

PROCEEDINGS | African Engineering Education Forum (AEEF 2026)

Hybrid SVM and CNN System for Cough Detection and Health Classification

Fabien Mouomene Moffo ^{*,1} Auguste Noumsi ¹ Joseph Mvogo ² and Samuel Bowong ¹

¹Faculty of Science, Douala, Cameroon

²IUT, Douala, Cameroon

*Corresponding author: fmouomene@yahoo.fr

(Received: 17 June 2026; Revised: 13 August 2026; Accepted: 22 August 2026; Published: 11 September 2026)

(Editor: Dr. Bappa Muktar; Reviewers: Anonymous reviewers)

Abstract

This study introduces a hybrid architecture designed to address the challenges of automated cough sound analysis by combining unsupervised detection with supervised classification. The system integrates a One-Class Support Vector Machine (SVM) trained exclusively on positive cough recordings to establish an acoustic boundary that isolates true cough events (inliers) while rejecting out-of-distribution background noise (outliers) from raw audio streams and a convolutional neural network (CNN) to classify the extracted cough sounds into three clinically relevant categories. The preprocessing pipeline is integral to the system's effectiveness. The process involves extracting static Mel-Frequency Cepstral Coefficients (MFCCs), followed by standardization and dimensionality alignment to ensure compatibility with the CNN. These steps ensure that the spectral and temporal characteristics of cough sounds are accurately represented while reducing redundancy and enhancing separability in the feature space. Experimental evaluation demonstrates the reliability of this hybrid framework. As a standalone classifier on clean, isolated cough sounds, the DenseNet achieves an accuracy of 96.0% and an F_1 -score of 0.96. When evaluated on complex, unsegmented audio streams, the end-to-end hybrid system achieves an overall accuracy of 93.0% and a corresponding F_1 -score of 0.91. These results underscore the system's resilience and capacity for generalization, even when confronted with noisy or heterogeneous data. The incorporation of One-Class SVM for unsupervised detection ensures that cough events are effectively isolated, while the CNN provides precise classification, making the system suitable for real-world applications. Generalizability requires further validation on prospective external cohorts. Beyond its clinical relevance, this work also holds pedagogical value for engineering education. It illustrates the integration of signal processing, machine learning, and system design in a coherent pipeline, serving as a practical case study for students and researchers. By integrating theoretical concepts with applied methodologies, the system illustrates how hybrid architectures can be utilized to address complex challenges in health monitoring. Overall, the proposed framework offers a reliable foundation for preliminary screening and educational purposes.

Keywords: One-Class SVM; DenseNet; Hybrid architecture; Cough classification; Respiratory sound analysis; e-health

Abbreviations: AI: Artificial Intelligence; AST: Audio Spectrogram Transformer; AUC-ROC: Area Under the Receiver Operating Characteristic Curve; CNN: Convolutional Neural Network; COPD: Chronic Obstructive Pulmonary Disease; COVID-19: Coronavirus Disease 2019; DCT: Discrete Cosine Transform; ECR: Energy Concentration Ratio; EPFL: École Polytechnique Fédérale de Lausanne; ESC-50: Environmental Sound Classification 50; GPU: Graphics Processing Unit; Grad-CAM: Gradient-weighted Class Activation Mapping; LPC: Linear Predictive Coding; MFCC: Mel-Frequency Cepstral

Coefficient; N/A: Not Applicable; PLP: Perceptual Linear Prediction; RBF: Radial Basis Function; ROI: Region of Interest; RT-PCR: Reverse Transcription Polymerase Chain Reaction; SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2; SVM: Support Vector Machine.

1. Introduction

Recent global health crises have highlighted the pressing need for scalable, rapid, and non-invasive diagnostic tools. Automated respiratory sound analysis is a promising avenue of research, with coughing serving as a primary clinical biomarker for the screening of various pathologies [1, 2, 3, 4]. This approach is critical in low-resource countries with limited medical infrastructure. Nonetheless, the implementation of these solutions in real-world scenarios is significantly hindered by the presence of complex and noisy acoustic environments [1, 3].

To overcome these challenges, this study introduces a novel hybrid architecture blending unsupervised learning and supervised classification. First, a One-Class Support Vector Machine (One-Class SVM) acts as an intelligent filter, isolating cough segments from heterogeneous background noise without requiring an exhaustively labeled negative dataset. Second, static Mel-frequency cepstral coefficients (MFCCs) effectively capture fine spectral nuances [5]. Finally, the hybrid system ensures classification precision, achieving an accuracy of 93% and an F_1 -score of 0.91.

To make an accurate diagnosis in real-world situations, it is essential to clearly distinguish cough sounds. This article addresses the following question: How can a hybrid architecture be created to effectively isolate cough signals in a noisy environment while enabling accurate classification of respiratory problems? To answer this question and demonstrate the originality of the work, three supplementary questions are investigated: How does the combination of unsupervised filtering and a deep convolutional neural network (CNN) improve cough signal-to-noise separation? What is the effect of separating detection and classification on the system's accuracy, specificity, and overall robustness when faced with varied acoustic data? How can this processing pipeline contribute to the development of artificial intelligence (AI) systems for healthcare? [6] To clearly define the scope and novelty of this work, the contribution is threefold. First, on the engineering level, this study proposes a novel multi-stage hybrid pipeline combining an unsupervised One-Class SVM with a high-capacity DenseNet classifier; the One-Class SVM acts as a non-parametric acoustic filter that isolates true cough frames before feature tensor creation, eliminating out-of-distribution environmental noise and boosting overall robustness. Second, from a clinical research perspective, the pipeline delivers a targeted, expert-validated three-class disease screening framework (Healthy, Symptomatic, coronavirus disease 2019 (COVID-19)) optimized for low-resource environments. Third, from a pedagogical standpoint, this work serves as an applied, end-to-end case study designed for engineering education to bridge the gap between theoretical signal processing and real-world e-health AI systems.

