Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Minimal-Preprocessing Paradigm for Robust COVID-19 Detection from Cough Audio Using Deep Learning

Chee Chin Lim

cclim@unimap.edu.my

Universiti Malaysia Perlis

Leow Bin Toh

Taylor's University

Megat Syahirul Amin Megat Ali

Universiti Teknology MARA

Jason Teh Lei Yik

Universiti Malaysia Perlis

Article

Keywords:

Posted Date: May 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9394179/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 2nd, 2026. See the published version at https://doi.org/10.1038/s41598-026-68570-y.

Minimal-Preprocessing Paradigm for Robust COVID-19 Detection from Cough Audio Using Deep Learning

Chee Chin Lim[1,*], Leow Bin Toh[2,*] , Megat Syahirul Amin Megat Ali[3], Jason Teh Lei Yik[1]

[1] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia.

[2] Faculty of Innovation and Technology, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia.

[3] Microwave Research Institute, Universiti Teknology MARA, Selangor, Malaysia.

* Corresponding authors: Chee Chin Lim (cclim@unimap.edu.my), Leow Bin Toh (leowbin.toh@taylors.edu.my).

Abstract

The COVID-19 pandemic has highlighted the need for rapid, accessible, and non-invasive screening tools to complement conventional diagnostic methods. While audio-based analysis of cough signals has emerged as a promising approach, most existing studies rely on complex preprocessing pipelines aimed at noise suppression and signal enhancement. However, the impact of such preprocessing on model generalisation remains insufficiently understood. This study challenges the prevailing assumption that extensive preprocessing is necessary for robust audio-based diagnosis. We propose a systematic evaluation of preprocessing strategies, investigating whether minimal preprocessing can preserve diagnostically relevant acoustic information and improve model performance. Cough recordings were transformed into time–frequency representations using short-time Fourier transform (STFT) and Mel-frequency cepstral coefficients (MFCCs) and classified using multiple convolutional neural network (CNN) architectures, including ResNet, Inception, and MobileNet variants. Experimental results demonstrate that minimal preprocessing using Fast Fourier Transform (FFT) without additional filtering consistently outperforms conventional noise-reduction techniques. The best-performing model, ResNet50V2, achieved a testing accuracy of 86.79%, with balanced precision, sensitivity, specificity, and F1-score. Furthermore, the combination of STFT and MFCC features provided complementary information, improving classification robustness. These findings indicate that preserving full spectral information is critical for effective deep learning–based audio diagnosis, and that aggressive preprocessing may inadvertently suppress subtle pathological signatures. This work establishes a minimal-preprocessing paradigm for cough-based COVID-19 detection and provides practical insights for developing robust, scalable, and non-invasive diagnostic systems.

Introduction

The COVID-19 pandemic has had a profound impact on various aspects of people's lives worldwide, including their lifestyle, economy, social interactions, and education. Given its high contagiousness, the symptoms of COVID-19 can vary significantly among individuals, often taking approximately 14 days to manifest. Recent data from the last 28 days (8th May to 4th June 2023) reveal approximately 1.7 million new COVID-19 cases and 10,000 deaths worldwide, indicating a 38% decrease in cases and a 47% decrease in deaths compared to the previous 28 days. Reverse transcription polymerase chain reaction (RT-PCR) testing remains the gold standard for SARS-CoV-2 detection; however, it is time-consuming, costly, and requires trained healthcare professionals. Additionally, discomfort may be experienced by certain individuals, especially children.

Governments worldwide have provided free RT-PCR tests to combat COVID-19, but the costs are substantial and average $23 per test. Researchers and medical practitioners propose alternative, cost-effective, and user-friendly methods. One method involves using human audio signals, particularly coughing sounds, to detect COVID-19, offering a potential solution to the limitations of RT-PCR testing. This is crucial as the increasing number of cases demands more precise, affordable, and efficient testing methods.

Accordingly, this study addresses the following research questions:

(1) Can convolutional neural network-based models reliably differentiate between COVID-19-positive and COVID-19-negative individuals using cough sound recordings?

(2) How do different preprocessing and feature extraction strategies influence the generalisation performance of CNNs in cough-based COVID-19 detection?

(3) Does the choice of CNN architecture significantly affect classification accuracy?

(4) Does the integration of patient demographic data improve diagnostic performance compared to audio-only models?

Based on these questions, the study tests the hypotheses that CNNs trained on cough sound features can achieve classification performance above baseline levels, that spectro-temporal feature representations enhance model generalisation, that deeper CNN architectures outperform lighter models, and that incorporating demographic information improves classification accuracy. While prior studies emphasize noise reduction and signal conditioning, this work challenges the necessity of complex preprocessing pipelines. We hypothesize that excessive filtering may suppress subtle yet diagnostically relevant acoustic biomarkers. Therefore, this study introduces a minimal-preprocessing paradigm, investigating whether preserving raw spectral information can improve generalization and robustness in cough-based COVID-19 detection

Literature Survey

Recent years have witnessed growing interest in audio-based diagnostic systems for COVID-19, driven by the need for scalable, non-invasive, and cost-effective screening tools. Cough sounds, in particular, have been shown to encode physiological changes associated with respiratory infections, motivating extensive research into machine learning–based cough analysis. While numerous studies report promising results, critical limitations remain in terms of robustness, generalizability, and clinical readiness.

Audio Preprocessing and Signal Variability

Most existing studies employ preprocessing pipelines involving noise reduction, segmentation, and data augmentation to mitigate the effects of environmental noise and device heterogeneity. Techniques such as spectral gating, Butterworth filtering [6], and neural-network-based denoising have been shown to improve signal-to-noise ratios and classification performance. However, models trained under controlled conditions often degrade significantly when exposed to real-world recordings collected using heterogeneous devices and environments.

Although segmentation of cough events improves interpretability and reduces irrelevant acoustic content, the lack of standardized segmentation protocols introduces additional variability across studies. Moreover, aggressive preprocessing and augmentation strategies risk suppressing subtle disease-related biomarkers, raising concerns about over-optimisation and dataset dependency.

Feature Extraction Strategies: Strengths and Limitations

Handcrafted acoustic features such as Short-time Fourier transform (STFT), Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Mel-frequency Cepstral Coefficients (MFCCs), Chroma, and spectral contrast remain popular due to their interpretability and effectiveness on small datasets. Nevertheless, multiple comparative studies highlight their limited robustness to noise and recording variability. Deep learning–based feature extraction from time–frequency representations, particularly CNN-derived embeddings from spectrograms, has shown strong capability in modelling complex spectral–temporal characteristics of cough sounds. Existing studies employ architectures such as 1D-CNNs, CRNNs, and deep CNN models (e.g., ResNet, Inception, VGGish), alongside audio texture descriptors including local binary or ternary patterns, fractal dimension–based features, and biologically inspired filter banks, to capture fine-scale acoustic structure associated with respiratory conditions. However, these methods are highly dependent on dataset size, class balance, and preprocessing design, and often sacrifice interpretability, motivating the present study's systematic, signal-processing-driven evaluation of spectro-temporal features combined with carefully selected CNN architectures to improve robustness and generalisation.

