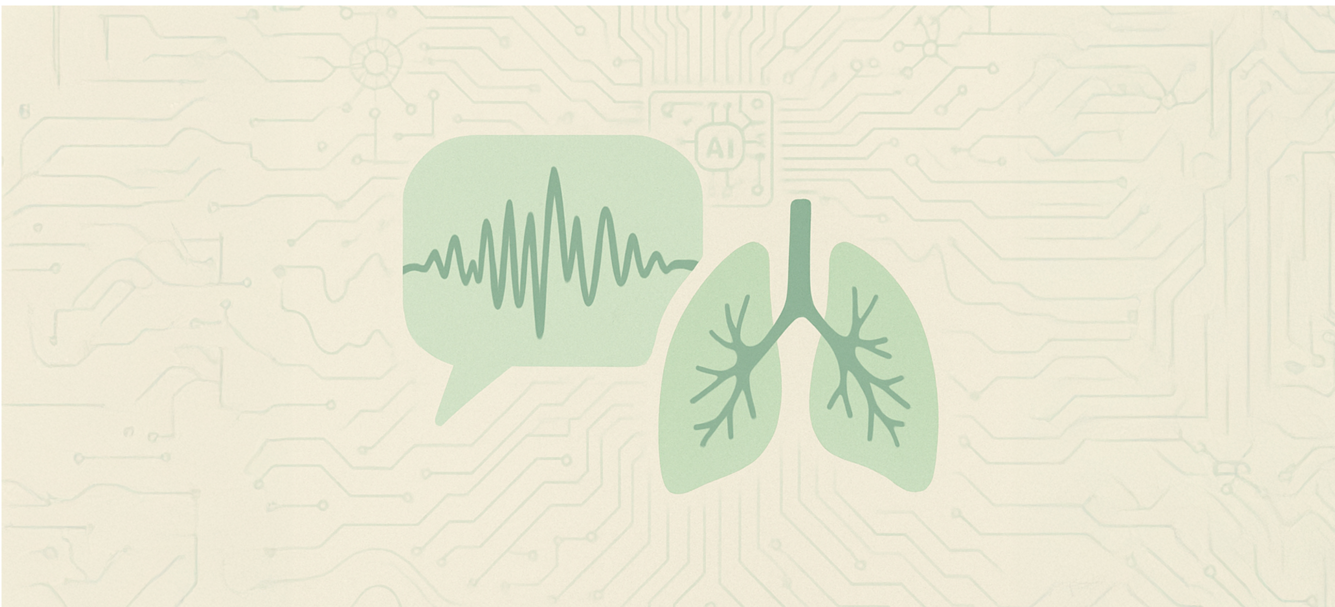
Detecting COPD Through Speech Analysis: A Dataset and Machine Learning Approach

Cuno Sankey-Olsen
Aalborg University
Department of Computer Science
Aalborg, Denmark
csanke19@student.aau.dk

Rasmus Hvass Olesen
Aalborg University
Department of Computer Science
Aalborg, Denmark
rolese20@student.aau.dk

Tobias Oliver Eberhard
Aalborg University
Department of Computer Science
Aalborg, Denmark
teberh23@student.aau.dk



ABSTRACT

This study explores the application of voice recordings and machine learning to support early detection of Chronic Obstructive Pulmonary Disease (COPD). Audio data were collected from 96 participants through three vocal tasks, and features were extracted using openSMILE and SpeechBrain. Four models were tested across multiple data configurations. Results show that SVM and Random Forest models performed consistently well, especially with openSMILE features. While limitations include reliance on self-reported diagnoses and inconsistent task execution, the findings suggest that voice-based analysis has potential as a non-invasive, scalable screening tool for COPD.

KEYWORDS

Chronic Obstructive Pulmonary Disease, COPD, Artificial Intelligence, AI, Machine Learning, ML, Detection, Voice Analysis, Speech Analysis, Signs of COPD, Digital Health, Voice-based diagnosis, Smartphone-based, Data Collection, Speech Features

1 INTRODUCTION

Recent advancements in mobile health and artificial intelligence (AI) have demonstrated the growing potential of machine learning in supporting medical diagnostics and empowering patients beyond traditional clinical environments. Mobile-first technologies and wearable devices are increasingly being used for private, in-home health assessments and pre-clinical consultations [7, 28, 41]. With the combination of sophisticated machine learning techniques and access to large-scale medical datasets, AI-based systems have achieved significant success in fields such as radiology, cardiology, and disease prediction [2].

One area that stands to benefit significantly from these technological advancements is the diagnosis of respiratory diseases, particularly Chronic Obstructive Pulmonary Disease (COPD). Affecting over 390 million people globally, COPD is a progressive and frequently underdiagnosed condition that severely impacts patients' quality of life. Research indicates that up to 80% of COPD cases remain undiagnosed, with many individuals experiencing symptoms for years before receiving a clinical diagnosis [51]. This trend is also reflected locally in Denmark, where the present study was conducted. According to the director of The Lung Association of Denmark (Lungeforeningen), approximately 400,000 people in Denmark are estimated to have COPD, but only about half have received an official diagnosis – largely because many are unaware they have the disease [40]. While relatively few studies have explored non-invasive approaches to identifying COPD, a notable exception is the work by Batliner et al. at the Technical University of Munich (TUM), who used speech analysis during COPD exacerbations to predict clinical status with an accuracy of 84% [52].

Recent advances in biomedical audio analysis indicate that biometric signals, such as speech may contain valuable markers of respiratory health. Subtle variations in vocal characteristics – such as airflow, pitch, and voice intensity – can serve as non-invasive indicators of underlying pulmonary conditions [29]. Leveraging machine learning techniques, these acoustic features can be systematically extracted and analyzed to support the detection of diseases like COPD – potentially even before symptoms become clinically apparent. This approach aligns with broader trends in mobile health and passive sensing technologies, which aim to extend diagnostic capabilities beyond traditional healthcare settings and into everyday environments [2].

1.1 Research Question

Building on these advancements, this study investigates the potential of machine learning models to detect COPD using voice recordings. The aim is to evaluate whether acoustic features extracted from speech can be used to distinguish individuals with COPD from those without. Specifically, the study poses the following research question:

> *(How) can COPD be detected through voice recordings captured with common mobile devices using machine learning techniques?*

Furthermore, this study investigates whether and how metadata such as age and gender, in combination with different machine learning models and audio feature extraction toolkits, influence the accuracy and performance of COPD detection across various demographic groups.

2 BACKGROUND

This section provides an overview of COPD, including its progression, impact on quality of life, and challenges in diagnosis. Furthermore, it explores the role of machine learning in medicine, with a specific focus on its applications in speech analysis for COPD status.

2.1 COPD in the Medical Context

Chronic Obstructive Pulmonary Disease (COPD) is a progressive respiratory disorder characterized by limited airflow function, primarily caused by chronic bronchitis and damaged alveoli, called emphysema. It affects over 390 million people globally and caused around 3.5 million deaths in 2021, which were roughly 5% of all deaths globally [51]. COPD is a major public health challenge, with mortality rates expected to rise due to aging populations and persistent exposure to risk factors such as smoking tobacco, air pollution levels, and occupational hazards [11].

