and' ana.' This suggests that the model can produce a high rate of false-positive results in clinical applications. When the ensemble learning approaches were examined, the Voting Ensemble and Weighted Voting Ensemble models showed high discriminative performance for the word 'araba' with AUC values of 0.704 ± 0.081 and 0.703 ± 0.081, respectively. These results show that, unlike previous findings for phonetic units, the ensemble models achieved higher AUC values for the word 'araba.' This suggests that the acoustic features of the word 'araba' may be more discriminative for asthma detection. The Logistic Regression algorithm showed a balanced performance, with an AUC of 0.685 ± 0.078 and a specificity value of 0.600 ± 0.133. This model exhibits a more balanced profile between sensitivity and specificity than previous word analyses, thereby providing a reliable option for clinical applications. The CatBoost algorithm demonstrated strong predictive capacity with an accuracy of 79.9 ± 3.4% and an AUC value of 0.673 ± 0.092. This result demonstrates that the proposed algorithm, which is optimized for processing categorical variables, can also effectively model complex patterns in audio data. In contrast, the K-NN model demonstrated the lowest performance among all models, with an accuracy of 48.8% ± 6.5% and an AUC value of 0.521 ± 0.048. This result confirms the limitations of neighborhood-based approaches in high-dimensional and complex datasets, such as voice analysis, as observed in previous “wording” analyses.

The results of the analysis of the word ‘araba’ showed that, compared with previous phonetic units, the ensemble models achieved higher AUC values (0.684-0.704), indicating that this word may be a stronger acoustic biomarker for asthma diagnosis. This suggests that phonetic units containing multiple vowel and consonant combinations may more clearly reflect the effects of asthma on the voice.

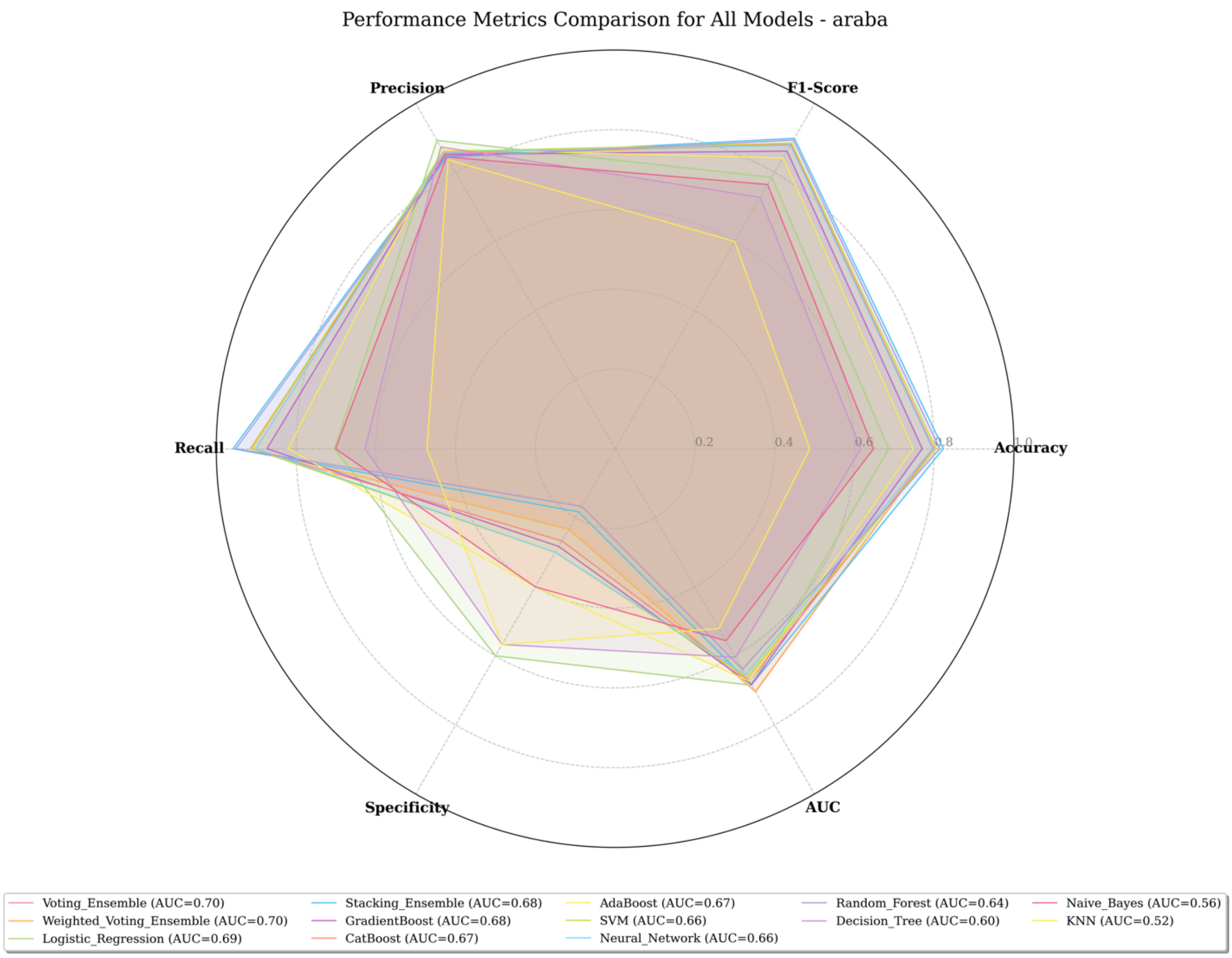
Table 3

Machine learning model performance evaluation metrics for the sound “araba”

Model	Accuracy	F1-Score	Precision	Recall	Specificity	AUC
Naive_Bayes	0.648 ± 0.065	0.765 ± 0.053	0.845 ± 0.022	0.701 ± 0.075	0.400 ± 0.062	0.556 ± 0.083
KNN	0.488 ± 0.065	0.600 ± 0.070	0.837 ± 0.044	0.472 ± 0.076	**0.567 ± 0.143**	0.521 ± 0.048
Decision_Tree	0.616 ± 0.045	0.728 ± 0.040	0.874 ± 0.035	0.627 ± 0.058	**0.567 ± 0.143**	0.604 ± 0.071
Neural_Network	0.796 ± 0.040	0.879 ± 0.025	0.859 ± 0.025	0.901 ± 0.041	0.300 ± 0.135	0.655 ± 0.041
CatBoost	0.799 ± 0.034	0.882 ± 0.020	0.855 ± 0.023	0.912 ± 0.020	0.267 ± 0.122	0.673 ± 0.092
AdaBoost	0.747 ± 0.045	0.842 ± 0.029	0.867 ± 0.028	0.820 ± 0.049	0.400 ± 0.143	0.673 ± 0.075
GradientBoost	0.770 ± 0.041	0.862 ± 0.026	0.852 ± 0.019	0.873 ± 0.037	0.283 ± 0.085	0.683 ± 0.056
Random_Forest	0.814 ± 0.029	0.894 ± 0.017	0.844 ± 0.014	0.951 ± 0.028	0.167 ± 0.075	0.639 ± 0.081
SVM	0.805 ± 0.020	0.885 ± 0.013	0.861 ± 0.014	0.912 ± 0.025	0.300 ± 0.085	0.663 ± 0.052
Logistic_Regression	0.686 ± 0.055	0.786 ± 0.044	**0.893 ± 0.034**	0.704 ± 0.063	0.600 ± 0.133	0.685 ± 0.078
Voting_Ensemble	0.796 ± 0.027	0.881 ± 0.016	0.850 ± 0.015	0.915 ± 0.023	0.233 ± 0.082	**0.704 ± 0.081**
Stacking_Ensemble	**0.823 ± 0.015**	**0.899 ± 0.009**	0.847 ± 0.010	**0.958 ± 0.018**	0.183 ± 0.062	0.684 ± 0.076
Weighted_Voting_Ensemble	0.796 ± 0.027	0.881 ± 0.016	0.850 ± 0.015	0.915 ± 0.023	0.233 ± 0.082	0.703± 0.081