Hybrid approaches that fuse handcrafted features with deep learning embeddings have emerged as a promising compromise, offering improved generalization while retaining partial interpretability. Scopus reports indicate that such hybrid representations consistently outperform single-feature strategies, particularly in cross-dataset evaluations, suggesting that complementary acoustic cues are critical for reliable COVID-19 detection.

CNN Architecture and Optimization Challenges

Convolutional neural networks, including ResNet, EfficientNet, and CNN–LSTM hybrids, dominate the current literature. While deeper architectures frequently report high accuracy, these gains often come at the expense of increased computational complexity and reduced robustness. Notably, many studies fail to systematically examine the impact of architectural optimisation, training dynamics, and hyperparameter selection on generalization performance.

A recurring limitation identified across Scopus-indexed studies is the tendency to report performance on a single dataset, without cross-dataset or external clinical validation. This practice inflates reported accuracy and obscures the true reliability of the models. Consequently, the absence of optimised CNN frameworks explicitly designed to balance accuracy, generalization, and deployability remains a critical research gap.

Generalization, interpretability, and clinical readiness

Despite encouraging experimental results, the transition from research prototypes to clinically viable tools remain limited. Dataset bias, class imbalance, and evolving virus variants continue to undermine model stability. Furthermore, the opaque nature of deep CNN models presents a significant barrier to clinical adoption. While explainable AI techniques such as Grad-CAM and attention mechanisms have been explored, their integration into cough-based diagnostic systems is still nascent.

Scopus literature consistently emphasizes that diagnostic accuracy alone is insufficient for clinical acceptance. Models must demonstrate robustness across populations, transparency in decision-making, and feasibility for real-time deployment on resource-constrained devices.

In summary, existing research confirms the feasibility of audio-based COVID-19 detection but reveals a lack of systematically optimised CNN frameworks that address preprocessing variability, feature robustness, architectural efficiency, and generalization simultaneously. Table 1 provides a critical synthesis of Scopus-indexed studies between 2020 and 2025 on cough-based COVID-19 detection, highlighting persistent limitations in generalization, preprocessing standardization, and CNN optimisation. Despite high reported accuracies, most existing systems lack robustness and deployment readiness, motivating the optimised CNN framework proposed in this study. This gap motivates the development of an optimised convolutional neural network tailored for accurate and reliable COVID-19 diagnosis from acoustic cough signatures.

Study Category	**Representative Methods (Scopus)**	**Key Strengths**	**Critical Limitations**	**Identified Research Gaps**
Data Sources	COUGHVID, Coswara, DiCOVA, COVID-19 Sounds	Large-scale crowdsourced data; rapid data collection; global reach	Severe class imbalance; self-reported labels; device and environment bias	Need for curated, clinically validated, and demographically diverse datasets
Pre-processing	Spectral gating, Butterworth filters, wavelet denoising, SpecAugment	Improved SNR; enhanced robustness under controlled noise	No standard pipeline; aggressive filtering may suppress disease biomarkers; poor cross-dataset transfer	Systematic evaluation of minimal yet robust preprocessing strategies
Segmentation	RMS thresholding, hysteresis comparator, cough-event isolation	Removes irrelevant audio; improves interpretability	Inconsistent segmentation criteria; loss of temporal cues	Unified, cough-phase–aware segmentation methods
Feature Extraction (Handcrafted)	MFCC, Chroma, ZCR, spectral contrast	Interpretable; effective on small datasets	Limited robustness to noise and device variability	Deep learning based cough detection camera, Hybrid representations combining acoustic priors with learned features
Feature Extraction (Deep)	CNN spectrogram embeddings, audio texture analysis	Captures complex non-linear biomarkers	Data-hungry; reduced interpretability	Lightweight and explainable feature learning frameworks

Classification (Traditional ML)	SVM, Random Forest, MLP, XGBoost	Efficient; good baseline generalization	Limited capacity for complex patterns	Optimised hybrid pipelines integrating ML and DL
Classification (Deep Learning)	ResNet, EfficientNet, CNN-LSTM, CRNN	High accuracy in controlled settings	Overfitting; high computational cost; dataset dependency	CNN optimisation focused on generalisation, not depth
Ensemble & Multi-modal	Model ensembles; cough + breath + speech fusion	Improved robustness and accuracy	Increased complexity; deployment challenges	Scalable, deployable ensemble strategies
Generalization & Validation	Cross-validation; limited external testing	Demonstrates feasibility	Poor real-world robustness; lack of cross-dataset evaluation	Cross-dataset and population-level validation protocols
Explainability	Grad-CAM, attention maps, LIME, SHAP	Improves trust and interpretability	Rarely integrated into final systems	Explainable CNNs for clinical acceptance
Deployment Readiness	MobileNet, lightweight CNNs	Feasible for mobile screening	Accuracy–efficiency trade-offs unresolved	Optimised CNNs balancing accuracy, robustness, and efficiency

Table 1. Study design table for the evaluation of COVID-19 cough-based classification

Methods

This study adopts a structured signal-processing and deep learning pipeline for classifying COVID-19 and non-COVID-19 cases using cough audio recordings. As illustrated in Figure. 1, the overall workflow comprises four main stages: dataset preparation and cough sound acquisition, minimal and conventional audio pre-processing, feature extraction using time–frequency representations (STFT and MFCC), and classification using convolutional neural networks (CNNs) such as ResNet50V2. Notably, the framework emphasises a minimal-preprocessing paradigm, where Fast Fourier Transform (FFT) is applied without additional filtering to preserve the full spectral characteristics of cough signals. To systematically evaluate the impact of signal conditioning on model performance, multiple preprocessing strategies were investigated and compared under consistent experimental settings, highlighting the effect of signal preservation versus noise reduction on classification robustness and generalisation.