2.1.1 ***COPD Stages and the Progression of the Disease.*** COPD is not a static condition but rather a spectrum of disease that is commonly classified into several stages according to the severity of airflow limitation [18]. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines divide COPD into four stages: Mild (GOLD 1), moderate (GOLD 2), severe (GOLD 3), and very severe (GOLD 4). COPD may be almost asymptomatic in the early stage, with patients experiencing only subtle symptoms, such as mild shortness of breath during exertion or the occasional cough [50]. As the condition progresses to advanced stages, the limitation of airflow becomes increasingly pronounced. People with moderate to severe disease commonly have a chronic cough, greater volume of sputum expectoration, and shortness of breath that considerably restricts their ability to function normally. This gradual decline illustrates the progressive component of COPD and shows why the initial signs are so frequently ignored or attributed to becoming older or being unfit [18].

2.1.2 ***Living with COPD and Quality of Life.*** COPD can affect people's lives beyond just a decline in lung function that can be measured. Due to intermittent symptoms in the early stages, many people do not recognize the progression of their condition. Later stages are often associated with substantial physical limitations that negatively affect people's activities of daily living, work, exercise, and socializing [3]. Patients may report chronic breathlessness, fatigue, and frequent respiratory infections, all of which have detrimental effects on physical activity and may lead to psychological comorbidities such as anxiety and depression [18]. For most people, living with COPD means a life of ongoing adjustment to changing abilities – modifying daily activities, depending on oxygen therapy, or attending pulmonary rehabilitation programs to preserve the quality of life. The complex relationship between physical symptoms and emotional status underscores the need for comprehensive care that responds to both medical and psychosocial concerns [3].

2.1.3 ***The Burden of Undiagnosed COPD.*** As previously presented in this paper, COPD is a major global health burden, with approximately 60–86% of cases remaining undiagnosed [21]. Undiagnosed COPD is linked to poor clinical outcomes, including significantly higher rates of exacerbations, hospitalization, respiratory-related mortality, and impaired quality of life. Delayed diagnosis also accelerates disease progression and the likelihood of increased comorbid conditions [35]. This source further describes that detection is critical since the treatment of mild-to-moderate COPD, the stage that most individuals have before diagnosis, can slow the progression of the disease. Quitting smoking, medication, and lung rehabilitation have revealed evident benefits in preserving lung function, minimizing exacerbations, and enhancing quality of life. Nevertheless, the paper highlights issues, such as insufficient use of spirometry, overlooking early symptoms, and overemphasis on tobacco hazards, which hinder the timely diagnosis of patients. Guidelines now recommend targeted case-finding in high-risk individuals (e.g., adults over the age of 35–40 years with respiratory symptoms, smoking history, or environmental exposure). Technological advances in portable spirometry devices and validated risk assessment questionnaires maximize detection efficiency but are limited by their high cost and low availability. Proactive case-finding strategies, coupled with risk factor evaluation and testing, are required to reduce undiagnosed COPD, improve patient outcomes, and alleviate pressures on healthcare systems [35].

2.1.4 ***Under- and Overdiagnosis of COPD.***

COPD is a common condition yet is often misdiagnosed with both under- and overdiagnosis that can lead to challenges to patient care and the healthcare systems. An underdiagnosis occurs when a patient living with COPD has not yet been identified with the disease [14]. In many cases, individuals may live with COPD for years before receiving a diagnosis, during which their condition can progress untreated [24]. A significant factor contributing to underdiagnosis is the failure to perform spirometry, which is used to confirm irreversible airflow obstruction and is a key characteristic of COPD [23].

Another concern related to the often delayed diagnosis and treatment of COPD is the stigma associated with the disease, largely due to its strong link to smoking. Individuals with COPD are frequently assumed to be smokers and may experience stigmatization from others as a result [53]. This perception frames COPD as a preventable and self-inflicted condition, which can lead to feelings of guilt or self-blame among those affected. Such stigma has the potential to negatively influence both the willingness of individuals to seek medical attention and the quality of care they receive, ultimately affecting the diagnosis and treatment of the disease [53].

An overdiagnosis of COPD happens when a patient is misdiagnosed with COPD when, in reality, they have a different disease or health complication [23]. Overdiagnosis occurs when a spirometry test is not performed, and healthcare personnel diagnose the patient based on other tests or symptoms. Overdiagnosis often occurs in older patients, whose natural age-related decline in lung function may be misinterpreted as COPD. Additionally, comorbidities like asthma or bronchiectasis can mimic COPD symptoms, leading to misdiagnosis [23].

2.2 Machine Learning in Medicine

Machine learning algorithms leverage statistical methods and computational techniques to identify patterns in large datasets, including text, numerical data, images, and audio. Machine Learning underpins various modern applications, such as search engines, recommendation systems, and AI chatbots [22], as well as numerous advancements in medical diagnostics and treatment. Advances in processing power, memory, and storage have enabled computers to analyze and identify patterns in vast amounts of medical data –analyses that would have been infeasible just a few years ago [13]. An example of machine learning in this field is the automation of electrocardiogram (EKG) interpretation by cardiologists. Machine learning algorithms analyze EKG test results, identifying patterns to assist in diagnosing cardiac conditions [13]. In radiology, machine learning is used for the automated detection of lung nodules in chest X-rays. Both this and automated electrocardiogram interpretation illustrate how technology can assist trained medical professionals. In these cases, the computer approximates the diagnostic capabilities of a physician with high accuracy, but at a significantly faster pace and fewer resources [13].

2.3 Machine Learning & Speech Analysis

Speech Analysis utilizes speech and voice characteristics to extract and classify audio signals [6]. This is achieved through feature extraction techniques such as energy, jitter, and spectral features – including flux, roll-off, and centroid – among many others. The extracted features are then stored as structured data, enabling their use in training machine learning models [5].

2.3.1 ***The Role of Speech Analysis in Medicine.***

Speech Analysis has gained interest in the medical field, as paralinguistic analysis can assess a wide range of health conditions due to the complexity of speech production in relation to overall health [39]. This includes both neurological and respiratory health, as even slight changes in these areas can affect a person's ability to control the vocal apparatus, thereby altering acoustic properties [29]. The unique insights derived from pathological speech changes, combined with the ease of collecting, storing, and programmatically analyzing speech data with AI [48], have contributed to the growing role of speech analysis in medicine. This trend is further reflected in the work of Milling et al., where AI-based speech analysis is emerging as a powerful tool for detecting a wide range of diseases – potentially becoming as integral to medical diagnostics as blood tests [36].

2.3.2 ***Speech Analysis in COPD.***

With the increasing application of AI technology in COPD management, modern tools are frequently used for diagnosis, treatment, and post-diagnosis care. However, speech analysis particularly in the early diagnosis of COPD, remains relatively underutilized [25]. One of the most notable studies in this field comes from the Technical University of Munich (TUM), where researchers have explored the use of speech analysis to assess the clinical and functional status of participants with COPD during and after exacerbation [52]. In this study, scientists programmatically analyzed voice recordings of participants with COPD both upon admission for critical treatment and after discharge. The research led to the development of a machine learning model capable of predicting patient readiness for discharge with 84% accuracy, based on data from 50 participants with COPD in the category of GOLD 2, GOLD 3, and GOLD 4 [52].

3 METHODS

This section outlines the methodological approach taken to collect and analyze data for the purpose of developing a machine learning model capable of detecting COPD based on voice recordings. It describes the procedures used for participant recruitment, data collection, audio processing, and complementary qualitative interviews. This process was meant to ensure both technical consistency and ethical compliance while capturing a diverse range of phonatory and respiratory characteristics from individuals with and without COPD. The following subsections provide an overview of the dataset, recording protocols, and interview techniques employed during the study.