Figure 7

Radar graphic for "araba"



Analysis of the word 'gelecek'

In the classification analyses performed on the word 'gelecek' presented in Table 4 and Figure 8, the Stacking Ensemble model exhibited the highest discriminative performance with an accuracy of 82.0% ± 3.1% and an AUC value of 0.702 ± 0.028. This result confirms the superior ensemble learning approaches observed in the previous phonetic units for the word 'gelecek.' The voting ensemble and weighted voting ensemble models demonstrated similarly high performance, with AUC values of 0.717 ± 0.062 and 0.716 ± 0.062, respectively.

In the model-based evaluation, the Logistic Regression algorithm exhibited remarkable discriminative power, with an AUC value of 0.709 ± 0.049. This result, similar to the Logistic Regression performance for the word 'araba,' suggests that this model can effectively discriminate vocal characteristics between the asthma and control groups. However, a specificity value of 0.583 ± 0.091 indicates that false-favorable rates may increase.

The SVM algorithm demonstrated strong performance with an accuracy of 79.1% ± 2.0% and an AUC value of 0.643 ± 0.069. This model is characterized by its capacity to effectively separate complex sound features in high-dimensional spaces. Consistent with its performance in previous phonetic units, the discriminative power of the SVM for the word ‘gelecek’ was moderate to high.

Remarkably, the Decision Tree algorithm showed limited performance with an accuracy of 55.3% ± 7.4% and an AUC of 0.551 ± 0.053. This confirms the inability of a single decision tree to model complex audio features, which has been observed in previous studies. Similarly, the K-NN model also showed limited discriminative capacity, with an accuracy of 59.6 ± 6.5% and an AUC value of 0.606 ± 0.071. This result supports the limited effectiveness of neighborhood-based approaches in complex datasets, such as voice analysis, as observed for the words ‘aaa’, ‘ana’ and’ araba.’

The analysis results of the word ‘gelecek’ show that compared with the previous phonetic units, the ensemble models reach higher AUC values (0.702-0.717), indicating that this word may be a stronger acoustic biomarker. In particular, the 0.717 AUC value of the Voting Ensemble model is the highest discriminative value among all the phonetic units analyzed thus far. This suggests that the variety of vowel and consonant sounds and the complexity of articulatory movements in the word ‘gelecek’ may capture the effects of asthma on the voice more clearly.

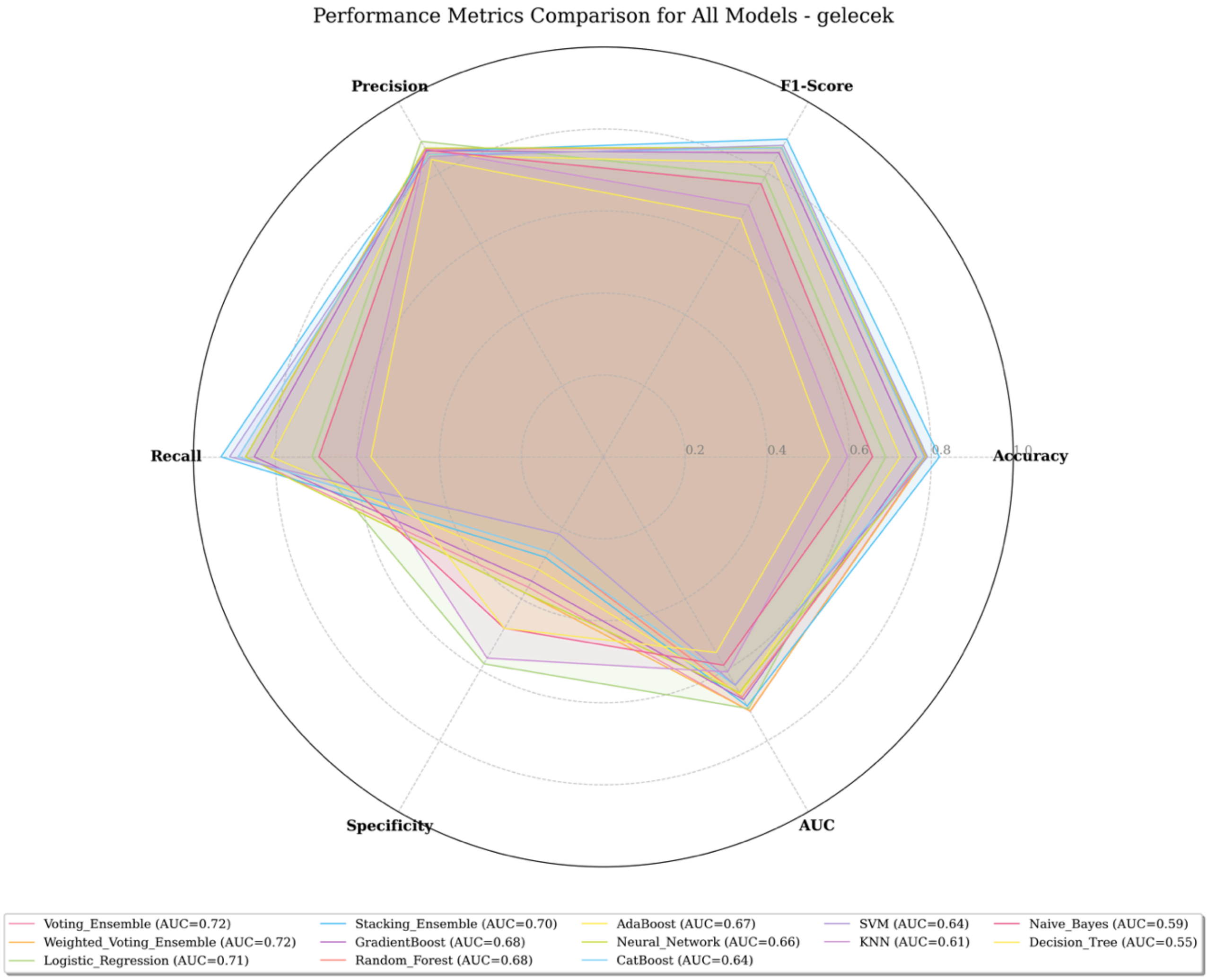
The high AUC values observed for the word ‘gelecek’ suggest that words with multiple syllables and different phonetic features are more effective biomarkers for asthma diagnosis. This finding emphasizes the diagnostic value of phonetic diversity in clinical practice.