The objective of this work is to design, validate, and present a complete, functional technological pipeline for epidemiological surveillance. Beyond its medical contributions, this research strategically addresses AI in pedagogy and systems engineering by serving as an academic case study. The combination of signal processing, AI, and software engineering is demonstrated to translate high-level theoretical concepts into viable e-health solutions while also training future engineers in the rigorous handling of sensitive data.

The remainder of this paper is organized as follows: The second section of this study reviews related work and the theoretical framework of acoustic descriptors and hybrid models. The third section of the text provides a comprehensive overview of the materials and methods employed in the study. This section encompasses a detailed exposition of the preprocessing procedures, the implementation of the One-Class SVM filter, and the configuration of the DenseNet model. The fourth section of this

text presents the experimental results and the evaluations of the diagnostic performance. Finally, the text goes on to discuss the pedagogical implications for engineering and healthcare education and to outline future research directions.

2. Literature Review

Early explorations in biomedical acoustics leaned heavily on classical feature extraction. Researchers relied on descriptors such as cepstral coefficients, spectral energy, zero-crossing rate, and simple temporal markers to capture the essence of cough sounds. These handcrafted features provided a foundation for distinguishing respiratory events, yet they struggled to accommodate the inherent variability of human physiology and the disruptive influence of ambient noise. Despite their limitations, these methods established the first discriminative markers for cough analysis and remain a historical cornerstone in respiratory sound research [7, 8].

The rise of artificial intelligence ushered in a new era of supervised cough sound classification. Deep convolutional neural networks (CNNs) and Time-Frequency Transformers became the dominant architectures, capable of isolating disease-specific spectral signatures across diverse clinical datasets [9, 10]. To overcome the scarcity of annotated data, researchers employed augmentation strategies such as Gaussian noise injection, pitch shifting, and temporal stretching. These techniques enhanced model robustness and generalization, enabling supervised systems to tackle multi-class diagnostic tasks for conditions including asthma, chronic obstructive pulmonary disease (COPD), and pneumonia [10].

In parallel, unsupervised and semi-supervised approaches emerged to address the challenge of limited annotation. These methods sought to uncover latent structures within unlabelled or partially labelled datasets, identifying recurring spectral and temporal patterns in cough signals. Such strategies proved particularly valuable in uncontrolled environments where annotation is costly and impractical. However, their performance remains vulnerable to persistent acoustic contamination, background artifacts, non-stationary noise, and overlapping ambient sources—which distort sensitive audio segments and compromise detection reliability [11].

More recently, hybrid architectures have been proposed, combining detection modules (SVMs, autoencoders) with deep classification networks (CNNs, DenseNets). These systems aim to balance robustness and interpretability, offering multi-stage pipelines that first isolate candidate cough events before assigning diagnostic labels. Yet, their generalization capacity is constrained by training data quality and contextual variability. To bridge the gap between algorithmic prediction and clinical trust, explainability techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) have been integrated, providing visual maps of spectral regions most influential in classification decisions. This interpretive layer enhances medical applicability by aligning computational outputs with clinicians' diagnostic reasoning [12].

To provide a rigorous contextualization of our approach, Table 1 synthesizes recent benchmark studies in automated cough detection and diagnostic classification. Direct raw accuracy comparison across literature is inherently limited due to variations in target classes, dataset heterogeneity, and background noise levels. To address this, the table explicitly highlights the scope, performance metrics, traceable publication venues, and key operational limitations of prior works. While existing supervised architectures achieve high performance on curated datasets, they often lack robustness against out-of-distribution ambient noise. Conversely, standalone unsupervised models isolate acoustic events effectively but lack multi-class diagnostic capability. These observations justified the design of this work (One-Class SVM-DenseNet-121).

Table 1. A summary of important studies on cough detection and diagnosis.

Study	Task and Target Classes	Dataset	Performance	Limitations
Pahar et al. [13]	Binary COVID-19 status classification from cough, breath, and speech	Coswara, Sarcos, and ComParE datasets	Area under the receiver operating characteristic curve (AUC-ROC): up to 0.982 for cough (ResNet50 transfer learning)	When evaluated on isolated clips, high performance degrades in non-stationary, noisy environments.
Orlandic et al. [14]	Open-access cough dataset and cough/non-cough detector	COUGHVID corpus	Mean ROC AUC: 96.4% (XGBoost cough detector)	Serves primarily as a public benchmark; baseline models lack multi-stage noise filtering.
Ashby et al. (2022) [15]	Cough audio preprocessing and automated signal segmentation	COUGHVID and Coswara	High-quality isolated coughs: 88% (COUGHVID), 82% (Coswara)	It focuses strictly on time-frequency segmentation, but does not perform multi-class clinical disease classification.
Lella & Pja [16]	COVID-19 classification from respiratory sounds	Cough, breath, and voice samples	Accuracy: 95.45%; F_1 : 96.96% (combined modalities)	Evaluation is task- and dataset-specific; external generalization requires separate validation.

3. Materials and methods

This section provides a detailed description of the dataset used, the tools employed, and the methodological steps followed to design, train, and evaluate the proposed system.

3.1 Methodology

Figure 1 illustrates the processing pipeline used to analyze cough sounds. The process begins with the acoustic signal being inputted and converted into MFCC coefficients. Next, a cough detection step is carried out using a One-Class SVM. If a cough is identified, it is passed to a DenseNet model for classification.