3.1 Data Collection

The primary dataset in this study consists of audio recordings from human participants, including both individuals diagnosed with COPD and a control group without any known respiratory conditions. Recordings were collected using microphones on smartphones (iPhone 13 Pro, iPhone 12 Pro, iPhone 10) in quiet, controlled environments. See Figure 2.

Participants were primarily recruited in person through visits to local activity centers, lung cafés, lung choirs, and COPD-specific fitness groups in Aalborg and Aarhus. A smaller number of participants were recruited online through COPD-related Facebook groups, Google Forms, and E-mail. Recordings were conducted either in person or remotely by the participants themselves. In total, data was collected from 48 participants with varying stages of COPD (GOLD 1 to 4) and 48 participants in the control group, all aged between 26 and 88 years, with each group having a median age of 72. Details are further described in Table 1 and Figure 1.

Statistic	COPD	Non-COPD
COPD GOLD Stages		
GOLD 1	14 (29.2%)	-
GOLD 2	14 (29.2%)	-
GOLD 3	14 (29.2%)	-
GOLD 4	1 (2.1%)	-
GOLD unknown	5 (10.4%)	-
Gender Distribution		
Females	41 (age: 49-85)	36 (age: 34-88)
Males	7 (age: 64-74)	12 (age: 26-83)
Total	**48**	**48**
Median Age	**72**	**72**

Table 1: COPD and Non-COPD Statistics

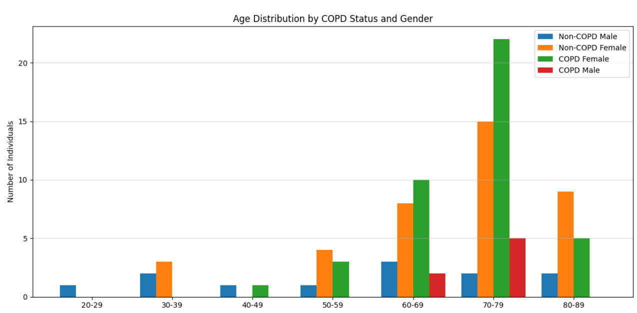


Figure 1: Age Profile of COPD Participants

Each participant was asked to perform three specific vocal tasks designed to capture a broad spectrum of respiratory and phonatory characteristics.

(1) Sustained vowel pronunciation: Participants continuously pronounced a sequence of Danish vowels – "A, E, O, Ø, and Å" – to assess vocal steadiness and airflow control.
(2) Reading aloud: Participants read the short fable The North Wind and the Sun aloud. This test was selected for its standardized structure.
(3) Coughing: Participants were instructed to cough three times into the microphone to capture forced expiratory sounds typically affected by respiratory illness.

This data protocol was adapted by Batilener et al. [52].

3.2 Dataset Preparation

The average duration of each participant session was approximately 1–2 minutes. Recordings were manually segmented into individual audio files for each task and stored in WAV format at a sample rate of 44.1 kHz. The only preprocessing method used was silence trimming, using a Python library called "pydub". In addition, Adobe Audition was used to reduce background noise in a few audio recordings where excessive noise was present. File naming conventions and metadata, such as age, gender, and COPD status, were stored in a structured format to support model training. This structure also included additional, unused metadata, such as the COPD stage and remarks or comments made during the recording sessions. These supplementary data points could be leveraged in future work to train models with alternative focuses or to support more fine-grained analyses. An example can be seen in snippet 1 of the JSON metadata file.

```
{
"name": "00012",
"silence_thresh": -40,
"age": 72,
"gender": "M",
"COPD_status": 1,
"COPD_GOLD": 3,
"comment": "This person
could not read the script,
so he told a short story instead"
}
```

Listing 1: Metadata for Audio Recordings

3.3 Interview Data

To complement the quantitative audio data, short, informal, semi-structured interviews were conducted with a subset of participants after they had completed their recordings. These followed the qualitative methodology of Brinkmann and Kvale, using an interview guide designed to elicit reflections on respiratory health, recent symptoms, and participants' experiences with the recording process [8]. Given the participant demographic, the interviews were conducted in a conversational and approachable manner to promote comfort and openness [47]. All interviews were audio-recorded and anonymized in accordance with GDPR and institutional ethical guidelines.

Although the interview data was not used to train the machine learning models, it played an important role in validating the interpretation of the audio recordings and enriched the overall dataset. The conversations also confirmed several assumptions drawn from existing literature and consultations with medical professionals – particularly the insight that many individuals may unknowingly live with COPD for years before receiving a formal diagnosis. Additionally, the interviews revealed a generally positive attitude among participants, with many expressing appreciation for contributing to research that has the potential to benefit future generations of patients.

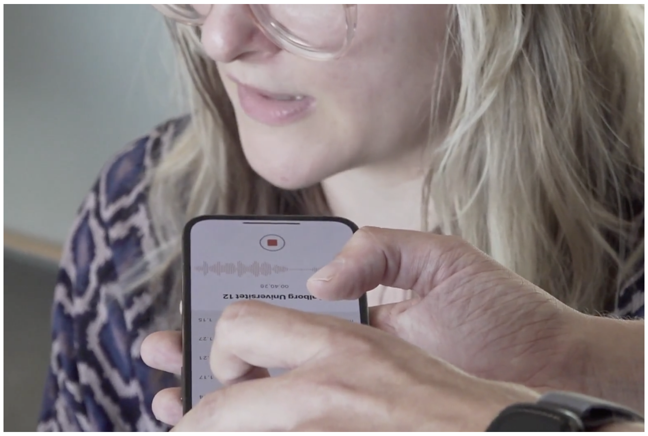


Figure 2: An example of an audio recording

3.4 Data Processing

To evaluate robustness, generalizability, and overall performance, the data processing pipeline incorporated two different toolkits for audio feature extraction: openSMILE and SpeechBrain. Additionally, four distinct machine learning models were tested: Random Forest, Support Vector Machine (SVM), Logistic Regression, and a Neural Network. Various configurations were explored to better understand model behavior and generalizability, including models trained exclusively on female participants, models excluding age metadata, and models including both male and female data. This comparative approach was designed to identify potential biases and assess the influence of demographic features on prediction accuracy. Model evaluation was carried out using GridSearchCV in combination with StratifiedGroupKFold cross-validation, enabling a thorough comparison of average precision, recall, and F1-score across different hyperparameter settings and data splits [32, 33]. A visualization of this process is shown in figure 3.

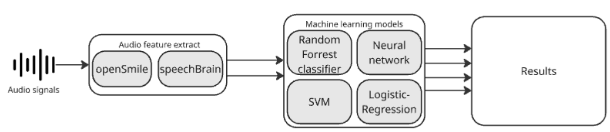


Figure 3: A general overview of the COPD prediction workflow

3.4.1 *Audio Feature Extraction.*

After collecting audio recordings from both COPD and non-COPD participants, two open-source toolkits — SpeechBrain and openSMILE — were used to extract voice features potentially indicative of respiratory illness [6, 43]. SpeechBrain is a deep learning-based speech processing toolkit built on PyTorch. In this study, speaker embeddings were generated, which are dense vector representations that capture subtle vocal characteristics, such as articulation, prosody, and vocal tract dynamics [42]. These embeddings were produced using pre-trained models trained on the VoxCeleb1 and VoxCeleb2 datasets, which include over one million utterances from thousands of speakers in diverse real-world conditions [38]. In the context of COPD, these embeddings may reflect subtle respiratory-related changes in speech patterns, such as:

- Shortness of breath affecting phrasing and intonation
- Altered speech rhythm due to reduced airflow
- Weakened voice intensity

OpenSMILE, on the other hand, offers a more traditional feature extraction approach. Unlike SpeechBrain, it does not rely on deep learning or pre-trained models. Instead, it extracts a wide range of well-documented acoustic features directly from the audio signal. These include:

- Jitter and shimmer, which reflect vocal stability and are often affected by respiratory strain
- Energy and pitch variation, which may be reduced in individuals with impaired breath control
- Speech rate and pause patterns, which can change due to respiratory fatigue

While openSMILE provides less abstract representations than SpeechBrain, its features are highly interpretable and commonly used in speech research, demonstrations, and prototyping applications [6]. The features extracted by both toolkits were used as input to various machine learning models to evaluate whether vocal patterns could be used to distinguish between individuals with and without COPD.