Table 4

Machine learning model performance evaluation metrics for the sound “gelecek”

Model	Accuracy	F1-Score	Precision	Recall	Specificity	AUC
Naive_Bayes	0.657 ± 0.028	0.769 ± 0.020	0.864 ± 0.024	0.694 ± 0.022	0.483 ± 0.097	0.587 ± 0.062
KNN	0.596 ± 0.065	0.709 ± 0.060	0.867 ± 0.033	0.603 ± 0.077	**0.567 ± 0.097**	0.606 ± 0.071
Decision_Tree	0.553 ± 0.074	0.671 ± 0.079	0.838 ± 0.022	0.567 ± 0.104	0.483 ± 0.111	0.551 ± 0.053
Neural_Network	0.788 ± 0.031	0.872 ± 0.019	0.870 ± 0.018	0.873 ± 0.023	0.383 ± 0.085	0.664 ± 0.050
CatBoost	0.782 ± 0.032	0.871 ± 0.018	0.852 ± 0.023	0.891 ± 0.014	0.267 ± 0.122	0.644 ± 0.085
AdaBoost	0.724 ± 0.012	0.829 ± 0.007	0.850 ± 0.028	0.810 ± 0.029	0.317 ± 0.162	0.669 ± 0.031
GradientBoost	0.764 ± 0.017	0.857 ± 0.010	0.863 ± 0.030	0.852 ± 0.032	0.350 ± 0.186	0.684 ± 0.021
Random_Forest	0.782 ± 0.017	0.871 ± 0.008	0.853 ± 0.022	0.891 ± 0.017	0.267 ± 0.133	0.677 ± 0.047
SVM	0.791 ± 0.020	0.878 ± 0.011	0.847 ± 0.014	0.912 ± 0.015	0.217 ± 0.085	0.643 ± 0.069
Logistic_Regression	0.689 ± 0.063	0.789 ± 0.051	**0.889 ± 0.026**	0.711 ± 0.069	0.583 ± 0.091	0.709 ± 0.049
Voting_Ensemble	0.785 ± 0.050	0.870 ± 0.033	0.867 ± 0.024	0.873 ± 0.049	0.367 ± 0.113	**0.717 ± 0.062**
Stacking_Ensemble	**0.820 ± 0.031**	**0.895 ± 0.018**	0.861 ± 0.023	**0.933 ± 0.017**	0.283 ± 0.125	0.702 ± 0.028
Weighted_Voting_Ensemble	0.788 ± 0.047	0.871 ± 0.031	0.870 ± 0.021	0.873 ± 0.049	0.383 ± 0.100	0.716 ± 0.062

Figure 8
Radar graphic for "gelecek"



Analysis of the word 'ordu'

In the classification analyses performed on the word 'ordu' presented in Table 5 and Figure 9, the Stacking Ensemble model exhibited the highest overall performance with an accuracy of 79.6% ± 3.0% and an AUC of 0.678 ± 0.043. This result confirms the superiority of ensemble learning approaches for generalizing complex acoustic patterns, as observed in previous phonetic units. In the model-based evaluation, the SVM algorithm showed strong discriminative capacity with an accuracy of 79.0% ± 4.6% and an AUC value of 0.686 ± 0.077. This result confirms that SVM can discriminate effectively in high-dimensional acoustic feature space, as observed for the word 'gelecek.' However, the limitation in the specificity values is a factor to be considered in clinical applications. The Random Forest model showed high sensitivity in detecting asthma cases, with a sensitivity of 88.7 ± 5.4%. However, it showed a significant limitation in identifying healthy individuals, with a specificity of 0.200 ± 0.067%. This unbalanced performance profile is consistent with the low specificity problem observed for Random Forest in previous "wording" analyses. This highlights the risk of producing a high rate of false positive results in clinical applications. The Logistic Regression algorithm showed moderate success in discriminating between asthmatic and healthy individuals, with an AUC of 0.665 ± 0.035 and specificity value of 0.533 ± 0.135. This result is consistent with the Logistic Regression performance observed for the words 'gelecek' araba' araba,' confirming that the model has balanced discriminative power in voice-based biomarkers. When the ensemble learning approaches were analyzed, the Weighted Voting Ensemble model with an AUC value of 0.681 ± 0.041 minimally exceeded the AUC value of 0.680 ± 0.041 for the Voting Ensemble model. This demonstrates that, as in previous word analyses, the model weighting strategy for the word 'ordu' only maximizes classification performance. The K-NN algorithm exhibited limited discriminative capacity, with an accuracy of 54.5% ± 6.4% and an AUC value of 0.573 ± 0.106. This result confirms the limited effectiveness of neighborhood-based approaches in high-dimensional and complex datasets, such as voice analysis, which has been consistently observed in all previous word analyses. The results of the analysis of the word 'ordu' show that, compared to the last phonetic units, the community models exhibit slightly lower discriminative power than the word 'gelecek' (0.702-0.717), but higher discriminative power than the words 'aaa' (0.627-0.649) and 'ana' (0.639-0.650), with AUC values in the range 0.678-0.686. This suggests that the phonetic features of the word 'ordu' may reflect the effects of asthma on the voice at a medium-high level.