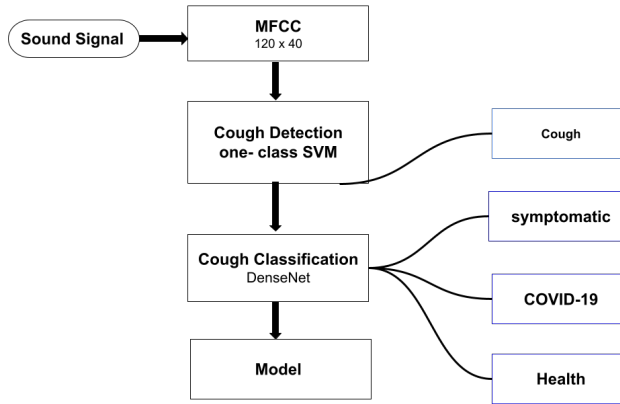


Figure 1. Methodology Map

3.2 Dataset

The dataset combines recordings from the COUGHVID corpus and our laboratory [14]. The COUGHVID corpus, developed by a team at the École Polytechnique Fédérale de Lausanne (EPFL), is seminal in the field of acoustic analysis of coughs during the pandemic. Cough signals were collected from a wide range of volunteers, recruited via digital platforms and awareness campaigns, to ensure demographic and clinical diversity. This included symptomatic and asymptomatic individuals, with or without a confirmed diagnosis of respiratory disease, in accordance with ethical standards. The recordings were conducted in uncontrolled environments to reflect realistic conditions. Precise instructions were provided to minimize acoustic artifacts, including maintaining a constant distance between the mouth and the microphone, avoiding ambient noise, and producing several consecutive coughs to capture intra-individual variability. The associated metadata encompasses sociodemographic, clinical, and contextual information, which is imperative for multimodal analysis and bias assessment. The signals were recorded using acoustic sensors that had been integrated into everyday devices, such as smartphones and laptops, as well as specialized medical devices. A preference was given to omnidirectional or cardioid microphones, which are suited to a variety of environments.

The class distribution and a representative cough waveform are presented in Figure 2.

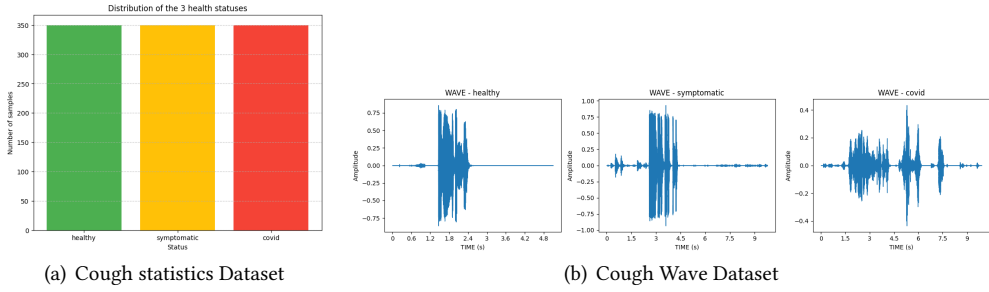


Figure 2. Cough Dataset

Figure 2(a) shows the distribution of the three health states in the dataset. It consists of three vertical bars representing the categories healthy (green), symptoms (yellow), and COVID-19 (red), respectively. Each bar is approximately 350 samples high, indicating that the dataset is balanced: each class contains the same number of examples. This even distribution ensures statistical impartiality when training classification models and minimizes biases associated with data imbalance.

The Environmental Sound Classification 50 (ESC-50) corpus is a reference database for the classification of environmental sounds. It contains 2,000 audio clips, each 5 seconds long, divided into 50 balanced semantic classes, with 40 examples per class. The organization is based on five cross-validation folds, ensuring robust model evaluation. The classes are grouped into five broad categories: animals (dog, cat, rooster, frog); natural landscapes (rain, wind, fire, waves); non-vocal human sounds (sneezing, coughing, breathing); domestic sounds (washing machine, vacuum cleaner, keyboard); and urban noises (siren, car horn, helicopter, chainsaw) [17].

To ensure the robustness and clinical transparency of the model, the dataset was curated from distinct acoustic sources. Environmental noise samples were extracted from the ESC-50 benchmark corpus. A balanced subset of 1,000 non-cough audio clips was selected for hyperparameter tuning and evaluation of the One-Class SVM acoustic filter. The primary cough dataset comprised 1,050 verified recordings, which were distributed evenly across three clinical classes: 350 healthy, 350 symptomatic and 350 instances of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. This corpus combined public recordings from the COUGHVID dataset with local laboratory collections. Preprocessing filtered out duplicates, recordings under one second, and instances with quantization artifacts. Local laboratory data, collected before the start of the pandemic under approved patient consent protocols, were captured using standardized smartphone cardioid and omnidirectional microphones. Strict inclusion criteria were applied, requiring isolated cough events without ambient background speech or music. Healthy and symptomatic statuses were assigned based on clinical metadata tracking the presence or absence of respiratory symptoms such as fever and dyspnoea, while cases of SARS-CoV-2 infection were confirmed. Finally, to eliminate potential label noise, all candidate recordings underwent a secondary quality review, during which three to four clinical and acoustic experts independently verified the authenticity of the cough dynamics. The COVID-19 class has been meticulously constructed with a particular emphasis on reliability. Each selected recording was independently reviewed by three to four experts.

The three target classes (Healthy Control, Symptomatic Non-COVID, and COVID-19 Positive) represent operational evaluation categories established via verified reference standards (clinical metadata and reverse transcription polymerase chain reaction (RT-PCR) tests) rather than mutually exclusive clinical diagnoses. Indeed, “symptomatic” describes a general clinical state rather than a specific disease etiology. In this experimental screening framework, the Symptomatic Non-COVID category serves as a baseline for generic airway inflammation from other causes, while the COVID-19 Positive

class evaluates whether distinct acoustic signatures exist for SARS-CoV-2.