3.4.2 *Models.*

Given the manual labeling of audio data, a supervised learning approach is suitable. Supervised models learn to map input features to known output labels, enabling predictions on new, unseen data. We employed four common supervised models: Logistic Regression, Random Forest, Support Vector Machine (SVM), and Neural Networks – each offering different strengths for this task [49].

Logistic Regression is a simple linear model ideal for binary classification. It is efficient and interpretable, making it a good baseline for distinguishing between COPD and non-COPD cases [30].

Random Forest is an ensemble model that combines multiple decision trees to enhance robustness and mitigate overfitting. It handles noisy data well and can capture more complex feature interactions [27].

The Support Vector Machine (SVM) excels at separating classes with a clear margin and performs well in high-dimensional feature spaces. With kernel functions, it can also capture non-linear patterns in the audio data [26].

Neural Networks are well-suited for modeling complex, nonlinear relationships. They can detect subtle patterns in voice and cough features but require more tuning and computational resources [20, 44].

Each model was implemented using Scikit-learn or TensorFlow, with training and evaluation performed using an 80/20 train-test split and StratifiedGroupKFold cross-validation to ensure reproducibility and account for group structure in the data [45, 46].

Table 2: Performance Metrics for COPD Prediction Models Across Datasets and Feature Sets (0 = non-COPD, 1 = COPD)

Model	Dataset	Precision (0)	Recall (0)	F1 (0)	Precision (1)	Recall (1)	F1 (1)	Accuracy
Random Forest								
	SpeechBrain (w/ age)	0.52	0.57	0.49	0.69	0.66	0.65	0.6034
	SpeechBrain (w/o age)	0.51	0.61	0.50	0.69	0.62	0.62	0.5938
	SpeechBrain Females (w/ age)	0.53	0.39	0.40	0.72	0.81	0.75	0.6582
	SpeechBrain Females (w/o age)	0.53	0.39	0.40	0.72	0.81	0.75	0.6582
	openSMILE (w/ age)	0.6427	0.5896	0.6120	0.6575	0.7036	0.6772	0.6508
	openSMILE (w/o age)	0.6391	0.5954	0.6131	0.6604	0.6978	0.6765	0.6509
	openSMILE Females (w/ age)	0.5801	0.5486	0.5578	0.6754	0.7034	0.6848	0.6378
	openSMILE Females (w/o age)	0.6025	0.5662	0.5774	0.6864	0.7199	0.6985	0.6530
SVM								
	SpeechBrain (w/ age)	0.53	0.57	0.53	0.69	0.66	0.66	0.6215
	SpeechBrain (w/o age)	0.53	0.57	0.54	0.69	0.66	0.66	0.6233
	SpeechBrain Females (w/ age)	0.45	0.44	0.42	0.74	0.75	0.74	0.6519
	SpeechBrain Females (w/o age)	0.53	0.57	0.54	0.69	0.66	0.66	0.6233
	openSMILE (w/ age)	0.6371	0.6809	0.6527	0.6953	0.6485	0.6659	0.6612
	openSMILE (w/o age)	0.6403	0.6566	0.6424	0.6853	0.6650	0.6700	0.6595
	openSMILE Females (w/ age)	0.5827	0.6429	0.6065	0.7108	0.6531	0.6766	0.6529
	openSMILE Females (w/o age)	0.5731	0.6660	0.6065	0.7171	0.6159	0.6548	0.6439
Logistic Regression								
	SpeechBrain (w/ age)	0.52	0.58	0.53	0.69	0.65	0.66	0.6145
	SpeechBrain (w/o age)	0.52	0.58	0.53	0.69	0.65	0.66	0.6146
	SpeechBrain Females (w/ age)	0.40	0.51	0.43	0.75	0.65	0.69	0.6111
	SpeechBrain Females (w/o age)	0.42	0.54	0.46	0.76	0.67	0.71	0.6288
	openSMILE (w/ age)	0.6328	0.6724	0.6485	0.6924	0.6502	0.6673	0.6602
	openSMILE (w/o age)	0.6402	0.6566	0.6424	0.6852	0.6649	0.6700	0.6595
	openSMILE Females (w/ age)	0.5791	0.6934	0.6262	0.7332	0.6200	0.6671	0.6578
	openSMILE Females (w/o age)	0.5886	0.6982	0.6315	0.7434	0.6287	0.6747	0.6657
Neural Network								
	SpeechBrain (w/ age)	0.44	0.54	0.46	0.63	0.54	0.57	0.5407
	SpeechBrain (w/o age)	0.45	0.58	0.49	0.64	0.53	0.56	0.5434
	SpeechBrain Females (w/ age)	0.35	0.64	0.44	0.74	0.44	0.55	0.52
	SpeechBrain Females (w/o age)	0.40	0.33	0.36	0.59	0.66	0.62	0.5077
	openSMILE (w/ age)	0.6404	0.6431	0.6308	0.6682	0.6544	0.6504	0.6452
	openSMILE (w/o age)	0.6384	0.6333	0.6230	0.6760	0.6653	0.6609	0.6497
	openSMILE Females (w/ age)	0.5716	0.6660	0.6128	0.7213	0.6249	0.6665	0.6533
	openSMILE Females (w/o age)	0.5737	0.6741	0.6135	0.7182	0.6187	0.6589	0.6464

4 RESULTS

This section presents the results of the experiments and performance evaluation. A total of 32 model outputs were generated by combining the four machine learning models with the two different feature extraction toolkits. Each combination was tested under four distinct data configurations:

- Including both male and female participants with age
- Including both male and female participants without age
- Female-only participants with age
- Female-only participants without age

This stratification enabled the exploration of how gender and age metadata influence model performance and generalizability. Each experiment yielded metrics for both COPD-positive (label: 1) and COPD-negative (label: 0) classes [1]. The metrics reported include:

- Precision: the proportion of true positives among all predicted positives.
- Recall: the proportion of true positives identified among all actual positives.
- F1 Score: the harmonic mean of precision and recall, providing a balance between the two.
- Accuracy: the overall proportion of correct predictions. [1]

[1]Class "0" corresponds to non-COPD individuals. Class "1" refers to those diagnosed with COPD

These metrics were calculated using 5-fold and 10-fold cross-validation. Cross-validation is a technique used to evaluate how well a machine learning model generalizes to unseen data. In a 5-fold cross-validation, the dataset is split into five equal parts (folds). The model is then trained on four folds and tested on the remaining one fold. This is repeated five times, each time using a different fold as the test set. The final performance metrics (e.g., accuracy, F1 score) are averaged across all five runs. The same logic applies to the 10-fold, but with the data split into ten parts instead of five. This provides a more stable estimate of performance, especially on smaller datasets, though it takes more time to compute [10].