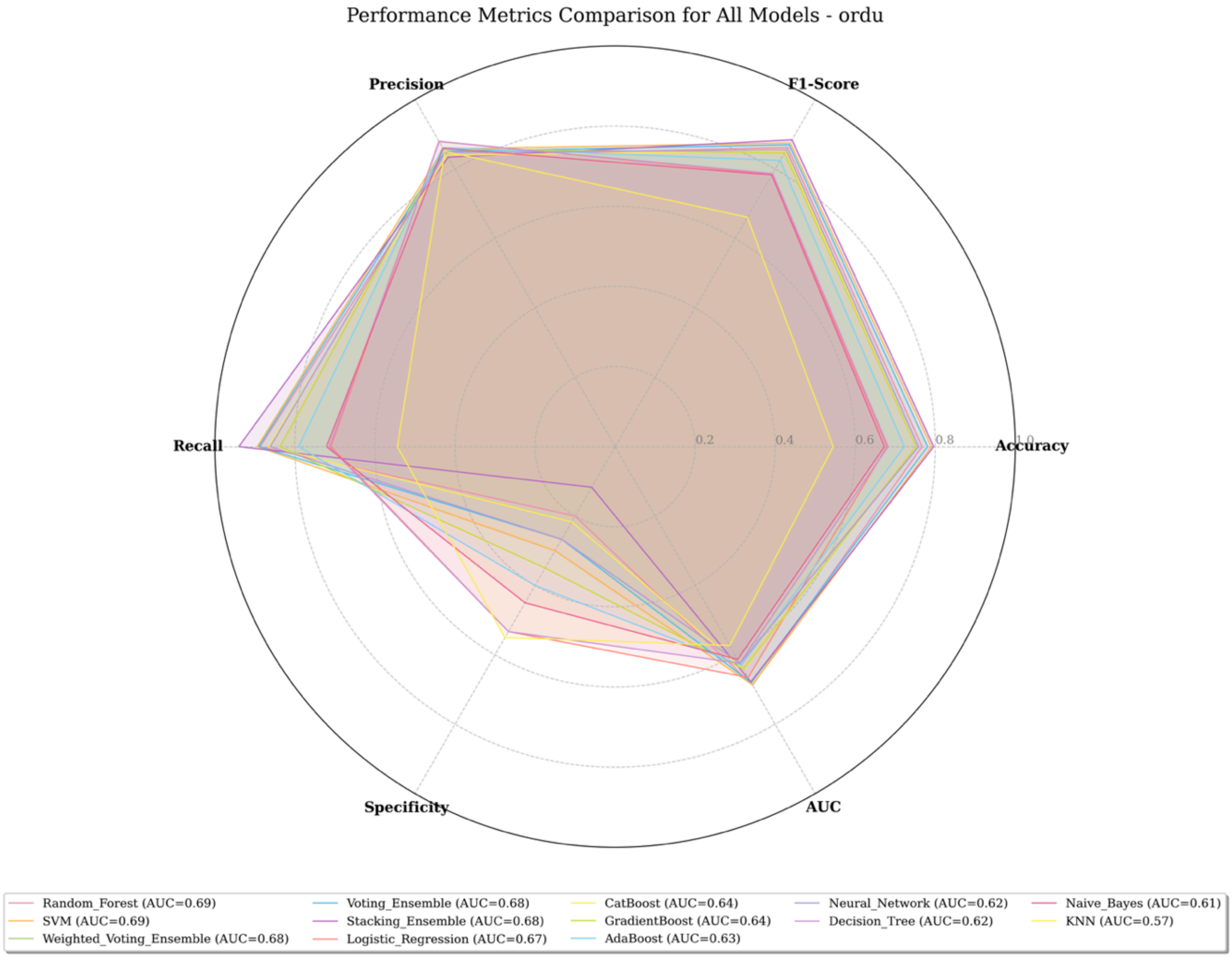
Table 5

Machine learning model performance evaluation metrics for the sound "ordu"

Model	Accuracy	F1-Score	Precision	Recall	Specificity	AUC
Naive_Bayes	0.673 ± 0.054	0.783 ± 0.044	0.860 ± 0.015	0.721 ± 0.064	0.450 ± 0.041	0.613 ± 0.076
KNN	0.545 ± 0.064	0.661 ± 0.059	0.851 ± 0.044	0.544 ± 0.073	**0.550 ± 0.145**	0.573 ± 0.106
Decision_Tree	0.682 ± 0.047	0.787 ± 0.035	0.879 ± 0.031	0.714 ± 0.044	0.533 ± 0.135	0.624 ± 0.070
Neural_Network	0.758 ± 0.026	0.855 ± 0.016	0.848 ± 0.017	0.862 ± 0.026	0.267 ± 0.097	0.625 ± 0.085
CatBoost	0.752 ± 0.027	0.852 ± 0.018	0.840 ± 0.018	0.866 ± 0.035	0.217 ± 0.113	0.642 ± 0.034
AdaBoost	0.723 ± 0.026	0.825 ± 0.021	0.862 ± 0.015	0.791 ± 0.041	0.400 ± 0.097	0.630 ± 0.042
GradientBoost	0.752 ± 0.044	0.847 ± 0.032	0.858 ± 0.012	0.837 ± 0.057	0.350 ± 0.062	0.641 ± 0.042
Random_Forest	0.767 ± 0.051	0.862 ± 0.033	0.839 ± 0.017	0.887 ± 0.054	0.200 ± 0.067	**0.687 ± 0.060**
SVM	0.790 ± 0.046	0.875 ± 0.027	0.858 ± 0.026	0.894 ± 0.033	0.300 ± 0.135	0.686 ± 0.077
Logistic_Regression	0.679 ± 0.027	0.784 ± 0.027	**0.880 ± 0.023**	0.710 ± 0.055	0.533 ± 0.135	0.665 ± 0.035
Voting_Ensemble	0.781 ± 0.025	0.870 ± 0.015	0.851 ± 0.015	0.890 ± 0.021	0.267 ± 0.082	0.680 ± 0.041
Stacking_Ensemble	**0.796 ± 0.030**	**0.884 ± 0.018**	0.834 ± 0.013	**0.940 ± 0.026**	0.117 ± 0.067	0.678 ± 0.043
Weighted_Voting_Ensemble	0.781 ± 0.025	0.870 ± 0.015	0.851 ± 0.015	0.890 ± 0.021	0.267 ± 0.082	0.681 ± 0.041

Figure 9

Radar graphic for "ordu"



Analysis of the word 'titiz'

In the classification analyses performed on the word 'titiz' presented in Table 6 and Figure 10, the Voting Ensemble and Weighted Voting Ensemble models exhibited the highest discriminative performance with AUC values of 0.736 ± 0.067 and 0.736 ± 0.068, respectively. These results confirm the superiority of ensemble learning approaches in generalizing complex acoustic patterns for the word 'fastidious,' as observed for all previous phonetic units. The Stacking Ensemble model showed superior performance, with an accuracy of 79.9% ± 4.0% and an AUC of 0.731 ± 0.064. In the model-based evaluation, the Neural Network algorithm exhibited a discriminative power close to that of the ensemble models with an accuracy of 79.9% ± 6.0% and an AUC value of 0.734 ± 0.056. This result shows that artificial neural networks can effectively model the acoustic features of the word 'titiz.' The SVM algorithm also demonstrated a strong classification capacity, with an accuracy of 76.4 ± 2.9% and an AUC of 0.718 ± 0.066. This performance profile confirms that the SVM can discriminate effectively in high-dimensional acoustic feature space, as observed for the words 'gelecek' and' ordu.' The Logistic Regression algorithm showed remarkable performance with an AUC of 0.710 ± 0.095 and specificity value of 0.567 ± 0.226. This result is very close to the AUC value of 0.709 obtained for the word' palecek, demonstrating that the Logistic Regression model exhibits consistent discriminative power across different phonetic units.

As observed in all previous wording analyses, the Random Forest model showed high sensitivity in detecting asthma cases, with a sensitivity of 87.7 ± 3.7%. However, it had a marked limitation in identifying healthy individuals, with specificity of 0.200 ± 0.085%. This unbalanced performance profile was also consistently observed for the words 'aaa', 'ana, ' 'araba, ' 'gelecek' and 'ordu, ' indicating that the Random Forest model tends to produce a high rate of false positives in clinical applications.

Simpler models, such as K-NN and Decision Tree, exhibited limited discriminative capacity with AUC values of 0.583 ± 0.088 and 0.577 ± 0.051, respectively. This result confirms that simple algorithms are inadequate for complex and high-dimensional datasets, such as voice analysis, as consistently observed in all previous word analyses.