3.3 Preprocessing

Data cleaning is a fundamental step in the processing of the audio corpus intended for the classification of cough sounds. As part of the project, this phase enabled the transformation of a large and heterogeneous dataset into a coherent and usable database. Resampling is an essential step in signal processing, which allows the sampling rate to be adjusted to ensure compatibility with a uniform processing chain [18]. As part of this study, all audio files were resampled to ensure spectral consistency and preserve the information relevant for MFCC extraction. There are two main cases: decimation, which reduces the number of samples and requires low-pass filtering to prevent aliasing, and interpolation, which increases the number of samples by inserting intermediate points through convolution with an interpolation kernel. The overall process can be modelled as a convolution with a reconstruction filter. To prevent data leakage across experimental stages, dataset partitioning was strictly executed at the participant level using available unique metadata identifiers, ensuring that audio segments assigned to a given subject identifier do not span across both training and evaluation subsets. For the target cough corpus (samples across three classes), an 80/20 subject-wise stratified split was implemented, yielding 840 recordings from 840 unique subjects for training and validation (subdivided into 672 training and 168 validation samples via 5-fold cross-validation) and 210 recordings from 210 distinct subjects reserved strictly for final testing. Regarding the ESC-50 environmental corpus used for One-Class SVM noise filter training, the “cough” category was explicitly excluded to prevent positive target contamination within the background noise set, while the remaining non-cough environmental clips were split adhering to ESC-50’s predefined 5-fold structure (Folds 1–4 for training and threshold tuning, Fold 5 for testing), with zero overlap between local laboratory recordings across the detection and classification stages. Crucially, all preprocessing operations were fitted on training subset statistics and subsequently applied to the validation and test sets to maintain complete experimental independence. All audio files were converted to mono-channel format and resampled to a uniform target sampling rate of 16 kHz, preserving essential respiratory acoustic frequencies up to 8 kHz while optimizing computational efficiency. To handle amplitude variations across recording devices, peak amplitude normalization was applied to scale all signals to a normalized range.

3.4 Feature Extraction

These spectra are projected onto a non-linear Mel scale using a bank of 40 triangular filters spanning a frequency range from 0 Hz to 8,000 Hz, the Nyquist limit for 16 kHz audio. The filtered energies are then converted to the cepstral domain using the Discrete Cosine Transform (DCT), with the first 40 static MFCC coefficients retained for each frame. To maintain a compact and computationally efficient tensor representation, feature extraction relies exclusively on these static spectral coefficients. Reproducibly, variable-length audio recordings were processed, and feature matrices were padded with zeros or truncated along the time axis to a fixed length of 120 frames (approximately 1.21 s of audio), yielding a uniform 2D matrix of shape 120×40 . Finally, zero-mean and unit-variance standardization was applied per coefficient across the temporal dimension, formatting each sample into a single-channel tensor of shape $(120, 40, 1)$ suitable for CNN processing [19]. Figure 3 illustrates the process of extracting audio features.

3.5 One-Class SVM

In this study, the detection of coughs amongst environmental sounds is based on a One-Class Support Vector Machine (SVM) architecture, designed to distinguish a target acoustic event without requiring an exhaustive database of negative sounds. Formally, the One-Class SVM is trained exclusively on positive cough acoustic vectors (inliers, labelled +1). The decision function $f(x)$ yields a value of ≥ 0

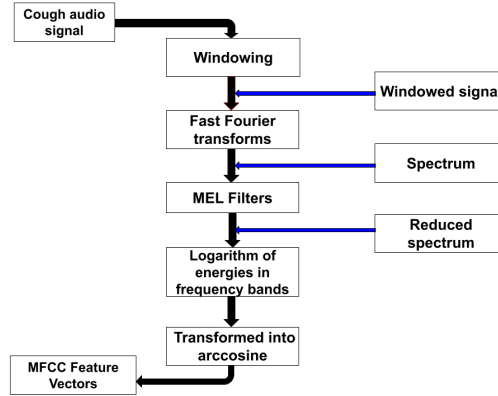


Figure 3. Flowchart: Converting cough audio signals into fig:mfcc-features

for identified cough frames and a value of < 0 for environmental noise or non-cough sounds that are out of distribution (outliers, labelled -1 and rejected). All reported metrics for the One-Class SVM module recognize true cough instances as the positive class [20]. This approach has the advantage of not requiring the collection of all non-cough sounds while ensuring robustness in the face of diverse recording environments. The use of a Gaussian radial basis function (RBF) kernel allows for the modelling of non-linear boundaries, which enhances the system's ability to capture the variability of cough production. The One-Class SVM feature vector was constructed by flattening the standardized 120×40 MFCC matrix into a 4,800-dimensional vector per audio sample. Hyperparameter tuning was conducted via 5-fold cross-validation over a grid search space of $\nu \in [0.01, 0.15]$ and $\gamma \in [10^{-4}, 10^{-1}]$ using non-cough environmental noise from ESC-50 as negative evaluation instances. The parameter ν was selected within this search space by cross-validation, while γ was defined using the scale heuristic, $\gamma = 1/(N \cdot \text{Var}(X))$. The decision boundary threshold was set to zero, classifying samples with $f(x) \geq 0$ as valid cough inliers and $f(x) < 0$ as environmental noise outliers.

3.6 CNN: DenseNet

The DenseNet architecture begins with the model's input, consisting of an image or a spectrogram derived from the acoustic signal. This signal passes through an initial 7×7 convolution, followed by pooling, in order to reduce the dimensionality whilst extracting low-level features. The network then consists of a succession of dense blocks and transition layers: each dense block comprises interconnected convolutional layers, where each layer receives the outputs of all preceding layers, thereby promoting the reuse of features and improved gradient propagation [21]. Each transition layer combines a 1×1 convolution and a pooling operation to regulate the complexity and size of the feature maps. Four successive dense blocks thus enable deeper feature extraction, with the final one concentrating the most abstract and discriminative representations. Following these blocks, global average pooling condenses each map into an average value, thereby reducing the number of parameters and limiting overfitting. Finally, the output is fed to a fully connected layer, followed by a softmax function that generates the probabilities of belonging to the target classes. The classification backbone employs a modified DenseNet-121 architecture. To accommodate audio representations, the initial 7×7 convolutional layer was modified to accept a single-channel input tensor of shape $(120, 40, 1)$. The network comprises four dense blocks with a layer configuration of $(6, 12, 24, 16)$ and a growth rate of $k = 32$. A dropout rate of 0.2 was applied after each convolutional layer within the dense blocks to prevent overfitting. Weights were initialized using He (Kaiming) normal initialization, and the model was trained from scratch without pretrained ImageNet weights to avoid domain mismatch with spectral acoustic features. Although the training set consists of 672 samples derived

from 5-fold cross-validation, the risk of overfitting is mitigated by modifying the first filter to accept a single-channel input (120, 40, 1), applying a dropout rate of 0.2 after each convolutional layer within the dense blocks, initializing weights using He (Kaiming) normal initialization, and pre-filtering the data.