A summary of these performance metrics is presented in Table 2. This comprehensive comparison illustrates how different combinations of feature sets, model types, and demographic configurations affect model performance.

These findings are further explored in the Discussion section, which examines key limitations and reflects on the impact of factors such as gender imbalance, the inclusion of age metadata, differences between audio feature extraction toolkits, and other challenges encountered during the project.

4.1 Confusion Matrix

Figure 4 presents the averaged confusion matrix for the SVM model trained, using features extracted with the openSMILE toolkit. Due to its strong performance, this model was selected for detailed illustration through a confusion matrix. The matrix displays the model's predictions on individual audio chunks across all validation folds. Each chunk was labeled based on the participant's known COPD status (non-COPD and COPD).

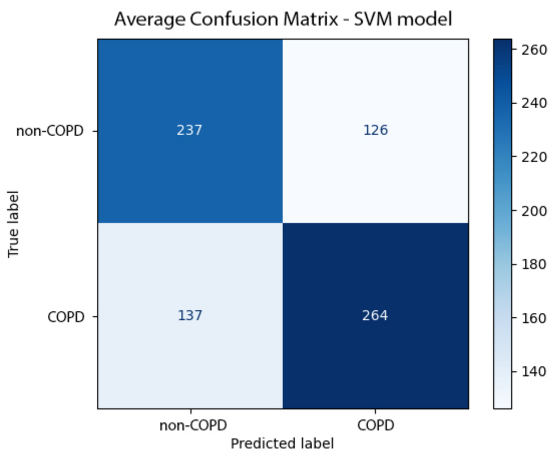


Figure 4: Averaged Confusion Matrix of SVM Model (openSMILE)

The vertical axis represents the true class labels, while the horizontal axis shows the predicted labels. The confusion matrix shows that the model correctly classified 237 non-COPD chunks and 264 COPD chunks. On the other hand, 126 non-COPD chunks were misclassified as COPD and 137 COPD chunks were misclassified as non-COPD. These outputs serve as the basis for evaluating the model's performance using several key metrics, from which accuracy, precision, recall, and F1-score can be calculated. These calculated values are presented in Table 2 for all models.

The color intensity in the matrix reflects the frequency of predictions, where darker shades represent higher counts. The confusion matrix provides an overview of how the SVM model performs in distinguishing between the two classes, serving as the basis for further analysis in the discussion section.

5 DISCUSSION

This section presents the key findings of the study, examining both the strengths and limitations of the methodological approach and model performance. It aims to contextualize the results presented in the previous section by examining potential sources of bias, data-related challenges, and the implications of demographic variation and model behavior. Additionally, this section outlines critical factors that may influence the generalizability and robustness of the findings while highlighting avenues for future research and improvement.

5.1 Method Limitations

While the methods employed in this study were carefully designed to ensure consistency and reliability, several limitations emerged during data collection and model development. These limitations may have impacted both the quality of the dataset and the generalizability of the machine learning models. The following subsections outline key challenges and considerations that should be addressed in the future development of our models.

5.1.1 ***Collecting Online Recordings.*** As mentioned in Section 3.1, recruitment was initiated by publishing a post in several Facebook groups aimed at reaching potential participants. Individuals were given the option to either record themselves or meet in person for assistance with the recording process. The majority chose to record themselves. However, many did not follow up, and some withdrew after reviewing the provided instructions for completing and submitting the recording.

It was initially assumed that a portion of users in these forums might have limited technical proficiency. To address this potential barrier, a short video guide was developed to clearly outline each step required for creating and submitting the recording. However, this did not have any effect on gathering more data from this demographic online.

5.1.2 ***Imbalanced Gender in Dataset.*** During the collection of the recordings from COPD and non-COPD participants, we attended multiple events where most participants were women, which resulted in an imbalance in our dataset. However, this imbalance in our dataset could also be a result of the prevalence of COPD in females has increased, and the number of females diagnosed with COPD in the United States now outnumbers males. A possible reason for this is females may be more susceptible to the effects of cigarettes compared to males [37]. The imbalance could limit our research, as vocal characteristics differ between men and women due to physiological factors that influence their voice patterns. A dataset dominated by women might affect the model's ability to generalize to the broader population, particularly male patients.

5.1.3 ***Unverified Self-Reported Diagnoses.*** A significant limitation of this study is the verification of participants' COPD status. The classification of individuals into COPD and non-COPD groups was based solely on self-reporting. No medical records, diagnostic test results (such as spirometry), or clinical documentation were obtained to confirm diagnoses. As a result, there is a risk of misclassification, especially among participants in the control group, where undiagnosed COPD may have gone unrecognized. This introduces potential noise into the training data, which could affect model performance.

5.1.4 ***Age Limitations.*** An age-related limitation was present during the data collection process, as the median age of the participants was 72. To ensure the most valid and generalizable results, the control group needed to be of a similar age, aligning with the typical age range at which individuals are commonly diagnosed with COPD [16]. However, completely excluding younger individuals would reduce the model's precision if used by a broader demographic. To address this, a smaller subset of younger participants in their late 20s, 30s and 40s was also included in the final model.

Another limitation related to age is that COPD symptoms typically begin to manifest in individuals over the age of 40 [9]. Additionally, research shows that many people experience symptoms for several years before receiving a formal diagnosis. This presents a potential issue in the dataset – especially at scale – as some individuals labeled as "non-COPD" may, in fact, unknowingly exhibit early signs of the disease without yet being diagnosed. As a result, the model may learn from mislabeled data, inadvertently treating early-stage COPD cases as healthy controls. This could reduce the model's overall precision and hinder its ability to accurately distinguish between healthy and affected individuals, particularly in the early stages of the disease.

5.1.5 ***Inclusion of Active Smokers.*** This limitation of the study concerns the composition of the control group, which includes individuals who are current or former smokers. While these participants reported no known respiratory conditions, it cannot be ruled out that some might have undiagnosed or early-stage COPD. This introduces a risk of false positives in the classification task, as the model may learn to associate smoking-related vocal characteristics with COPD.

However, excluding smokers entirely from the control group would have introduced another bias, where we would be comparing primarily non-smokers with a COPD group largely composed of participants with a history of smoking. This would risk training the model to distinguish between smoking status rather than the status of their disease. Including smokers in the control group, therefore, reflects a more realistic scenario, where early COPD is often undiagnosed, but also underlines the need for awareness in interpreting borderline cases.

5.1.6 ***Sustained Vowels.*** Another limitation relates to the variability in how participants performed the sustained vowel pronunciation task. The purpose of this task was to capture long, steady vowel sounds that could help highlight vocal or respiratory differences between individuals with and without COPD. However, several participants did not follow the instructions as intended. In some cases, vowels were pronounced quickly or with little effort to sustain the sound, while others varied in how long or clearly they vocalized the vowels. Furthermore, some participants were either uncomfortable with or unable to read aloud. In such cases, they instead shared a short anecdote or a personal story.

This inconsistency in task execution may have introduced differences in the recordings that are unrelated to COPD, which can affect how the machine learning models interpret and learn from the data. When the input quality varies from one participant to another, it becomes more difficult for the model to focus on the disease-related patterns. Although this kind of variation reflects real-world user behavior, it also highlights the importance of providing clearer instructions and possibly excluding recordings that do not meet the expected task format in future studies.