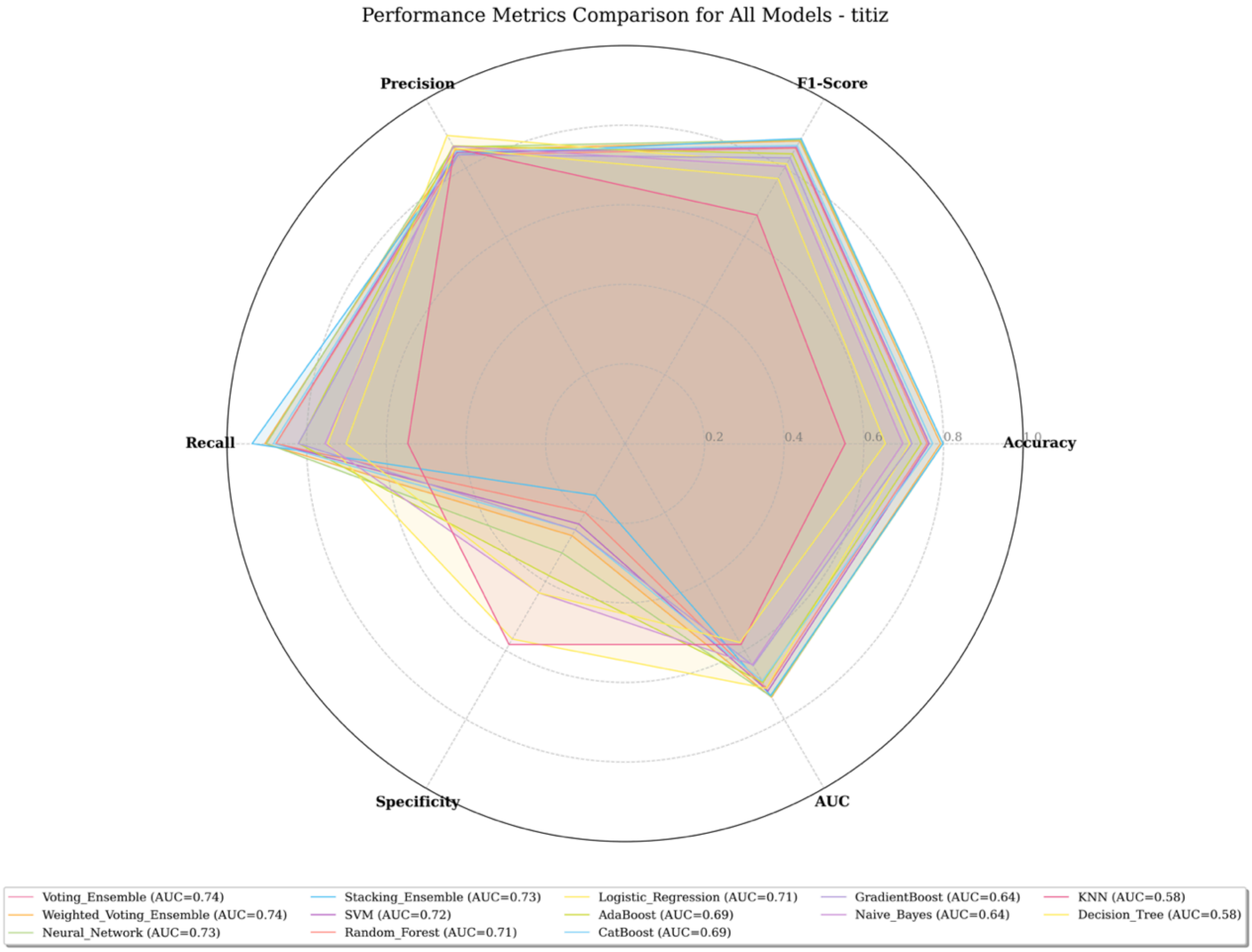
The analysis results of the word 'rigorous' show that, compared to the previous phonetic units, the ensemble models exhibit the highest discriminative power among all the words analyzed thus far, with AUC values in the range of 0.731-0.736. These values also exceed the AUC range of 0.702-0.717 obtained for the word 'gelecek,' indicating that 'titiz' may be the most powerful acoustic biomarker for asthma diagnosis. This suggests that the complexity of consonant sound combinations and articulatory movements in the word 'titiz' may reflect the effects of asthma on the voice more clearly.

Table 6

Machine learning model performance evaluation metrics for the sound "titiz"

Model	Accuracy	F1-Score	Precision	Recall	Specificity	AUC
Naive_Bayes	0.698±0.040	0.804±0.025	0.863±0.031	0.754±0.029	0.433±0.143	0.641±0.104
KNN	0.553 ± 0.088	0.663 ± 0.085	0.857 ± 0.035	0.546 ± 0.101	**0.583 ± 0.075**	0.583 ± 0.088
Decision_Tree	0.654 ± 0.033	0.769 ± 0.026	0.856 ± 0.031	0.701 ± 0.041	0.433 ± 0.143	0.577 ± 0.051
Neural_Network	**0.799 ± 0.060**	0.881 ± 0.037	0.861 ± 0.024	0.901 ± 0.052	0.317 ± 0.097	0.734 ± 0.056
CatBoost	0.773 ± 0.027	0.865 ± 0.019	0.848 ± 0.006	0.884 ± 0.041	0.250 ± 0.053	0.689 ± 0.066
AdaBoost	0.744 ± 0.019	0.841 ± 0.013	**0.864 ± 0.020**	0.820 ± 0.025	0.383 ± 0.113	0.693 ± 0.048
GradientBoost	0.721 ± 0.018	0.829 ± 0.014	0.838 ± 0.007	0.821 ± 0.029	0.250 ± 0.053	0.644 ± 0.019
Random_Forest	0.759 ± 0.033	0.857 ± 0.022	0.838 ± 0.015	0.877 ± 0.037	0.200 ± 0.085	0.711 ± 0.062
SVM	0.764 ± 0.029	0.860 ± 0.016	0.845 ± 0.024	0.877 ± 0.019	0.233 ± 0.133	0.718 ± 0.066
Logistic_Regression	0.715 ± 0.038	0.812 ± 0.025	0.894 ± 0.049	0.747 ± 0.041	0.567 ± 0.226	0.710 ± 0.095
Voting_Ensemble	0.793 ± 0.043	0.878 ± 0.028	0.853 ± 0.011	0.905 ± 0.048	0.267 ± 0.033	**0.736 ± 0.067**
Stacking_Ensemble	**0.799 ± 0.040**	**0.885 ± 0.025**	0.839 ± 0.012	**0.937 ± 0.047**	0.150 ± 0.062	0.731 ± 0.064
Weighted_Voting_Ensemble	0.793 ± 0.043	0.878 ± 0.028	0.853 ± 0.011	0.905 ± 0.048	0.267 ± 0.033	**0.736 ± 0.068**

Figure 10
Radar graphic for "titiz"



Analysis of the word 'ünlem'

In the classification analyses performed on the word 'ünlem' presented in Table 7 and Figure 11, the Stacking Ensemble model exhibited the highest overall performance with an accuracy of 81.4% ± 4.1% and an AUC of 0.716 ± 0.070. This result confirms the superiority of ensemble learning approaches in generalizing complex acoustic patterns for the word 'ünlem,' as observed for all previous phonetic units. The Voting Ensemble and Weighted Voting Ensemble models showed similarly high discriminative power, with AUC values of 0.726 ± 0.084 and 0.726 ± 0.083, respectively. In the model-based evaluation, the Logistic Regression algorithm performed remarkably well, with an AUC of 0.723 ± 0.085 and a specificity value of 0.633 ± 0.135. This result is similar to the values obtained for the words 'titiz' (0.710 AUC) and 'gelecek' (0.709 AUC), confirming that the Logistic Regression model can effectively discriminate voice characteristics between the asthma and control groups.

The SVM algorithm demonstrated strong classification capacity, with an accuracy of 79.4 ± 2.7% and an AUC value of 0.713 ± 0.041. This performance profile confirms that the SVM can discriminate effectively in high-dimensional acoustic feature space, as observed for the words 'titiz' (0.718 AUC) and 'gelecek' (0.643 AUC). The Random Forest model showed superior sensitivity in detecting asthma cases, with a sensitivity of 88.4 ± 2.7%, as consistently observed in all previous wording analyses. However, it exhibited a marked limitation in identifying healthy individuals, with a specificity of 0.300 ± 0.145%. This uneven performance profile suggests that the Random Forest model tends to produce a high rate of false positives in clinical applications.