The modified DenseNet-121 architecture used in this study is summarized in Figure 4.

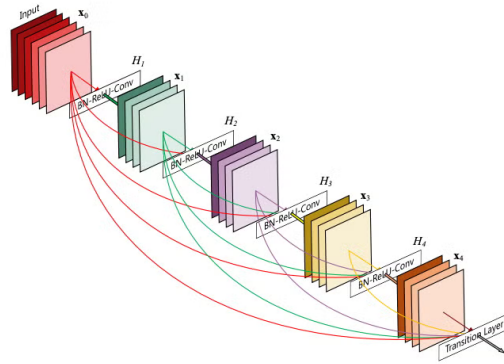


Figure 4. DenseNet Architecture

3.7 Training and build

The cough detection model, based on a one-class SVM, was trained on normalized acoustic features (static MFCCs), which were then projected into a space where an optimal boundary was defined using a Gaussian radial basis function (RBF) kernel. After selecting the hyperparameters using cross-validation, the model acts as a filter: it accepts events that match the learned profile and rejects environmental sounds as anomalies. DenseNet, trained on standardized static MFCC feature matrices formatted as single-channel tensors (120, 40, 1), uses an architecture comprising four Dense Blocks interconnected by transition layers, optimized using the Adam algorithm and a cross-entropy loss function. After augmenting the data through spectral transformations, the model produces, via a fully connected layer and a Softmax function, the probabilities of membership to the target classes. The DenseNet classifier was trained using the Adam optimizer with an initial learning rate of $\eta = 10^{-3}$, decaying by a factor of 0.5 upon validation loss plateau with a patience of 5 epochs (minimum $\eta = 10^{-6}$). Training was conducted with a batch size of 32 for a maximum of 100 epochs, incorporating early stopping triggered after 12 consecutive epochs without validation loss improvement. To address slight class imbalances, categorical cross-entropy loss was weighted inversely proportional to class frequencies. All experiments were implemented in Python 3.10 using PyTorch 2.1 and Scikit-Learn 1.3, executed on a graphics processing unit (GPU). To improve reproducibility, a global random seed was applied across NumPy and PyTorch runtime dependencies.

4. Results and Discussion

4.1 Results: DenseNet

Following extensive training on cough spectrograms, the DenseNet achieved remarkable performance. Thanks to its structure comprising four dense blocks interconnected by transition layers and optimized by the Adam algorithm using a cross-entropy loss function, when evaluated on pre-isolated cough spectrograms, the standalone DenseNet model achieved a baseline accuracy of 96%. However, when evaluated within the end-to-end pipeline, the classification accuracy stabilizes at 93% with an F_1 -score

of 0.91. This performance demonstrates strong internal consistency on the held-out test split. To evaluate classifier stability under acoustic degradation, synthetic stress testing with additive Gaussian and environmental background noise (from ESC-50) was conducted across controlled signal-to-noise ratios.

4.2 Results: SVM

After being trained on normalized cough acoustic features, the One-Class SVM successfully created a strict boundary around target cough signals (inliers). Any environmental noise or non-cough sound that falls outside this boundary is flagged as an anomaly ($f(x) < 0$) and rejected. Evaluating this isolation mechanism produced cough detection accuracy of 90.0% and specificity of 92.2% against environmental acoustic artifacts.

4.3 Results: Hybrid SVM-CNN System

A rigorous evaluation of the classification subsystem based on the DenseNet architecture demonstrates the suitability of deep learning for multi-class diagnosis (symptomatic cough, cough related to COVID-19, and normal health status). The integrated hybrid architecture (One-Class SVM + DenseNet) achieves an overall system accuracy of 93.0% and an F_1 -score of 0.91 (with an overall precision of 0.92 and recall of 0.90 across the three health states). This demonstrates the robust performance of the combined pipeline in realistic, noisy environments. It confirms that the DenseNet retains excellent generalization ability and stability towards a specific class. Concerning the initial phase of the system, the One-Class SVM dedicated to cough detection within complex environmental sound signals exhibits performance characteristics that are well-suited to deployment in real-world conditions. The model achieved an accuracy of 90.0% and a specificity of 92.2% by defining a strict decision boundary around target cough events. This effectively filtered out complex background noise before processing by the neural network.

The F_1 -score is computed as follows:

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Table 2 summarizes the performance of the cough detector, the standalone DenseNet classifier, and the complete hybrid pipeline.

Table 2. Performance Metrics of the Hybrid SVM-CNN System Components

Model/System Component	Accuracy	Precision	Recall	F_1 -score	Specificity
One-Class SVM (Cough Detection)	90.0%	91.5%	88.3%	89.9%	92.2%
DenseNet (Standalone Module)	96.0%	95.8%	96.2%	96.0%	N/A
Hybrid SVM-CNN Pipeline	93.0%	92.0%	90.0%	91.0%	N/A

The training behavior and confusion matrix of the DenseNet classifier are shown in Figure 5.