5.1.7 ***Post-Processing of Audio.*** During the post-processing of the collected audio recordings, several samples contained background noise or music that needed to be removed to prevent confusion for the models. However, even when applied carefully, noise reduction tools can unintentionally eliminate important speech features. This may result in the loss of relevant information critical for accurately detecting patterns associated with COPD, potentially degrading the model's performance.

Having discussed the limitations of the methods utilized, we now turn to the performance and interpretation of our model.

5.2 Performance of the Models

This section evaluates and compares the performance of the various machine learning models trained to classify COPD from voice recordings. The discussion is structured around model types, feature sets, and the impact of demographic information.

Reviewing the results of the trained models in Table 2, one of the first notable observations is the overall consistency in performance across most models. However, the neural network—particularly when using the SpeechBrain toolkit, exhibits the lowest accuracy among all evaluated models. Several factors likely contribute to this outcome. Neural networks generally require substantially larger datasets to generalize effectively and avoid overfitting. A commonly cited rule of thumb suggests that a neural network needs at least ten times more training samples than the number of parameters in the model [15]. When using openSMILE, the eGeMAPSv02 feature set was applied, which includes up to 88 acoustic features — potentially with even more trainable parameters depending on the network architecture [4]. This suggests that a minimum of approximately 880 samples would be necessary to fully exploit a neural network's capacity. In contrast, the SpeechBrain model employs a deep learning-based feature extractor (spkrec-xvect-voxceleb) that outputs 512-dimensional speaker embeddings and likely contains thousands of trainable parameters [17]. Consequently, several thousand recordings would be required for effective training — far more than were available in this study.

In addition to data limitations, the architectural design of neural networks may also influence performance. Unlike traditional machine learning models, neural networks are often considered "black boxes," as the internal processes by which inputs are transformed into outputs are not easily interpretable. This lack of transparency complicates efforts to diagnose performance issues, which may stem from data scarcity, overfitting, or suboptimal model design—such as the number and type of layers, the number of nodes per layer, or the choice of activation functions. It is, therefore, plausible that the neural network implemented in this study was not optimally configured for the relatively small dataset and high-dimensional feature representation, further contributing to its lower accuracy [31].

When evaluating models for the early detection of diseases such as COPD, achieving a high recall for positive cases is particularly important. In this context, it is generally preferable to produce false positives – identifying healthy individuals as potentially having COPD – than to miss actual cases, as the latter could delay necessary treatment. However, this prioritization must be properly balanced to avoid overwhelming false positive rates. Notably, the Random Forest models trained exclusively on female participants using the SpeechBrain toolkit stood out with a recall of 81% for COPD class (1). Despite this strong performance in identifying positive cases, the same models showed a recall of only 39% for non-COPD cases, suggesting a significant bias toward predicting the presence of the disease. This imbalance indicates a high rate of false positives for healthy individuals. Such skewed performance is consistent across most models trained using SpeechBrain, likely due, as previously mentioned, to the fact that deep learning-based toolkits require substantially larger datasets to generalize effectively and maintain balanced predictions across classes.

When evaluating the top-performing model triads – Random Forest, SVM, and Logistic Regression trained with the openSMILE feature set – all demonstrate accuracy within the range of 64–66%. Notably, the SVM model trained on both male and female participants, without incorporating age metadata, achieves a well-balanced recall: 66.5% for COPD cases and 65.6% for non-COPD cases. Logistic Regression shows a similar trend, while Random Forest tends to prioritize recall for COPD cases (72%) at the cost of reduced performance for non-COPD cases. These differences can be attributed to the fundamental mechanics of each algorithm. Random Forest, an ensemble method based on decision trees, is generally robust against overfitting and does not require feature scaling [19]. However, this robustness may come at the expense of nuanced predictions in smaller datasets, especially for underrepresented classes. Like Neural Networks, Random Forest models typically perform better with larger datasets, where their ensemble nature can leverage greater variance [12]. Conversely, SVM and Logistic Regression may perform adequately on smaller datasets but are more sensitive to issues such as class imbalance and the absence of detailed feature scaling. While these models are also more prone to overfitting under these conditions, the use of validation techniques such as StratifiedGroupKFold helps mitigate these risks by preserving class distributions across training and test folds [34].

Finally, in examining gender-specific classification, the models generally demonstrate slightly higher precision in detecting COPD when trained exclusively on female participants. Performance also varies depending on the inclusion of age metadata. This trend is likely influenced by the dataset's imbalance, with a significantly higher number of female recordings. Moreover, the relatively small overall dataset size may limit the generalizability and robustness of the models. These observations suggest that model performance could benefit not only from a more balanced and diverse dataset – particularly one with a greater proportion of male participants – but also from a larger volume of samples overall. Notably, the SVM model appears to utilize age metadata more heavily in its decision-making process, especially regarding COPD recall, where a performance difference of approximately 4 percentage points is observed. This highlights a potential area where feature scaling could enhance model accuracy, particularly if the model is overly sensitive or biased toward raw age values.

5.2.1 ***SVM Confusion Matrix***. As shown in Figure 4, the confusion matrix illustrates the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions.

When diagnosing COPD, balancing false positives and false negatives is crucial, as it reflects the broader challenge of underdiagnosis and overdiagnosis. In this case, the model had a recall of 65.83% for the COPD class, meaning it failed to identify about 34.17% of actual COPD cases (false negatives). This is concerning, as missing a COPD diagnosis early can delay treatment and worsen the patient's condition. Although the model's precision was slightly higher at 67.69%, the number of false positives (126) remains significant and could lead to unnecessary stress for the user.

5.2.2 ***Proof of Concept***. While the current findings highlight promising avenues for classifying COPD from voice recordings, the modest accuracy and class imbalance underscore limitations in the existing dataset. A pronounced gender skew and a limited overall sample size constrain both the generalizability and robustness of the results. Future research should prioritize expanding the dataset – particularly by including more male participants and increasing the total number of samples. Furthermore, incorporating domain knowledge for feature selection or exploring multimodal data fusion (e.g., combining audio with sensor or questionnaire data) could further enhance model performance.

An additional opportunity lies in utilizing the collected COPD-GOLD status data (see listing 1). Training models specifically on early-stage cases (GOLD 1–2), combined with a larger and more diverse dataset, could potentially enhance the ability to detect early signs of COPD more effectively. Another promising direction would be to train models capable of distinguishing between GOLD stages. Such a system could be valuable for monitoring disease progression in diagnosed patients, assessing whether their condition is improving or worsening – potentially estimating lung function remotely without requiring a hospital visit.

These findings demonstrate the feasibility of using voice features to classify COPD, particularly when combined with metadata such as age. While the models are not yet robust enough for clinical deployment, the observed performance, especially in female participants, provides compelling proof of concept. Future work with larger, more balanced datasets and additional feature engineering could further improve predictive accuracy and generalizability.

6 CONCLUSION

This study explored the feasibility of utilizing voice recordings and machine learning techniques to detect Chronic Obstructive Pulmonary Disease (COPD). Audio data were collected from individuals diagnosed with COPD as well as from a control group without known respiratory conditions. Four classification models — Logistic Regression, Support Vector Machine (SVM), Random Forest, and a Neural Network were trained on acoustic features extracted using both the openSMILE and SpeechBrain toolkits.

The findings demonstrate that several models, particularly those utilizing openSMILE features, are capable of distinguishing between COPD and non-COPD cases with moderate accuracy, recall, and F1-scores. While performance varied depending on model architecture and metadata configurations such as gender and age, the results provide a compelling proof of concept for leveraging voice analysis in COPD screening. In particular, the SVM and Random Forest models showed consistently balanced classification outcomes across multiple experimental setups.