Remarkably, unlike the previous phonetic units, the Decision Tree algorithm showed a relatively higher specificity for the word 'ünlem' with an AUC of 0.659 ± 0.132 and a specificity value of 0.700 ± 0.215. This suggests that decision-tree-based approaches may be advantageous in terms of specificity for words with certain phonetic features.

The K-NN algorithm exhibited limited discriminative capacity, with an accuracy of 55.8 ± 4.8% and an AUC value of 0.652 ± 0.043. This result confirms the limited effectiveness of neighborhood-based approaches on high-dimensional and complex datasets, such as voice analysis, as consistently observed in all previous wording analyses.

The results of the analysis of the word 'ünlem' show that, compared to the previous phonetic units, the ensemble models exhibit a discriminative power slightly lower than the word 'titiz' (0.731-0.736), but close to the word 'gelecek' (0.702-0.717), with AUC values in the range 0.716-0.726. This suggests that the vowel and consonant combinations contained in the word 'ünlem' can significantly reflect the effects of asthma on the voice.

The overall performance profile of the models for the word 'ünlem,' particularly high sensitivity but limited specificity in most models, emphasizes the importance of balancing sensitivity and specificity in the diagnosis of asthma. This balance requires the development of optimized models that can detect a high proportion of true asthma cases while minimizing false-positives in clinical practice.

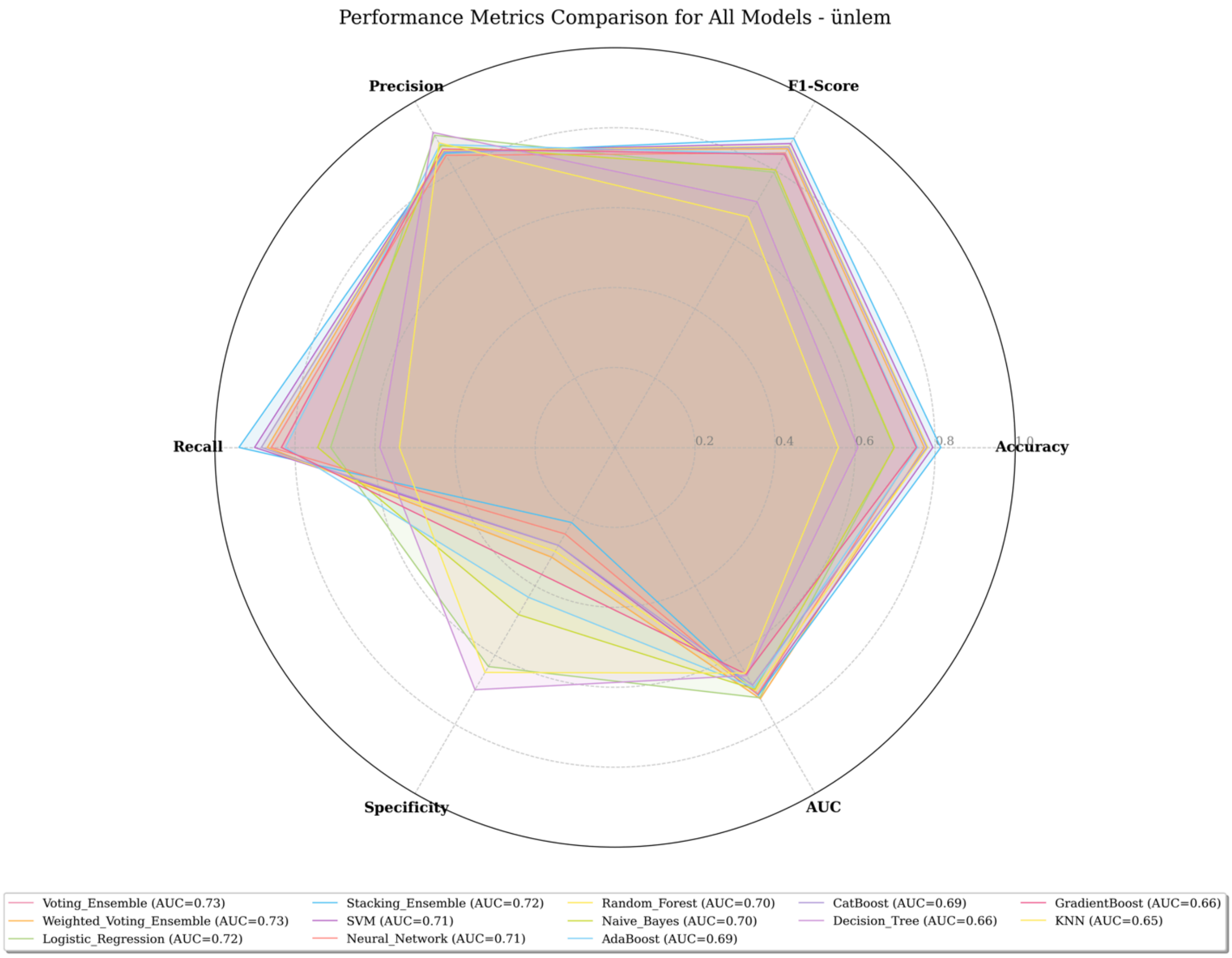
Table 7

Machine learning model performance evaluation metrics for the sound "ünlem"

Model	Accuracy	F1-Score	Precision	Recall	Specificity	AUC
Naive_Bayes	0.698±0.041	0.802 ± 0.027	0.872 ± 0.030	0.743 ± 0.027	0.483 ± 0.122	0.699 ± 0.085
KNN	0.558 ± 0.048	0.666 ± 0.048	0.880 ± 0.027	0.539 ± 0.064	0.650 ± 0.097	0.652 ± 0.043
Decision_Tree	0.607 ± 0.042	0.710 ± 0.040	**0.910 ± 0.062**	0.588 ± 0.063	**0.700 ± 0.215**	0.659 ± 0.132
Neural_Network	0.753 ± 0.030	0.851 ± 0.020	0.844 ± 0.016	0.859 ± 0.033	0.250 ± 0.091	**0.713 ± 0.063**
CatBoost	0.779 ± 0.042	0.868 ± 0.027	0.854 ± 0.023	0.884 ± 0.046	0.283 ± 0.125	0.687 ± 0.072
AdaBoost	0.756 ± 0.036	0.848 ± 0.021	0.875 ± 0.037	0.824 ± 0.019	0.433 ± 0.193	0.694 ± 0.085
GradientBoost	0.753 ± 0.055	0.847 ± 0.035	0.862 ± 0.036	0.834 ± 0.046	0.367 ± 0.172	0.655 ± 0.077
Random_Forest	0.782 ± 0.019	0.870 ± 0.012	0.858 ± 0.023	0.884 ± 0.027	0.300 ± 0.145	0.703 ± 0.087
SVM	**0.794 ± 0.027**	**0.878 ± 0.017**	0.856 ± 0.012	**0.901 ± 0.029**	0.283 ± 0.067	**0.713 ± 0.041**
Logistic_Regression	0.697 ± 0.041	0.795 ± 0.029	0.902 ± 0.036	0.711 ± 0.032	0.633 ± 0.135	0.723 ± 0.085
Voting_Ensemble	0.773 ± 0.043	0.863 ± 0.031	0.857 ± 0.010	0.869 ± 0.052	0.317 ± 0.033	0.726 ± 0.084
Stacking_Ensemble	0.814 ± 0.041	0.893 ± 0.025	0.850 ± 0.017	0.940 ± 0.037	0.217 ± 0.085	0.716 ± 0.070
Weighted_Voting_Ensemble	0.773 ± 0.043	0.863 ± 0.031	0.857 ± 0.010	0.869 ± 0.052	0.317 ± 0.033	0.726 ± 0.083