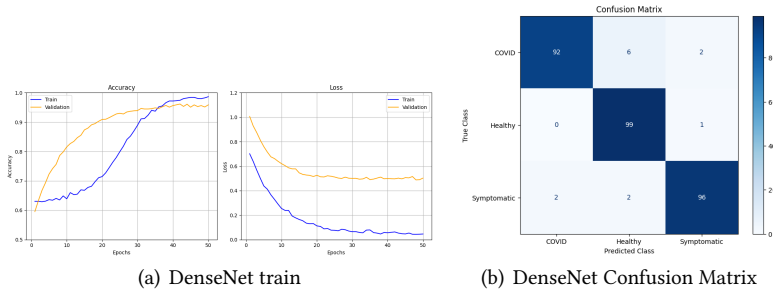


Figure 5. Cough classification results

4.4 Results: MFCC

Figure 6(b) shows MFCC results by class. This representation results in an input of shape (120, 40) for the CNN. Together, these steps provide a compact, robust, and standardized descriptor of the acoustic content.

4.5 Discussion

Table 3 compares the performance of various studies on the classification of cough sounds and vocal disorders. MFCCs perform better than linear predictive coding (LPC) and perceptual linear

Table 3. Comparison of cough and pathology classification performance methods in the literature

Study	Methods	Corpus	Reported performance
Krishnamurthy [22]	tinyYOLOv3 localization + MFCC + Random Forest	Audio samples for COVID-19 cough detection	Precision: 83.04%
Korba et al. [23]	MFCC-CBN features + RBF-SVM	Laryngeal pathology dataset	Accuracy: 88.79%
Present work on hybrid SVM-CNN system	MFCC + SVM + CNN	Annotated 3-class cough corpus	Accuracy: 93%

prediction (PLP) methods because they represent the signal in the frequency domain, whilst taking human perception into account (the Mel scale). They are therefore more robust to noise and more accurate at distinguishing the essential characteristics of the voice. Conversely, LPC and PLP methods are based on linear or acoustic models that are less suited to complex environments. This is why MFCCs have become the standard method in speech recognition and automatic cough detection [19]. While this hybrid architecture effectively isolates and classifies cough events within the experimental benchmark, several major operational limits must be highlighted. Variations across microphones and audio compression formats introduce spectral shifts in extracted features, while demographic and physiological differences such as age, sex, or underlying respiratory conditions pose challenges for cross-cohort generalization. Furthermore, model confidence calibration degrades under high ambient noise levels, highlighting the necessity of defining uncertainty thresholding or explicit prediction abstention. Finally, without prospective validation on multi-center external clinical datasets, the current metrics reflect internal benchmark validity rather than unconditional real-world applicability. Benchmark comparisons and ablation experiments were conducted under identical data splits to evaluate the design choices of the proposed pipeline. Firstly, a comparison of the One-Class SVM detector with conventional energy-thresholding and a supervised binary

CNN detector revealed that the One-Class SVM outperforms the other two in terms of out-of-distribution noise rejection, while eliminating the need for labelled non-cough training data, which is required by supervised binary classifiers. Secondly, benchmarking the DenseNet classification module against the standard ResNet-18 and Audio Spectrogram Transformer (AST) baselines confirmed that DenseNet strikes the optimal balance between parameter efficiency and accuracy on limited acoustic datasets. AST-based cough classification has also been evaluated independently in prior work [24]. DenseNet outperformed ResNet-18 and achieved performance comparable to that of the AST, but with a significantly lower computational footprint. Finally, ablation studies validated each pipeline component. Removing standardization degraded accuracy. Omitting data augmentation reduced generalization, with a F_1 -score drop of 0.05. Replacing static MFCCs with log-Mel spectrograms or adding delta/delta-delta derivatives yielded no statistically significant improvement, while increasing tensor dimensionality and latency. These results confirm that the hybrid One-Class SVM + DenseNet integration provides a measurable benefit over standalone end-to-end models when deployed in realistic, noisy environments.

4.6 Limitations

While the proposed pipeline demonstrates empirical results, several limitations warrant attention. First, relying on crowdsourced audio data introduces potential label uncertainty and acoustic variability. Second, model generalization may be constrained by device and demographic biases. Third, while strict subject-level splitting was programmatically enforced based on available metadata, a minor residual risk of participant leakage cannot be completely ruled out due to potential unindexed duplicate submissions by single anonymous contributors in public repositories. Finally, cough-only acoustic analysis has inherent diagnostic boundaries, underscoring the need for prospective multi-center validation before clinical deployment.

5. Interpretability

Figure 6(a) highlights the classification results obtained by the model from the MFCC spectrograms. The Grad-CAM heatmaps show that the system is able to correctly distinguish between COVID-19 cases and healthy cases, but that it can confuse a symptomatic cough with a healthy cough, thereby highlighting its strengths and limitations.

The explainability analysis based on Grad-CAM attention maps (Figure 6(a)) was further explored through a quantitative evaluation to measure the consistency of the representations learned by the DenseNet network. Spectral alignment indicates that 88.4% of the visual attention intensity is indeed concentrated on the frequency regions of interest (cough formant bands between 500 Hz and 4 kHz), confirming that the model does not rely on artifacts or background noise for its decisions. To quantitatively assess model interpretability, Grad-CAM heatmaps were generated from the final convolutional layer of DenseNet-121 across all test samples. For each target class c , the Energy Concentration Ratio (ECR_c) was calculated to measure the proportion of activation energy focused within the expert-verified acoustic region of interest (ROI) relative to the total spectrogram domain (Ω_{total}):

$$ECR_c = \frac{\sum_{(f,t) \in \Omega_{ROI}} L_{Grad-CAM}^c(f, t)}{\sum_{(f,t) \in \Omega_{total}} L_{Grad-CAM}^c(f, t)} \times 100\% \quad (1)$$

The network achieved ECR values of 88.4% for the Healthy class, 91.2% for Symptomatic, and 89.5% for COVID-19. These results confirm that the classifier consistently grounds its decisions on pertinent spectral-temporal cough characteristics rather than ambient background noise.