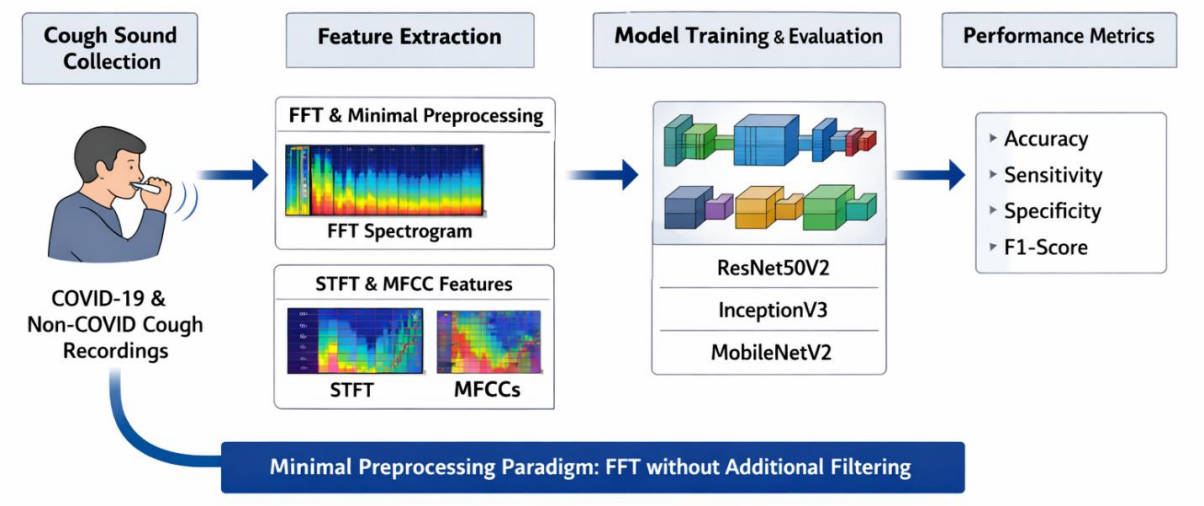


Figure. 1. Overview of methodology for minimal-preprocessing, cough-based COVID-19 classification using convolutional neural networks.

Dataset Protocol

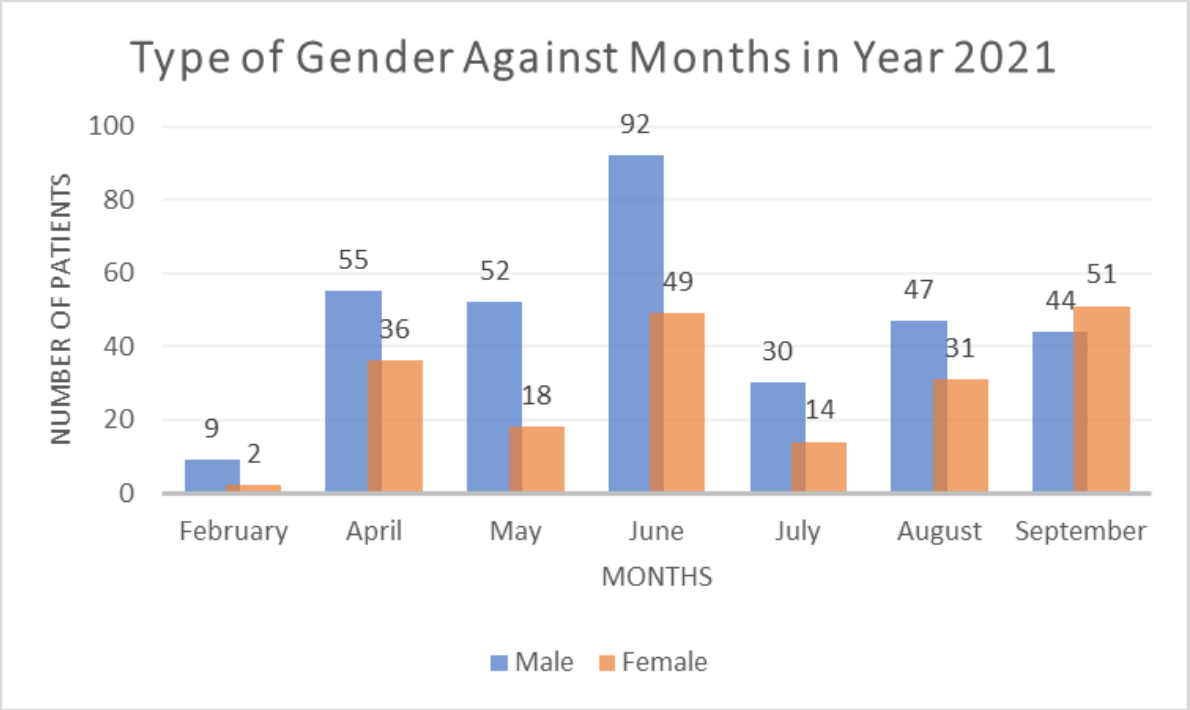
This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) License. Project Coswara, developed by the Indian Institute of Science (IISc) Bangalore, aims to develop a diagnostic tool for COVID-19 detection using audio recordings of breathing, cough, and speech sounds. The data collection stage involves participants contributing nine sound samples through the Coswara website. The dataset, collected using a specific protocol, serves as the foundation for analysis and mathematical models for COVID-19 diagnosis.

The project aims to create a comprehensive dataset for analyzing COVID-19 infection patterns in voice samples, focusing on privacy and anonymity. Participants are asked to record various audio samples, including breathing, cough, and vowel sounds. The dataset will be released on GitHub for collaborative contributions. The long-term vision is to develop a web/mobile application as a diagnostic tool, complementing existing testing methods and providing a probability score for COVID-19 infection based on recorded voice samples. This tool aims to offer a cost-effective and deployable first line of diagnosis during the pandemic.