Nevertheless, several limitations must be acknowledged. The relatively small sample size, gender imbalance, and reliance on self-reported diagnoses may have introduced bias and limited the generalizability of the results. Furthermore, some participants did not consistently perform the required tasks – such as sustained vowel pronunciation – potentially introducing noise into the feature extraction process. These challenges highlight the importance of standardized recording procedures and the use of clinically verified datasets in future research.

Despite these constraints, the project makes a meaningful contribution to the emerging field of speech analysis by demonstrating the potential of low-cost, non-invasive diagnostic tools based on voice. This approach holds promise for scalable, at-home screening of COPD. Future work should aim to significantly expand the dataset, especially by including more male participants and clinically validated cases, and explore whether machine learning models can not only detect COPD but also assess its severity or progression.

7 ACKNOWLEDGMENTS

We would like to express our sincere gratitude to all those who contributed to the success of this project. Special thanks go to Lungeforeningen and Nordkraft Sundhedscenter Aalborg for their support in helping us recruit participants and collect valuable recordings from individuals living with COPD. We also appreciate the cooperation of Aarhus Lung Choir and Aarhus Music School during our data collection process.

We are particularly thankful to Ulla Møller Weinreich from Lungemedicinsk Afdeling for her generous help in facilitating access to patients and connecting us with relevant contacts. We would also like to thank TV 2 Nord for their coverage of our project, as well as Peter Witten from AAU Kommunikation og Public Affairs for helping promote our project. Furthermore, we are deeply grateful to all participants, both with and without COPD, who generously contributed their time and voices to this study.

We are also grateful to Andreas Triantafyllopoulos from the University of Augsburg for his helpful insights and methodological guidance on our data collection tasks and codebase review. Lastly, we would like to thank our supervisor, Ilhan Aslan, for his continuous support, valuable feedback, and encouragement throughout the project.

8 APPENDIX

Appendix 1: Thesis summary: Detecting COPD Through Speech Analysis: A Dataset and Machine Learning Approach. See Appendix 1: Thesis summary: Detecting COPD Through Speech Analysis: A Dataset and Machine Learning Approach.

REFERENCES

[1] Nirajan Acharya. 2024. Understanding Precision, Recall, F1-score, and Support in Machine Learning Evaluation. Retrieved d. 29-05-2025 Link: https://medium.com/@nirajan.acharya777/understanding-precision-recall-f1-score-and-support-in-machine-learning-evaluation-7ec935e8512e.

[2] ACM. 2025.. CHI'25 - Proceedings of the 2025 CHI Conference on Human Factors in Computing Sytems.

[3] Mathews AM. 2023. The Functional and Psychosocial Consequences of COPD. *Respir Care* 18 (2023). https://doi.org/doi:10.4187/respcare.10542

[4] audEERING. N.D.. openSMILE Python. Retrieved d. 30-05-2025 Link: https://audeering.github.io/opensmile-python/.

[5] audEERING® GmbH. 2023. Components. Retrieved d. 28-02-2025 Link: https://audeering.github.io/opensmile/reference.html#components.

[6] audEERING® GmbH. 2023. openSMILE 3.0. Retrieved d. 28-02-2025 Link: https://www.audeering.com/research/opensmile/.

[7] Khai N. Truong et al. Brenna Li, Saba Tauseef. 2025. A Comparative Analysis of Information Gathering by Chatbots, Questionnaires, and Humans in Clinical Pre-Consultation. *ACM* (2025). https://doi.org/10.1145/3706598.3713613

[8] S. Brinkmann and S. Kvale. 2018. Doing Interviews. *Sage Publications* 2 (2018), 81–114. https://doi.org/10.4135/9781529716665

[9] Anuj K. Agarwal; Avais Raja; Brandon D. Brown. 2023. Chronic Obstructive Pulmonary Disease. Retrieved d. 20-05-2025 Link: https://www.ncbi.nlm.nih.gov/books/NBK559281/.

[10] Jason Brownlee. 2023. A Gentle Introduction to k-fold Cross-Validation. Retrieved d. 29-05-2025 Link: https://machinelearningmastery.com/k-fold-cross-validation/.

[11] Cleveland Clinic. 2024. Chronic Obstructive Pulmonary Disease (COPD). Retrieved d. 20-03-2025 Link: https://my.clevelandclinic.org/health/diseases/8709-chronic-obstructive-pulmonary-disease-copd.

[12] StackExchange DataScience. 2022. When to use Random Forest over SVM and vice versa? Retrieved d. 30-05-2025 Link: hhttps://datascience.stackexchange.com/questions/6838/when-to-use-random-forest-over-svm-and-vice-versa.

[13] Rahul C. Deo. 2015. Machine Learning in Medicine. Retrieved d. 20-03-2025 Link: https://doi.org/10.1161/CIRCULATIONAHA.115.001593.

[14] Nermin Diab, Andrea S. Gershon, Don D. Sin, and et. al. 2018. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine* 198, 9 (2018). https://doi.org/10.1164/rccm.201804-0621CI

[15] Eugene Dorfman. 2022. How Much Data Is Required for Machine Learning? Retrieved d. 30-05-2025 Link: https://postindustria.com/how-much-data-is-required-for-machine-learning/.

[16] Sebastian Tristan Ehlert Buchardt et. al. 2024. Patient characteristics and mortality across diagnostic settings in COPD. Retrieved d. 20-05-2025 Link: https://www.sciencedirect.com/science/article/pii/S0954611124003184?.

[17] Hugging Face. N.D.. Speech Brain - spkrec-xvect-voxceleb. Retrieved d. 30-05-2025 Link: https://huggingface.co/speechbrain/spkrec-xvect-voxceleb/blob/main/hyperparams.yaml.

[18] Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2021. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease (2021 REPORT). Retrieved d. 10-03-2025 Link: https://goldcopd.org/wp-content/uploads/2020/11/GOLD-REPORT-2021-v1.1-25Nov20_WMV.pdf.

[19] GeeksforGeeks. 2025. Logistic Regression Vs Random Forest Classifier. Retrieved d. 30-05-2025 Link: https://www.geeksforgeeks.org/logistic-regression-vs-random-forest-classifier/.

[20] GeeksforGeeks. 2025. What is a Neural Network? Retrieved d. 06-05-2025 Link: https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/.

[21] Andrea S. Gershon, Deva Thiruchelvam, Kenneth R. Chapman, Shawn D. Aaron, Matthew B. Stanbrook, Jean Bourbeau, Wan Tan, and Teresa To. 2018. Health Services Burden of Undiagnosed and Overdiagnosed COPD. *Chest* 153, 6 (2018), 1336–1346. https://doi.org/10.1016/j.chest.2018.01.038

[22] Karen Hao. 2018. What is machine learning? Retrieved d. 20-03-2025 Link: https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/.

[23] Terence Ho, Ruth P. Cusack, Nagendra Chaudhary, Imran Satia, and Om P. Kurmi. 2019. Under- and over-diagnosis of COPD: a global perspective. *Breathe (Sheff)* 15, 1 (2019). https://doi.org/10.1183/20734735.0346-2018

[24] Nick Hopkinson. 2016. COPD: the disease that too often goes undetected. Retrieved d. 23-05-2025 Link: https://www.imperial.ac.uk/news/176135/copd-disease-that-often-goes-undetected/.