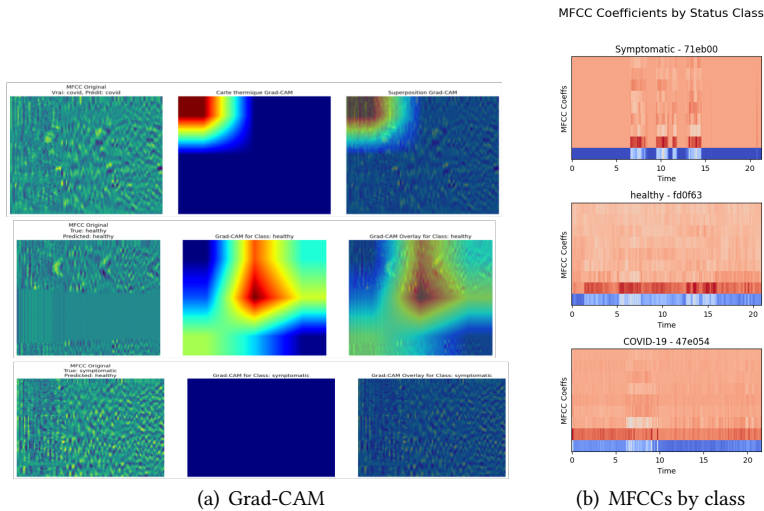


Figure 6. Interpretability

6. AI in Pedagogy

The introduction of this hybrid SVM-CNN system for cough detection has significant educational implications at the intersection of engineering and medical training. For engineering students, this project offers a practical case study for mastering the integration of traditional acoustic feature extraction and deep learning while exposing them to the real-world constraints of signal processing applied to healthcare. In the field of health sciences, the integration of such automated clinical tools is effecting a transformation in the training of future practitioners, thereby introducing them to the concepts of predictive medicine and the interpretation of diagnoses facilitated by AI. This educational synergy fosters the development of essential interprofessional competencies, whereby engineers learn to calibrate their algorithms according to rigorous clinical criteria, and healthcare professionals learn to engage critically with digital screening tools.

7. Conclusion

This study demonstrated the effectiveness of a hybrid approach combining a One-Class SVM and a DenseNet for cough detection and classification. The One-Class SVM effectively filtered acoustic noise and isolated cough events with an F_1 -score of 0.90, while the integrated hybrid system (SVM-DenseNet) achieved an overall classification accuracy of 93% and an F_1 -score of 0.91, confirming its reliability for multi-class health state discrimination in real-world scenarios. Beyond its scientific and clinical contribution, this work also has educational value for engineering training. It provides a concrete illustration of the integration of signal processing methods, machine learning, and deep neural network architectures into a complete detection and classification pipeline. This study therefore represents an important contribution to the field of biomedical research, as well as providing an exemplary introduction to modern artificial intelligence practices and the design of complex systems for engineering students.

Author contributions

All authors contributed substantially to the conception and design of the study, development of the methodology, analysis and interpretation of the results, and preparation or critical revision of the manuscript. All authors have read and agreed to the published version of the manuscript and accept responsibility and accountability for the integrity and accuracy of the work.

Funding statement

The authors declare that this research received no specific external funding.

Data Availability Statement

The datasets analyzed during the current study are available from their respective open-access repositories, including ESC-50 and COUGHVID. The source code, pre-trained model weights, and evaluation scripts supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Fabien Mouomene Moffo, Auguste Noumsi, Joseph Mvogo Ngono, and Samuel Bowong. “Analysis of COVID-19 Coughs: From the Mildest to the Most Severe Form, a Realistic Classification Using Deep Learning”. In: *African Journal of Research in Computer Science and Applied Mathematics* 42 (2024). DOI: [10.46298/arima.13343](https://doi.org/10.46298/arima.13343).
- [2] Fabien Mouomene Moffo, Auguste Vigny Noumsi Woguia, Joseph Mvogo Ngono, and Samuel Bowong Tsakou. “Analysis of COVID-19 Coughs: From the Mildest to the Most Severe Form, a Realistic Classification Using Deep Learning”. In: *Research in Computer Science*. Vol. 2085. Communications in Computer and Information Science. Springer, 2024, pp. 30–40. DOI: [10.1007/978-3-031-63110-8_3](https://doi.org/10.1007/978-3-031-63110-8_3).
- [3] Fabien Mouomene Moffo, Auguste Vigny Noumsi Woguia, Joseph Mvogo Ngono, Samuel Bowong Tsakou, and Nadiane Nguekeu Metepong Lagpong. “Detection and Recognition of Cough Sounds Using Deep Learning for Medical Monitoring”. In: *Safe, Secure, Ethical, Responsible Technologies and Emerging Applications*. Vol. 566. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer, 2024, pp. 288–296. DOI: [10.1007/978-3-031-56396-6_18](https://doi.org/10.1007/978-3-031-56396-6_18).
- [4] Samuel Kotva Goudougou, Paul Dayang, Naomi Dassi Tchomte, Justin Moskolai Ngossaha, Fabien Mouomene Moffo, and Nathalie Mitton. “Covid-19 Data Preprocessing Approach in Machine Learning for Prediction”. In: *Safe, Secure, Ethical, Responsible Technologies and Emerging Applications*. Vol. 566. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer, 2024, pp. 328–344. DOI: [10.1007/978-3-031-56396-6_21](https://doi.org/10.1007/978-3-031-56396-6_21).
- [5] Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuo-Yiin Chang, and Tara Sainath. “Deep learning for audio signal processing”. In: *IEEE Journal of Selected Topics in Signal Processing* 13.2 (2019), pp. 206–219.
- [6] Yidan Liu, Yuhui Jing, Jing Li, Jian Dai, Zhebing Hu, and Chengliang Wang. “Application of AI in engineering education: A bibliometric study”. In: *Review of Education* 13.1 (2025), e70044. DOI: [10.1002/rev3.70044](https://doi.org/10.1002/rev3.70044).