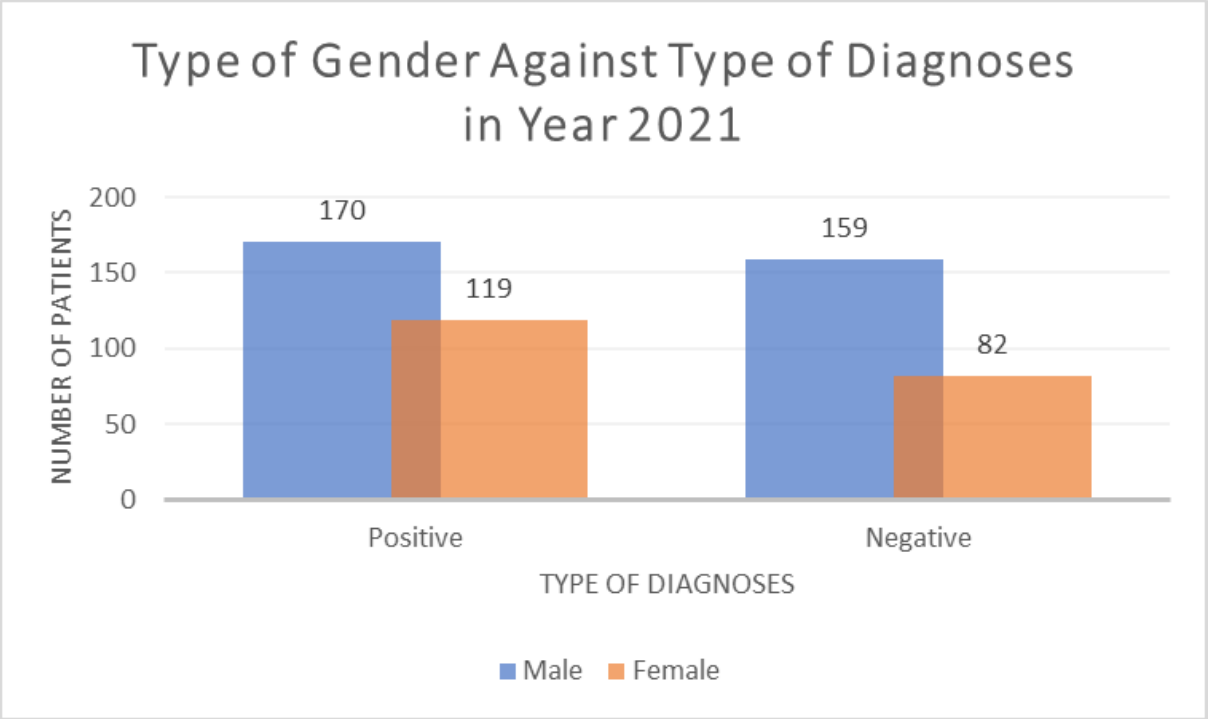
Sample Data

The project analyzes 530 patient audio recordings from the Coswara dataset, covering a seven-month period from February to September. The data includes both COVID-19 and non-COVID-19 diagnoses, with a mix of heavy and shallow cough recordings, bringing the total to 1060.

Figure 2(a) illustrates the variation in patient participation month-to-month. The lowest participation occurred in February (9 males, 2 females), while the highest was in June (141 patients, 92 males, 49 females). Participation declined significantly in July (30 males, 14 females), followed by a slight increase in August (47 males, 31 females) and September (44 males, 51 females). Figure 2(b) shows that in 2021, there were more male patients (329) than female patients (201). Among the male patients, 52% tested positive for COVID-19, while among the female patients, 59% tested positive. The remaining patients tested negative. These findings indicate a higher prevalence of COVID-19 among male patients and emphasize the importance of considering factors such as age, underlying health conditions, and exposure to the virus when analyzing the disease's spread among different populations.



(a) Type of Gender Against Months in Year 2021



(b) Type of Gender Against Type of Diagnoses in Year 2021

Figure. 2. (a) Type of Gender Against Months in Year 2021 (b) Type of Gender Against Type of Diagnoses in Year 2021

Pre-processing

Unlike conventional approaches that apply multiple layers of noise reduction, this study explicitly investigates a minimal-preprocessing strategy, where FFT is used without additional filtering to preserve the full spectral characteristics of cough signals. Audio pre-processing was conducted to improve signal quality and to assess the influence of different noise reduction and frequency enhancement techniques on classification performance. Four pre-processing strategies were evaluated: (i) Fast Fourier Transform (FFT) without additional filtering, (ii) Butterworth filtering, (iii) wavelet denoising, and (iv) pre-emphasis filtering.

Fast Fourier Transform (FFT)

Fast Fourier Transform (FFT) was applied as a standalone pre-processing approach to convert time-domain cough signals into the frequency domain. The discrete Fourier transform of an input signal x[n] is given by:

$$x[k] = \sum_{n=0}^{N-1} x\,[n] e^{\frac{-j2\pi kn}{N}} \qquad (1)$$

$X[k]$ represents the frequency component at index k, N is the total number of samples, and j is the imaginary unit. FFT enables efficient computation of spectral characteristics with a computational complexity of $O(N \log N)$, making it suitable for large-scale audio analysis. In this study, FFT was intentionally evaluated without additional filtering to examine whether minimal pre-processing could preserve diagnostically relevant acoustic features while improving model generalisation.

Butterworth Filtering

Butterworth filters were employed to investigate the effect of frequency-selective noise suppression on cough signal classification. The magnitude response of an n-order Butterworth filter is defined as

$$H(j\omega) = \frac{H_0}{\sqrt{1+\varepsilon^2(\frac{\omega s}{\omega p})^{2n}}} \qquad (2)$$

where H_0 denotes the reference gain, ε represents passband ripple, and ωs and ωp correspond to stopband and passband cutoff frequencies, respectively. Low-pass, high-pass, and band-pass configurations were evaluated using cutoff frequencies informed by prior cough-analysis studies. Lower cutoff frequencies of 45 Hz, 150 Hz, and 200 Hz were tested, while the upper cutoff frequency was fixed at 4 kHz [20] to capture dominant cough spectral content. An additional upper bound of 10 kHz was examined to assess potential high-frequency contributions.

Wavelet Denoising

Wavelet denoising was applied to suppress non-stationary noise while preserving transient cough characteristics. The denoised signal x_{denoised} is expressed as

$$x_{denoised} = W^{-1}(T(W(x))) \quad (3)$$

where $W(\cdot)$ and $W^{-1}(\cdot)$ denote the wavelet decomposition and reconstruction operators, respectively, and $T(\cdot)$ represents the thresholding function applied to wavelet coefficients. Decomposition levels of 2 and 3 were evaluated, corresponding to threshold values of 0.01 and 0.1, respectively. This approach enables selective attenuation of noise components while retaining clinically relevant time–frequency structures in cough signals.

Pre-emphasis Filtering

Pre-emphasis filtering was employed to amplify high-frequency components and compensate for spectral attenuation introduced during recording. The filter is defined as

$$y[n] = x[n] - \alpha * x[n-1]) \quad (4)$$

where $y[n]$ is the output signal, $x[n]$ is the input signal, and α is the pre-emphasis coefficient, set to 0.9 in this study. This operation enhances high-frequency information associated with cough onset and turbulence, which may carry discriminative cues for disease classification.

Feature Extraction

Feature extraction was performed to transform pre-processed cough audio signals into informative time–frequency representations suitable for deep learning–based classification. Two complementary spectral representations were investigated: Short-Time Fourier Transform (STFT) and Mel-frequency cepstral coefficients (MFCCs). These representations were selected due to their demonstrated effectiveness in capturing both transient and perceptually relevant acoustic characteristics of cough sounds.