Figure 11

Radar graphic of "ünlem"



Discussion

In this study, the performance of 13 machine learning models for asthma diagnosis using different sound samples ('aaa, ' 'ana,' 'araba,' 'gelecek, ' 'ordu, ' 'titiz,' 'ünlem') was evaluated using Accuracy, F1-Score, Precision, Recall, Specificity and AUC (area under the curve) metrics. The findings revealed that ensemble models (Stacking Ensemble, Voting Ensemble, Weighted Voting Ensemble) generally demonstrate superior performance. For example, the Stacking Ensemble model achieved the best results with 82.3% accuracy and 0.684 AUC for the word 'araba' and 82.0% accuracy and 0.702 AUC for the word 'gelecek.' This result confirms that the ensemble models can better generalize complex patterns. In addition, complex models such as neural networks and random forests have yielded remarkable results. In particular, the neural network model demonstrated high sensitivity with 90.1% recall rates for the words 'araba' and 'titiz' respectively. However, the low specificity values of these models resulted in limitations in practical applications due to the increase in false-positive rates. Models such as SVM and Logistic Regression also performed well, with generally high precision and AUC values. For example, the SVM model ranked second for the word 'gelecek, ' with an AUC value of 0.643. In contrast, simpler models, such as the naive Bayes and K-NN models, demonstrated limited performance owing to low accuracy and specificity values. In particular, K-NN had low values, such as 59.6% accuracy and an AUC of 0.606 for the word 'gelecek.' This indicates that they are inadequate for capturing complex patterns in audio data. In conclusion, the Stacking Ensemble model stood out as the most promising asthma diagnostic model, with high recall and AUC values.

However, the low specificity values of some models emphasize the need to control false-positive rates. Furthermore, the use of MFCC and additional acoustic features confirmed that MFCC is an effective feature extraction method for asthma diagnosis. On the other hand, the fact that deep learning techniques could not be evaluated in this study owing to the limited dataset provides an opportunity for future research.

Conclusions

In this study, the effectiveness of voice analysis-based machine learning models for asthma diagnosis was comprehensively evaluated. Within the scope of the study, a total of 13 different machine learning models were developed for different sound samples ('aaa, ' 'ana,' 'araba,' 'gelecek, ' 'ordu, ' 'titiz,' 'ünlem'), and their performances were compared. The findings showed that ensemble learning approaches (ensemble models) exhibit superior performance in asthma diagnosis.

The stacking ensemble model obtained the best overall performance. In particular, it achieved high AUC values (68.4% and 70.2%) for sound examples such as 'araba' and 'gelecek.' This result confirms that ensemble models can better generalize complex sound patterns. The Voting Ensemble and Weighted Voting Ensemble models showed similar performances; however, low specificity values were identified as a limitation to be considered in practical applications.

Complex models, such as Neural Networks and Gradient Boosting, have also demonstrated remarkable results. In particular, the Neural Network model provided high sensitivity with recall rates of 90.1% for the 'araba' and 'titiz' sound samples. However, the low specificity of these models should be considered a factor that may increase false-positive results in clinical applications.

The use of Mel-Frequency Cepstral Coefficients (MFCCs) and additional acoustic features (spectral center, spectral bandwidth, etc.) was confirmed to be an effective feature extraction method for asthma diagnosis.

The important limitations of our study include the limited and single-center dataset of 284 patients with asthma diagnosed using the GINA criteria and 60 healthy volunteers. This may have affected the generalizability of the models to larger and more diverse populations. Nevertheless, this study demonstrates the potential of voice analysis-based approaches as cost-effective screening tools for early detection, especially in areas with limited access to health services. The high-sensitivity values of the community models support this potential.

This study has important implications for clinical practice. In particular, the high sensitivity values of the community models suggest that they can be used as asthma screening tools. A voice analysis-based diagnostic approach may provide a cost-effective alternative for early diagnosis, particularly in areas with limited access to healthcare services. It may also enable remote and continuous monitoring of asthma symptoms in patients' daily lives through smartphone applications.

Several directions for future research are identified. First, validating the models in multicenter, larger, and more diverse populations (differing in age, ethnicity, and asthma phenotypes) is critical for enhancing generalizability, a need also highlighted by community-based studies, such as Gunawardana et al. (2024). Second, expanding the dataset could further improve model performance by enabling the application of deep learning techniques (e.g., CNN, RNN, Transformer architectures), as studies such as Chen et al. (2025)

demonstrated the potential of DNNs in related tasks, albeit with different datasets and objectives. Adapting existing models to other respiratory diseases (e.g., COPD, pneumonia, and bronchitis) using a transfer learning approach has emerged as a valuable research area.

Data Sharing: Contact the authors within three years of the date of publication to request the datasets used in the study. Data sharing will be performed while maintaining participant confidentiality.

Ethics Committee Approval: Ethics committee approval was received for this study from the ethics committee of İstanbul Yedikule Chest Diseases and Thoracic Surgery Training and Research Hospital (Date: 14.12.2023, No: 2023-433).

Informed Consent: Written informed consent was obtained from all participants who participated in this study.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- M.G., M.A.U.; Data Acquisition- N.A., C.O., E.Y.N.; Data Analysis/Interpretation- M.G.; Drafting Manuscript- M.G., M.A.U., N.A., C.O., E.Y.N.; Critical Revision of Manuscript- M.G.; Final Approval and Accountability- M.G., M.A.U., N.A., C.O., E.Y.N.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: This study was carried out with the financial support of MedCase Yazılım Teknolojileri A.Ş, which was used for data collection processes. The sponsoring organisation did not influence the design of the study, data analysis, interpretation of the results, or writing of the manuscript.

Acknowledgements: The author(s) would like to thank Oğuz Akbaş (MonitorCRO) for his invaluable contributions and support in the scientific writing of this article.

Author Details

Murat Gezer

[1] İstanbul University, Faculty of Science, Department of Computer Science, İstanbul, Türkiye

0000-0002-7286-3943 murat.gezer@istanbul.edu.tr

Mehmet Atilla Uysal

[2] University of Health Sciences Türkiye, Yedikule Hospital for Chest Diseases and Thoracic Surgery, Department of Chest Diseases, İstanbul, Türkiye

0000-0002-0430-498X

Neval Alagöz

[2] University of Health Sciences Türkiye, Yedikule Hospital for Chest Diseases and Thoracic Surgery, Department of Chest Diseases, İstanbul, Türkiye

0000-0002-4203-4562

Can Ortak

[3] Medcase Corp, İstanbul, Türkiye

0009-0001-3299-5562

Elif Yelda Niksarlıoğlu

[2] University of Health Sciences Türkiye, Yedikule Hospital for Chest Diseases and Thoracic Surgery, Department of Chest Diseases, İstanbul, Türkiye

0000-0002-6119-6540

References

Alam, M. Z., Simonetti, A., Brillantino, R., Tayler, N., Grainge, C., Siribaddana, P., Nouraei, S. A. R., Batchelor, J., Rahman, M. S., Mancuzo, E. V., Holloway, J. W., Holloway, J. A., & Rezwan, F. I. (2022). Predicting pulmonary function from the analysis of voice: A machine learning approach. *Frontiers in Digital Health*, 4, 750226. https://doi.org/10.3389/fdgth.2022.750226