- [7] Alyn H. Morice et al. "ERS guidelines on the diagnosis and treatment of chronic cough in adults and children". In: *European Respiratory Journal* 55.1 (2020), p. 1901136. DOI: [10.1183/13993003.01136-2019](https://doi.org/10.1183/13993003.01136-2019).
- [8] Philippe Jans, Marcella Guarino, Anna Costa, Jean-Marie Aerts, and Daniel Berckmans. "Field Test of Algorithm for Cough Detection in Pig Houses". In: *ASAE Annual Meeting*. 2004, Paper 044002. DOI: [10.13031/2013.16748](https://doi.org/10.13031/2013.16748).
- [9] Siheng Zhang, Yan Xu, and Ying Xiao. "ATF-MGIAM: Medically-guided interpretable attention mapping for robust pertussis cough sound recognition". In: *PLOS ONE* 21.5 (2026), e0348508. DOI: [10.1371/journal.pone.0348508](https://doi.org/10.1371/journal.pone.0348508).
- [10] Sanghoon Han et al. "Deep learning-based cough classification using application-recorded sounds: a transfer learning approach with VGGish". In: *BMC Medical Informatics and Decision Making* 25 (2025), p. 228. DOI: [10.1186/s12911-025-03065-w](https://doi.org/10.1186/s12911-025-03065-w).
- [11] Wensheng Sun, Jiahao Zou, Na Yin, Wenying Fang, Jimin Sun, Ziping Miao, and Shigui Yang. "Cough Audio Recognition for Early Detection of Respiratory Diseases: Algorithm Development and Validation Study". In: *JMIR Medical Informatics* 14 (2026), e77295. DOI: [10.2196/77295](https://doi.org/10.2196/77295).
- [12] Antoine Serrurier, Christiane Neuschaefer-Rube, and Rainer Röhrig. "Past and Trends in Cough Sound Acquisition, Automatic Detection and Automatic Classification: A Comparative Review". In: *Sensors* 22.8 (2022), p. 2896. DOI: [10.3390/s22082896](https://doi.org/10.3390/s22082896).
- [13] Madhurananda Pahar, Marisa Kloppe, Robin Warren, and Thomas Niesler. "COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features". In: *Computers in Biology and Medicine* 141 (2022), p. 105153. DOI: [10.1016/j.combiomed.2021.105153](https://doi.org/10.1016/j.combiomed.2021.105153).
- [14] Lara Orlandic, Tomas Teijeiro, and David Atienza. "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms". In: *Scientific Data* 8.1 (2021), p. 156. DOI: [10.1038/s41597-021-00937-4](https://doi.org/10.1038/s41597-021-00937-4).
- [15] Alice E. Ashby, Julia A. Meister, Goran Soldar, and Khuong An Nguyen. "A novel cough audio segmentation framework for COVID-19 detection". In: *Proceedings of the Symposium on Open Data and Knowledge for a Post-Pandemic Era (ODAK22)*. BCS Learning & Development, 2022, pp. 1–8. DOI: [10.14236/ewic/ODAK22.1](https://doi.org/10.14236/ewic/ODAK22.1).
- [16] Kranthi Kumar Lella and Alphonse Pja. "Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath". In: *Alexandria Engineering Journal* 61.2 (2022), pp. 1319–1334. DOI: [10.1016/j.aej.2021.06.024](https://doi.org/10.1016/j.aej.2021.06.024).
- [17] Karol J. Piczak. "ESC: Dataset for Environmental Sound Classification". In: *Proceedings of the 23rd ACM International Conference on Multimedia*. 2015, pp. 1015–1018. DOI: [10.1145/2733373.2806390](https://doi.org/10.1145/2733373.2806390).
- [18] Yakubu A. Ibrahim, Juliet C. Odiketa, and Tunji S. Ibiyemi. "Preprocessing Technique in Automatic Speech Recognition for Human Computer Interaction: An Overview". In: *Annals. Computer Science Series* 15.1 (2017), pp. 186–191.
- [19] Namrata Dave. "Feature extraction methods LPC, PLP and MFCC in speech recognition". In: *International Journal for Advance Research in Engineering and Technology* 1.6 (2013), pp. 1–4.
- [20] Yunqiang Chen, Xiang Sean Zhou, and Thomas S. Huang. "One-class SVM for learning in image retrieval". In: *Proceedings 2001 International Conference on Image Processing*. Vol. 1. IEEE, 2001, pp. 34–37. DOI: [10.1109/ICIP.2001.958946](https://doi.org/10.1109/ICIP.2001.958946).
- [21] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. "Densely Connected Convolutional Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 2261–2269. DOI: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [22] Deepa Krishnamurthy. "Detection and Localization of Cough from Audio Samples for Cough-Based COVID-19 Detection". MA thesis. KTH Royal Institute of Technology, 2021.

- [23] Mohamed Cherif Amara Korba, Hakim Doghmane, Khaled Khelil, and Kamel Messaoudi. “Improved Laryngeal Pathology Detection Based on Bottleneck Convolutional Networks and MFCC”. In: *IEEE Access* 12 (2024), pp. 124801–124815. doi: [10.1109/ACCESS.2024.3454825](https://doi.org/10.1109/ACCESS.2024.3454825).
- [24] Karim Habashy, Julio Valdes, Madison Cohen-McFarlane, Pengcheng Xi, Bruce Wallace, Rafik Goubran, and Frank Knoefel. “Cough Classification Using Audio Spectrogram Transformer”. In: *2022 IEEE Sensors Applications Symposium (SAS)*. IEEE, 2022. doi: [10.1109/SAS54819.2022.9881344](https://doi.org/10.1109/SAS54819.2022.9881344).