Short-Time Fourier Transform (STFT)

Short-Time Fourier Transform (STFT) was employed to analyse the non-stationary characteristics of cough signals by representing spectral variations over time. The STFT of a continuous-time signal x(t) is defined as

$$STFT\{x(t)\} = x(\tau, w) = \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-j\omega t}\, dt \quad (5)$$

where $w(t-\tau)$is a window function centred at time τ, and ωdenotes angular frequency. By applying a sliding window to the signal, STFT enables localisation of frequency content over time, which is particularly important for cough sounds that exhibit rapid temporal variations . The resulting spectrograms were used as two-dimensional inputs to convolutional neural networks, allowing spatial feature learning across time–frequency patterns.

Mel-Frequency Cepstral Coefficients (MFCCs)

MFCCs were extracted to model perceptually relevant spectral features of cough signals. MFCCs approximate the human auditory response by mapping linear frequency components onto the Mel scale, thereby emphasising frequency regions most relevant to human perception. This representation has demonstrated effectiveness in respiratory sound analysis and cough-based disease classification.

The MFCC extraction pipeline involves segmenting the signal into overlapping frames, applying a Hamming window, computing the Fourier transform, mapping the power spectrum to the Mel scale using a filter bank, applying logarithmic compression, and finally performing a discrete cosine transform (DCT). The conversion from physical frequency $f_{Hz}$to the Mel scale $f_{Mel}$is defined as

$$f_{Mel} = 2595 * log_{10}(1 + \frac{f_{Hz}}{700}) \quad (6)$$

In this study, a Hamming window of 1024 samples with 50% overlap (512 samples) was employed, and 13 MFCCs were extracted from each frame. These parameters were selected to balance temporal resolution and spectral detail while maintaining computational efficiency. An example MFCC representation of the COVID-19 cough signal is shown in Figure 3.

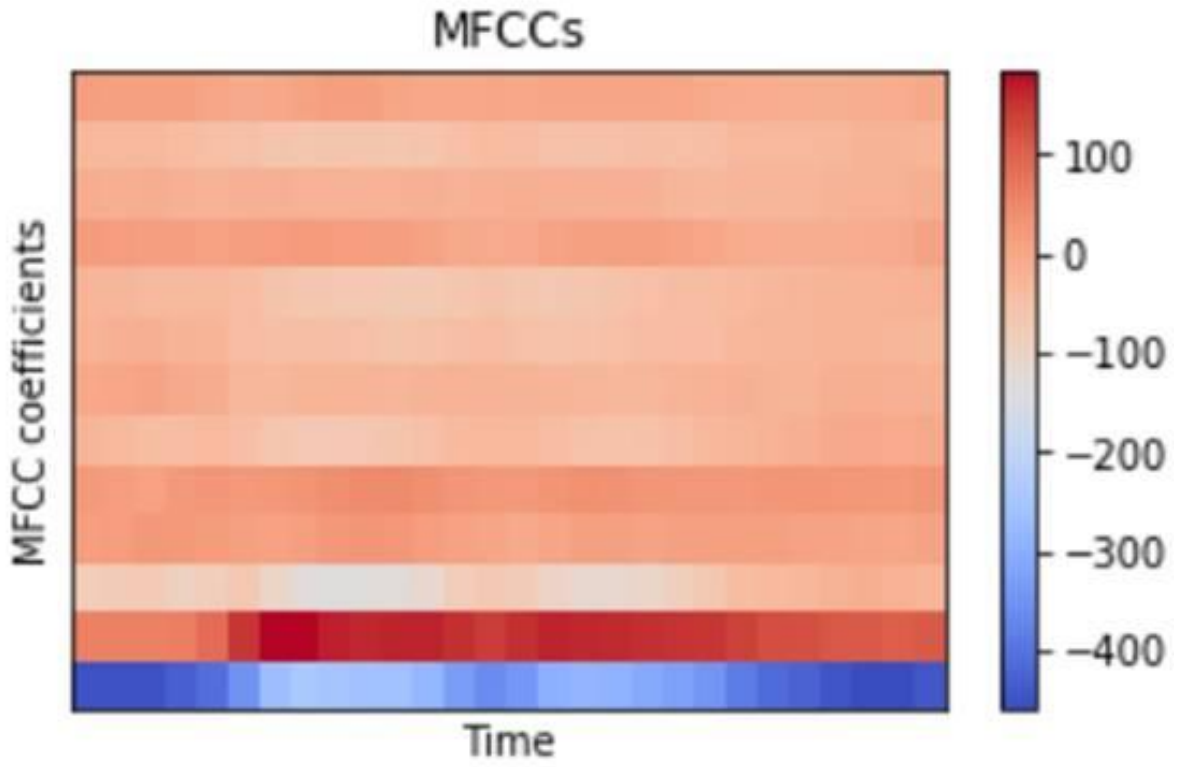


Figure. 3. Example MFCCs Image with Colorbar

Rationale for Feature Selection

STFT and MFCC representations were selected to capture complementary information from cough signals. STFT preserves fine-grained temporal–spectral dynamics, while MFCCs provide compact, perceptually motivated features that are robust to minor recording variations. Evaluating both representations enables systematic analysis of their impact on classification performance and generalisation in COVID-19 cough detection.

Deep Learning Models

Convolutional neural networks (CNNs) were employed for classifying COVID-19 and non-COVID-19 cough signals based on image-like time–frequency representations. All models were implemented using the TensorFlow–Keras framework and trained under consistent experimental conditions to enable fair comparison across architectures and configurations.

Model Input and Transfer Learning Framework

Feature representations generated from STFT and MFCC processing were resized to 224×224 pixels and treated as three-channel images. Pixel intensities were normalized to the range [0,1] to stabilise training. The dataset was partitioned into training, validation, and test subsets, and data generators were employed to load samples in batches while preserving class labels.

Batch sizes of 16, 32, and 64 were evaluated to assess their impact on convergence and generalisation. Training was conducted for 20, 50, and 100 epochs to examine learning dynamics and mitigate underfitting or overfitting.

A transfer learning strategy was adopted using pre-trained CNN backbones to leverage robust feature representations learned from large-scale image datasets. The following architectures were evaluated: InceptionV3, MobileNetV2, ResNet50, ResNet50V2, ResNet101V2, ResNet152V2.

CNN Architecture Descriptions

InceptionV3

InceptionV3 employs parallel convolutional paths with different kernel sizes within inception modules, enabling multi-scale feature extraction while maintaining computational efficiency. This architecture is effective at capturing both fine-grained and global patterns in spectrogram-based cough representations.