[25] et al. Hupo Bian, Shaoqi Zhu. 2024. Artificial Intelligence in Chronic Obstructive Pulmonary Disease: Research Status, Trends, and Future Directions –A Bibliometric Analysis from 2009 to 2023. *International journal of chronic obstructive pulmonary disease* (2024). https://doi.org/10.2147/COPD.S474402

[26] IBM. 2023. What are SVMs? Retrieved d. 23-04-2025 Link: https://www.ibm.com/think/topics/support-vector-machine.

[27] IBM. N.D.. What is random forest? Retrieved d. 23-04-2025 Link: https://www.ibm.com/think/topics/random-forest.
[28] Sachin R Pendse et al. Inhwa Song, SoHyun Park. 2025. ExploreSelf: Fostering User-driven Exploration and Reflection on Personal Challenges with Adaptive Guidance by Large Language Models. *ACM* (2025). https://doi.org/10.1145/3706598.3713883
[29] Noémie Elhadad Julia Hirschberg, Anna Hjalmarsson. 2010. "You're as Sick as You Sound": Using Computational Approaches for Modeling Speaker State to Gauge Illness and Recovery. Retrieved d. 20-03-2025 Link: https://link.springer.com/chapter/10.1007/978-1-4419-5951-5_13.
[30] Vijay Kanade. 2022. What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices.
[31] Matthew Kosinski. 2014. What is black box artificial intelligence (AI)? Retrieved d. 04-06-2025 Link: https://www.ibm.com/think/topics/black-box-ai.
[32] Scikit Learn. N.D.. GridSearchCV. Retrieved d. 23-05-2025 Link: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.
[33] Scikit Learn. N.D.. GridSearchCV. Retrieved d. 23-05-2025 Link: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html.
[34] Scikit Learn. N.D.. StratifiedGroupKFold. Retrieved d. 30-05-2025 Link: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedGroupKFold.html.
[35] Ching-Hsiung et al. Lin. 2023. Current Progress of COPD Early Detection: Key Points and Novel Strategies. *International journal of chronic obstructive pulmonary disease* 18 (2023). https://doi.org/doi:10.2147/COPD.S413969
[36] Manuel Milling, Florian B. Pokorny, Katrin D. Bartl-Pokorny, and Björn W. Schuller. 2022. Is Speech the New Blood? Recent Progress in AI-Based Disease Detection From Audio in a Nutshell. *Frontiers in Digital Health* 4 (2022), 886615. https://doi.org/10.3389/fdgth.2022.886615
[37] Kathryn M Milne, Reid A Mitchell, Olivia N Ferguson, Alanna S Hind, and Jordan A Guenette. 2024. Sex-differences in COPD: from biological mechanisms to therapeutic considerations. *Frontiers in Medicine* 11 (mar 2024), 1289259. https://doi.org/10.3389/fmed.2024.1289259
[38] Arsha Nagrani, Joon Son Chung, Weidi Xie, and Andrew Zisserman. 2020. Voxceleb: Large-scale speaker verification in the wild. *Computer Speech Language* (2020).
[39] Nicholas Cumminsa, Alice Bairda, Björn W. Schuller. [n. d.]. Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. Retrieved 19-03-2025 Link: https://www.sciencedirect.com/science/article/pii/S1046202317303717.
[40] TV2 Nord. 2025. TV2 Nord Nyhederne - 26th of May, 2025. Retrieved d. 29-05-2025 Link: https://www.tv2nord.dk/nyheder/26-05-2025/1930/1905-26-may-2025?t=934.
[41] Jiaqing Liu et al. Qiuyue (Shirley) Xue, Eric Steven Martin. 2025. ECG Necklace: Low-power Wireless Necklace for Continuous ECG monitoring. *ACM* (2025). https://doi.org/doi/10.1145/3706598.3713742
[42] Mirco Ravanelli, Titouan Parcollet, Adel Moumen, Sylvain de Langen, Cem Subakan, Peter Plantinga, Yingzhi Wang, Pooneh Mousavi, Luca Della Libera, Artem Ploujnikov, Francesco Paissan, Davide Borra, Salah Zaiem, Zeyu Zhao, Shucong Zhang, Georgios Karakasidis, Sung-Lin Yeh, Pierre Champion, Aku Rouhe, Rudolf Braun, Florian Mai, Juan Zuluaga-Gomez, Seyed Mahed Mousavi, Andreas Nautsch, Xuechen Liu, Sangeet Sagar, Jarod Duret, Salima Mdhaffar, Gaelle Laperriere, Mickael Rouvier, Renato De Mori, and Yannick Esteve. 2024. Open-Source Conversational AI with SpeechBrain 1.0. arXiv:2407.00463 [cs.LG] https://arxiv.org/abs/2407.00463
[43] Mirco Ravanelli, Titouan Parcollet, Peter Plantinga, Aku Rouhe, Samuele Cornell, Loren Lugosch, Cem Subakan, Nauman Dawalatabad, Abdelwahab Heba, Jianyuan Zhong, Ju-Chieh Chou, Sung-Lin Yeh, Szu-Wei Fu, Chien-Feng Liao, Elena Rastorgueva, François Grondin, William Aris, Hwidong Na, Yan Gao, Renato De Mori, and Yoshua Bengio. 2021. SpeechBrain: A General-Purpose Speech Toolkit. arXiv:2106.04624 [eess.AS] arXiv:2106.04624.
[44] Jean Rouat, Ramin Pichevar, and Stéphane Loiselle. 2005. *Perceptive, Non-linear Speech Processing and Spiking Neural Networks*. Springer Berlin Heidelberg, 317–337. https://doi.org/10.1007/11520153_14
[45] Scikit-learn. N.D.. LogisticRegression. Retrieved d. 16-04-2025 Link: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.
[46] Scikit-learn. N.D.. Scikit-learn - Machine Learning in Python. Retrieved d. 16-04-2025 Link: https://scikit-learn.org/stable/index.html.
[47] James Spradley. 1979. *The Ethnographic Interview*. Wadsworth,Cengage Learning.
[48] et. al. Stuart Cunningham. 2017. Cloud-Based Speech Technology for Assistive Technology Applications (CloudCAST). Retrieved d. 20-03-2025 Link: https://www.scopus.com/record/display.uri?eid=2-s2.0-85028872046&origin=resultslist.
[49] Simon Tavasoli. 2025. The Machine Learning Algorithms List: Types and Use Cases. Retrieved d. 16-04-2025 Link: https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article.
[50] WEBMD. N.D. COPD Stages and the GOLD Criteria. Retrieved d. 16-05-2025 Link: https://www.webmd.com/lung/copd/gold-criteria-for-copd.
[51] WHO. 2024. Chronic obstructive pulmonary disease (COPD). Retrieved d. 26-02-2025 Link: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd).
[52] Anton Batliner Björn W Schuller Thomas M Berghaus Wolfgang Mayr, Andreas Triantafyllopoulos. 2025. Assessing the Clinical and Functional Status of COPD Patients Using Speech Analysis During and After Exacerbation. Retrieved d. 20-03-2025 Link: https://doi.org/10.2147/COPD.S480842.
[53] Seoyoon Woo, Weijiao Zhou, and Janet L Larson. 2021. Stigma Experiences in People with Chronic Obstructive Pulmonary Disease: An Integrative Review. *International Journal of Chronic Obstructive Pulmonary Disease* 16 (2021). https://doi.org/10.2147/COPD.S306874