Aykanat, M., Kılıç, Ö., Kurt, B., & Saryal, S. B. (2020). Lung disease classification using machine learning algorithms. *International Journal of Applied Mathematics Electronics and Computers, 8*(4), 125–132. https://doi.org/10.18100/ijamec.799363

Barnes, P. J. (2008). Immunology of asthma and chronic obstructive pulmonary disease. *Nature Reviews Immunology, 8*(3), 183–192. https://doi.org/10.1038/nri2254

Bolat, E. (2021). *Use of machine learning methods in the classification of respiratory system diseases* [Unpublished doctoral dissertation]. Istanbul University Institute of Health Sciences.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Chen, M.-H., Lee, G., & Hung, L.-P. (2025). AI-driven data analysis for asthma risk prediction. *Healthcare, 13*, 774. https://doi.org/10.3390/healthcare13070774

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 28*(4), 357–366. https://doi.org/10.1109/TASSP.1980.1163420

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Springer. https://doi.org/10.1007/3-540-45014-9_1

Fesci, H., & Görgülü, Ü. (2005). Asthma and life. *Journal of School of Nursing, 1*(1), 77–83. https://dergipark.org.tr/tr/download/article-file/88607

Ganaie, M. A., Hu, M., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence, 115*, 105151. https://doi.org/10.1016/j.engappai.2022.105151

Global Initiative for Asthma. (2023). *Global strategy for asthma management and prevention*. https://ginasthma.org/2023-gina-main-report/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gunawardana, J., Viswakula, S., Rannan-Eliya, R. P., & Wijemunige, N. (2024). Machine learning approaches for asthma disease prediction among adults in Sri Lanka. *Health Informatics Journal, 30*(3), 14604582241283968. https://doi.org/10.1177/14604582241283968

Holgate, S. T., Wenzel, S., Postma, D. S., Weiss, S. T., Renz, H., & Sly, P. D. (2015). Asthma. *Nature Reviews Disease Primers, 1*, 15025. https://doi.org/10.1038/nrdp.2015.25

Kaplan, A., Cao, H., FitzGerald, J. M., Iannotti, N., Yang, E., Kocks, J. W. H., Kostikas, K., Price, D., Reddel, H. K., Tsiligianni, I., Vogelmeier, C. F., Pfister, P., & Mastoridis, P. (2021). Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *The Journal of Allergy and Clinical Immunology: In Practice, 9*(6), 2255–2261. https://doi.org/10.1016/j.jaip.2021.02.014

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (Vol. 2, pp. 1137–1143).

Lambrecht, B. N., & Hammad, H. (2015). The immunology of asthma. *Nature Immunology, 16*(1), 45–56. https://doi.org/10.1038/ni.3049

Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In *Proceedings of the 1st International Symposium on Music Information Retrieval* (pp. 1–13). https://doi.org/10.5281/zenodo.1416444

Malmberg, L. P., Sovijärvi, A. R. A., Paajanen, E., Piirilä, P., Haahtela, T., & Katila, T. (1994). Changes in frequency spectra of breath sounds during histamine challenge test in adult asthmatics and healthy control subjects. *Chest, 105*(1), 122–131. https://doi.org/10.1378/chest.105.1.122

Mayr, W., Triantafyllopoulos, A., Batliner, A., Schuller, B. W., & Berghaus, T. M. (2025). Assessing the clinical and functional status of COPD patients using speech analysis during and after exacerbation. *International Journal of Chronic Obstructive Pulmonary Disease, 20*, 137–147.

Ministry of Health of the Republic of Turkey, General Directorate of Public Health. (2024, August 1). *World Asthma Day May 7, 2024*. https://hsgm.saglik.gov.tr/tr/haberler/7-mayis-2024-duenya-astim-guenue.html

Petmezas, G., Cheimariotis, G.-A., Stefanopoulos, L., Rocha, B., Paiva, R. P., Katsaggelos, A. K., & Maglaveras, N. (2022). Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function. *Sensors, 22*(3), 1232. https://doi.org/10.3390/s22031232

Rabiner, L. R., & Schafer, R. W. (2007). *Introduction to digital speech processing*. Now Publishers. https://doi.org/10.1561/2000000001

Reichert, S., Gass, R., Brandt, C., & Andrès, E. (2008). Analysis of respiratory sounds: State of the art. *Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine, 2*, 45–58. https://doi.org/10.4137/CCRPM.S530

Rivas-Navarrete, J. A., Pérez-Espinosa, H., Padilla-Ortiz, A. L., Rodríguez-González, A. Y., & García-Cambero, D. C. (2025). Edge computing system for automatic detection of chronic respiratory diseases using audio analysis. *Journal of Medical Systems, 49*, 33. https://doi.org/10.1007/s10916-025-02154-7

Saygılı, M. (2019). Mobile health. In M. S. Yıldız (Ed.), *Advanced technology applications in healthcare* (pp. 211–238). Nobel Academic Publishing.

Schreur, H. J. W., Vanderschoot, J., Zwinderman, A. H., Dijkman, J. H., & Sterk, P. J. (1994). Abnormal lung sounds in patients with asthma during episodes with normal lung function. *Chest, 106*(1), 91–99. https://doi.org/10.1378/chest.106.1.91

Sezer, E., & Akıl, S. (2020). Criteria used by speech-language pathologists in Turkey for diagnosing childhood apraxia of speech. *Journal of Language, Speech and Swallowing Research, 3*(2), 153–174.

Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2021). Deep learning applications for COVID-19. *Journal of Big Data, 8*, 18. https://doi.org/10.1186/s40537-020-00392-9

Sterling, M., Rhee, H., & Bocko, M. (2014). Automated cough assessment on a mobile platform. *Journal of Medical Engineering, 2014*, 951621. https://doi.org/10.1155/2014/951621

Topaz, M., Zolnoori, M., Norful, A. A., Perrier, A., Kostic, Z., & George, M. (2022). Speech recognition can help evaluate shared decision making and predict medication adherence in primary care setting. *PLOS ONE, 17*(8), e0271884. https://doi.org/10.1371/journal.pone.0271884

World Health Organization. (2024, May 6). *Asthma*. https://www.who.int/news-room/fact-sheets/detail/asthma

Yadav, S., Keerthana, M., Gope, D., Maheswari, U. K., & Ghosh, P. K. (2020). Analysis of acoustic features for speech sound-based classification of asthmatic and healthy subjects. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6789–6793). IEEE. https://doi.org/10.1109/ICASSP40776.2020.9054062

Yadav, S., Nk, K., Gope, D., Krishnaswamy, U. M., & Ghosh, P. K. (2018). Comparison of cough, wheeze and sustained phonations for automatic classification between healthy subjects and asthmatic patients. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1400–1403).

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., & Moore, G. (2006). *The HTK book*. Cambridge University Engineering Department.

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM, 64*(3), 107–115. https://doi.org/10.1145/3446776

Zhang, Y., Huang, Q., Sun, W., Chen, F., Lin, D., & Chen, F. (2024). Research on lung sound classification model based on dual-channel CNN-LSTM algorithm. *Biomedical Signal Processing and Control, 94*, 106257. https://doi.org/10.1016/j.bspc.2024.106257