MobileNetV2

MobileNetV2 is a lightweight architecture designed for mobile and edge deployment. It utilises depthwise separable convolutions and inverted residual blocks with linear bottlenecks, significantly reducing parameter count and computational cost. This makes it suitable for real-time, resource-constrained COVID-19 screening applications.

ResNet Architectures (ResNet50, ResNet50V2, ResNet101V2, ResNet152V2)

ResNet models introduce residual connections that allow gradients to propagate effectively through deep networks, addressing the vanishing gradient problem. The V2 variants further improve optimisation by adopting pre-activation residual blocks, which enhance convergence and generalisation. Deeper versions (ResNet101V2 and ResNet152V2) offer increased representational capacity, while ResNet50V2 provides a balance between depth, accuracy, and computational efficiency.

For each architecture, the convolutional base was initialised with pre-trained weights and frozen during initial training to prevent disruption of learned low-level features. A custom classification head was appended, consisting of a flattening layer followed by a fully connected layer with SoftMax activation to enable binary classification.

Model Optimisation and Regularisation

Models were compiled using categorical cross-entropy loss and optimised with Adam, RMSprop, and stochastic gradient descent (SGD) optimisers. Multiple learning rates were evaluated to identify stable and efficient convergence behaviour.

To improve generalisation and reduce overfitting, L1 and L2 regularisation were applied to the classification layer, with regularisation strengths ranging from 10^{-1} to 10^{-20}. Additional architectural variations, including different pooling strategies and dense layer configurations, were explored to assess their influence on classification performance.

Training Strategy and Evaluation

Model training was monitored using validation accuracy, and the best-performing model weights were automatically saved via checkpointing. After training, the optimal model was evaluated on an unseen test dataset. Performance metrics included accuracy, precision, sensitivity (recall), specificity, and F1-score. A confusion matrix was generated to visualise class-wise prediction performance.

Results

Comparison of Pre-processing Methods

This section evaluates the impact of different audio pre-processing strategies on model performance, with particular emphasis on classification accuracy, robustness, and noise reduction effectiveness. Four pre-processing methods were compared: Butterworth filtering (2nd and 4th order), wavelet denoising, and pre-emphasis filtering. Testing accuracy, precision, sensitivity, specificity, and F1-score are reported to evaluate the effectiveness of Butterworth filtering (2nd and 4th order), wavelet denoising, and pre-emphasis filtering as summarized in Table 2.

Butterworth filtering consistently outperformed the other approaches. The 4th-order Butterworth filter (4BF) achieved the highest overall performance, with a testing accuracy of 83.96%, precision of 86.05%, sensitivity of 77.08%, specificity of 89.66%, and an F1-score of 81.32%. This represents a clear improvement over the 2nd-order Butterworth filter (2BF), which achieved a slightly lower accuracy (83.02%) and F1-score (80.00%).

Pre-processing	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
4th Order Butterworth Filter	83.96	86.05	77.08	89.66	81.32
2nd Order Butterworth Filter	83.02	85.71	75.00	89.66	80.00
Wavelet Denoising Filter	77.36	81.58	64.58	87.93	72.09
Pre-emphasis Filter	80.19	84.62	68.75	89.66	75.86

*Note: Mdl.(ResNet50V2), BS(32), Ep.(50), Loss(Categorical), Opt.(Adam)

Table 2. Impact of audio pre-processing strategies on cough-based COVID-19 detection model testing performance.

In terms of specificity, both Butterworth filters (2^{nd} and 4^{th} order) and the pre-emphasis filter, achieved the highest score of 89.66%, while wavelet denoising produced a slightly lower specificity at 87.93%. Wavelet denoising and pre-emphasis filtering demonstrated inferior performance, particularly in sensitivity and F1-score, indicating reduced effectiveness in preserving discriminative cough characteristics. These results suggest that moderate, frequency-selective filtering is more suitable for cough-based COVID-19 detection than aggressive denoising or spectral emphasis. Overall, the 4th-order Butterworth filter was identified as the most effective pre-processing technique and was therefore selected for further analysis. To further understand the observed superiority of Butterworth filtering, particularly the 4th-order configuration, it is necessary to examine how frequency selection influences model performance. Therefore, the effect of cutoff frequency is analyzed in the following section.

Effect of Cutoff Frequency in 4th-Order Butterworth Filtering

To further investigate the influence of frequency selection, multiple cutoff configurations of the 4th-order Butterworth filter were evaluated, including low-pass, band-pass, and high-pass designs. The results performance metrics are reported for low-pass, band-pass, and high-pass configurations to assess the influence of frequency range on cough-based COVID-19 detection are presented in Table 3.

Among all configurations, the low-pass filter with a 10 kHz cutoff frequency yielded the best overall performance, achieving an accuracy of 83.96%, precision of 86.05%, sensitivity of 77.08%, specificity of 89.66%, and an F1-score of 81.32%. These metrics indicate an effective trade-off between false-positive and false-negative rates, which is critical for screening applications.

In contrast, reducing the cutoff frequency to 4 kHz resulted in a substantial decline in accuracy (73.58%) and sensitivity (64.58%), indicating the loss of diagnostically relevant high-frequency information. Band-pass configurations exhibited moderate performance. For example, the 200–4 kHz band-pass achieved the highest sensitivity (95.83%), but it suffered from poor precision (66.67%) and specificity (60.34%), suggesting an increased false-positive rate. High-pass filters at 45 Hz and 150 Hz demonstrated competitive accuracy but lower sensitivity and F1-scores compared to the low-pass 10 kHz configuration.

These findings indicate that preserving a broader frequency spectrum, particularly higher-frequency components, is critical for capturing discriminative acoustic cues in cough sounds. Consequently, the 4th-order low-pass

Butterworth filter with a 10 kHz cutoff was selected as the optimal pre-processing configuration for subsequent experiments.

From a signal-processing perspective, cough sounds are broadband signals containing important high-frequency components associated with turbulent airflow and airway constriction. Limiting the frequency range to lower cutoffs removes these critical features, thereby degrading classification performance. The 10 kHz cutoff effectively preserves these components while reducing high-frequency noise.

Overall, the 4th-order low-pass Butterworth filter with a 10 kHz cutoff provided the most robust and balanced performance among filtering approaches and was therefore selected for further comparison. While the 10 kHz low-pass configuration demonstrated optimal performance among filtering approaches, the results also suggest that preserving a broader frequency spectrum is critical. This observation motivates a direct comparison between filtered and non-filtered (minimal preprocessing) approaches, as presented in the next section.

Cutoff Frequency	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
>10kHz	83.96	86.05	77.08	89.66	81.32
>4kHz	73.58	73.81	64.58	81.03	68.89
-45-4kHz	78.30	77.78	72.92	82.76	75.27
-150-4kHz	77.36	80.00	66.67	86.21	72.73
-200-4kHz	76.42	66.67	95.83	60.34	78.63
<45Hz	83.02	91.67	68.75	94.83	78.57
<150Hz	81.13	79.17	79.17	82.76	79.17
Low-pass “>”		Band-pass “-”		High-pass “<”	

*Note: Mdl.(ResNet50V2), BS(32), Ep.(50), Loss(Categorical), Opt.(Adam)

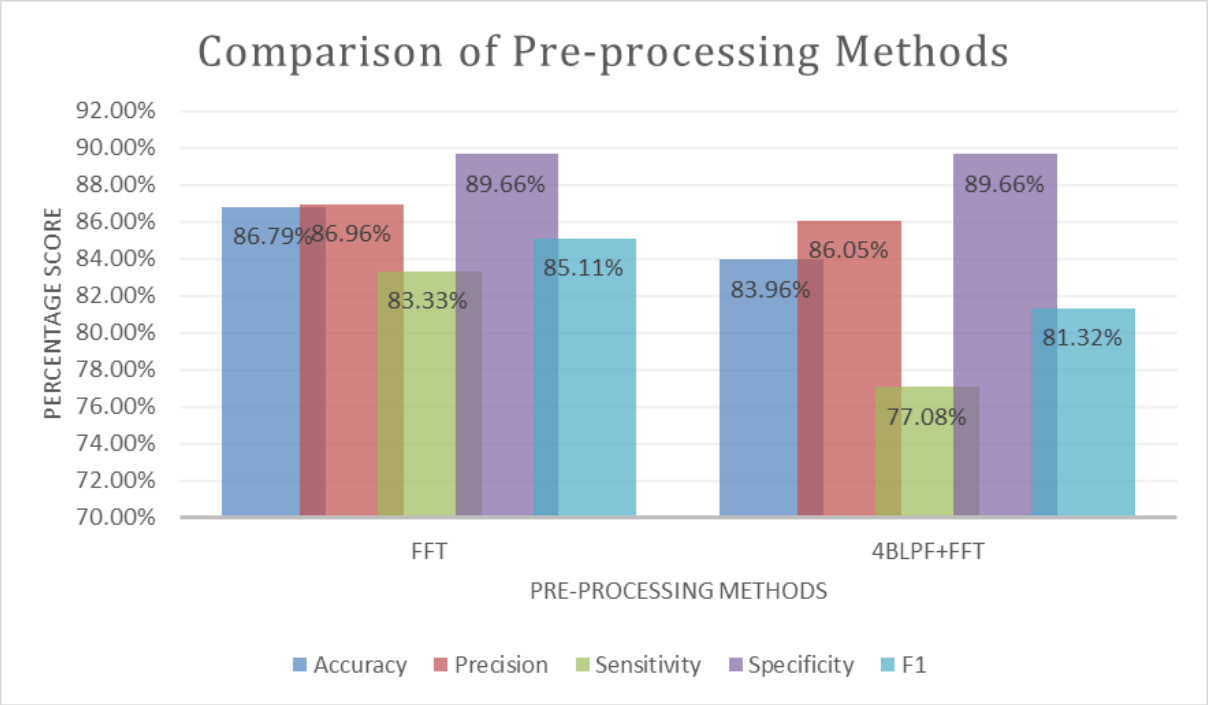
Table 3. Influence of cutoff frequency configuration in a 4th-order Butterworth filter on model testing performance.

3.3 Comparison of Filtering versus No Filtering in Pre-processing

This analysis compares two pre-processing strategies: (i) frequency-domain transformation using Fast Fourier Transform (FFT) alone and (ii) the combination of a 4th-order low-pass Butterworth filter with FFT (4BLPF+FFT). The objective was to determine whether additional filtering provides measurable performance benefits over minimal pre-processing.

As illustrated in Figure 4, the FFT-only approach consistently outperformed the filtered configuration across most evaluation metrics. FFT alone achieved the highest testing accuracy (86.79%), exceeding the 4BLPF+FFT approach by 0.91%. Similar trends were observed for precision, where FFT attained 86.96%, compared to 86.05% for 4BLPF+FFT. Sensitivity improved markedly with FFT-only pre-processing (83.33% versus 77.08%), indicating enhanced detection of COVID-19 positive cases. Both approaches achieved identical specificity (89.66%), suggesting comparable control over false-positive predictions. The F1-score further favoured FFT-only pre-processing (85.11% versus 81.32%).

These results demonstrate that minimal preprocessing can outperform conventional filtering strategies. While Butterworth filtering effectively reduces noise, it may also attenuate high-frequency components that carry diagnostically relevant information. In contrast, FFT-only preprocessing preserves the full spectral content of cough signals, enabling more effective feature learning by the CNN. This finding challenges the conventional assumption that noise reduction inherently improves classification performance and instead highlights the importance of signal preservation for robust deep learning-based diagnosis.



*Note: Mdl.(ResNet50V2), BS(32), Ep.(50), Loss(Categorical), Opt.(Adam)

Figure. 4. Comparison of filtering (4BLPF+FFT) and non-filtering (FFT-only) pre-processing strategies for cough-based COVID-19 classification.

These results indicate that while Butterworth filtering provides effective noise suppression, it can also attenuate diagnostically relevant high-frequency components of cough signals. FFT-only pre-processing preserves the full spectral content of cough sounds without introducing frequency-domain bias, enabling more effective feature learning by the CNN. Consequently, minimal pre-processing using FFT alone offers superior generalisation performance and was selected as the preferred configuration for subsequent experiments.

Impact of Time–Frequency Feature Fusion on COVID-19 Cough Classification

This section compares three feature extraction strategies for COVID-19 cough classification: MFCC-only, STFT-only, and the combined STFT + MFCC representation, to assess whether explicit time–frequency modeling improves classification performance.

As shown in Table 5, STFT + MFCC consistently outperformed both individual feature sets across all evaluation metrics. The combined representation achieved the highest testing accuracy (86.79%), compared with 83.96% for MFCC-only and 83.02% for STFT-only features. Precision increased from 86.05% (MFCC) and 87.04% (STFT) to 86.96%, while sensitivity improved markedly to 83.33%, exceeding MFCC (77.08%) and STFT (81.03%). Specificity remained high (89.66%) and identical to MFCC-only performance, indicating improved COVID-19 detection without increased false-positive rates. The F1-score further confirmed the advantage of feature fusion (85.11%) over MFCC (81.32%) and STFT (83.90%).

Feature Extraction	**Accuracy (%)**	**Precision (%)**	**Sensitivity (%)**	**Specificity (%)**	**F1-score (%)**
STFT	83.02	87.04	81.03	85.42	83.90
MFCC	83.96	86.05	77.08	89.66	81.32
STFT + MFCC	**86.79**	**86.96**	**83.33**	**89.66**	**85.11**

*Note: Mdl.(ResNet50V2), BS(32), Ep.(50), Loss(Categorical), Opt.(Adam)

Table. 4. Comparison of Feature Extraction Methods

Among the single-feature approaches, MFCC-only features exhibited high specificity but lower sensitivity, reflecting their emphasis on global spectral envelopes and perceptual frequency scaling, which may smooth

transient cough dynamics. In contrast, STFT-only features improved sensitivity, capturing localized time–frequency variations associated with cough onset and airflow turbulence, but showed reduced specificity, indicating greater susceptibility to false positives.

The superior performance of STFT + MFCC arises from their complementary characteristics. STFT preserves fine-grained temporal–frequency structures, while MFCCs provide compact, perceptually relevant spectral representations. Their integration yields a richer and more discriminative feature space, enabling more effective learning by CNNs. Overall, these results demonstrate that feature fusion offers a clear and consistent advantage, and STFT + MFCC was therefore selected as the optimal feature extraction strategy for subsequent experiments.

Comparison of Classification Model Architecture and Optimizations

The initial phase of the study involved benchmarking six distinct CNN architectures under standardized conditions (Image size: 224×224, Batch Size: 32, Epochs: 20). Based on Table 5, ResNet50V2 emerged as the optimal architecture, achieving a testing accuracy of 84% and a training loss of 0.0118. While ResNet152V2 achieved a comparable 83% accuracy, it required a 25% longer training duration (12 min 4 s vs. 9 min 39 s for ResNet50V2), making ResNet50V2 the more efficient candidate for edge-deployment in diagnostic tools. In contrast, architectures like MobileNetV2 (68% accuracy) and ResNet50 (59% accuracy) exhibited significant performance degradation, likely failing to capture the complex spectral features of cough acoustics within this specific dataset.

Models	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Testing Accuracy
InceptionV3	0.0300	100.00%	0.9545	62.50%	80.00%
MobileNetV2	0.0190	100.00%	1.1258	59.38%	68.00%
ResNet50	0.6095	65.20%	1.0435	44.79%	59.00%
ResNet50V2	0.0118	100.00%	1.3055	62.50%	84.00%

Table 5. Comparison of Model Architectures

Influence of Hyperparameter Tuning on Convergence

The diagnostic performance of the ResNet50V2 backbone was further enhanced through systematic hyperparameter optimization:

- **Batch Size Selection:** Utilizing a batch size of 32 minimized validation loss to 1.3055 and maximized testing accuracy to 84%, outperforming both smaller (BS=16, 73%) and larger (BS=64, 82%) configurations.
- **Distinct Loss:** Categorical cross-entropy achieved the best performance, with the highest testing accuracy (87%) and validation accuracy, indicating more stable convergence and better generalisation compared to other loss functions. In contrast, sparse categorical cross-entropy and Kullback–Leibler divergence showed lower validation performance, suggesting less effective optimisation for this classification task.
- **Epoch Optimization:** Increasing the training duration to 50 epochs improved the testing accuracy to 87%. However, further training to 100 epochs resulted in overfitting, evidenced by a drop in testing accuracy to 79% despite a reduction in training loss to 0.0004.
- **Optimization Algorithms:** The Adam optimizer proved superior to RMSprop and SGD, achieving the highest testing accuracy (87%) and the lowest validation loss (1.7098) due to its adaptive learning rate and momentum capabilities.

Fixed parameters	Tuning parameters		Training		Validation		Testing
			Loss	Acc.	Loss	Acc.	Acc.
Loss: Categorical, Ep: 20, Opt.: Adam	**Batch Size**	**16**	**0.0009**	1.0000	2.2560	0.6042	73%
		32	0.0118	**1.0000**	**1.3055**	**0.6146**	**84%**
		64	0.0077	1.0000	1.6225	0.5938	82%
BS: 32 Ep.: 50 Opt.: Adam	**Distinct Loss**	**Categorical**	0.0011	1.0000	1.7098	0.6354	87%
		Sparse	0.0012	1.0000	1.6831	0.6146	84%
		Kid	0.0009	1.0000	1.8990	0.5938	84%

BS: 32, Loss: Categorical, Opt.: Adam	**Epoch**	**20**	0.0118	1.0000	**1.3055**	0.6146	84%
		50	**0.0011**	**1.0000**	1.7098	**0.6146**	**87%**
		100	0.0004	1.0000	1.8016	0.5938	79%
BS: 32, Loss: Categorical Ep.: 50,	**Optimizer**	**Adam**	0.0011	**1.0000**	**1.7098**	**0.6146**	**87%**
		RMSprop	**0.0000**	1.0000	4.1493	0.5625	84%
		SGD	29.4750	0.7794	46.4610	0.6146	65%

Table. 6. Influence of Hyperparameter Tuning on the ResNet50V2 model

3.3 Regularization and Feature Integrity

To mitigate the 100% training accuracy observed across several trials—a potential indicator of overfitting—an L1 and L2 regularization parameter of 1×10^{-11} was applied to the Softmax activation. This parameter successfully regularized the weight distribution, maintaining an 87% accuracy while achieving a lower validation loss (1.5744) than the non-regularized baseline (1.7098). Furthermore, the omission of additional pooling and dense layers was found to be beneficial. By maintaining a fully convolutional pipeline, the model preserved high-level spatial information in the MFCC features, which are critical for distinguishing the "dry" impulsive signals of COVID-19 from the resonant "wet" sounds of non-COVID cases.

Figure 5 shows that three regularization parameters, 0.1, 0.01, and 0.001, failed to achieve 80% testing accuracy, achieving only 74%, 64%, and 79%, respectively. Only 0.001 achieved 100% training accuracy, while 0.1 and 0.01 achieved 67.40% and 89.46%, respectively. The training accuracy for other regularization factors was 100%. Most tested parameters were within the 80%-85% range, with "none", 1×10^{-8}, and 1×10^{-11} achieving the highest testing accuracy at 87% and 86%, respectively. The greatest score for validation accuracy was achieved by 1×10^{-11}, which achieved 63.54%, outperforming "none" with a validation accuracy of 61.46%. Additionally, 1×10^{-11} had a lower validation loss of 1.5744 compared to "none" with a higher value of 1.7098.

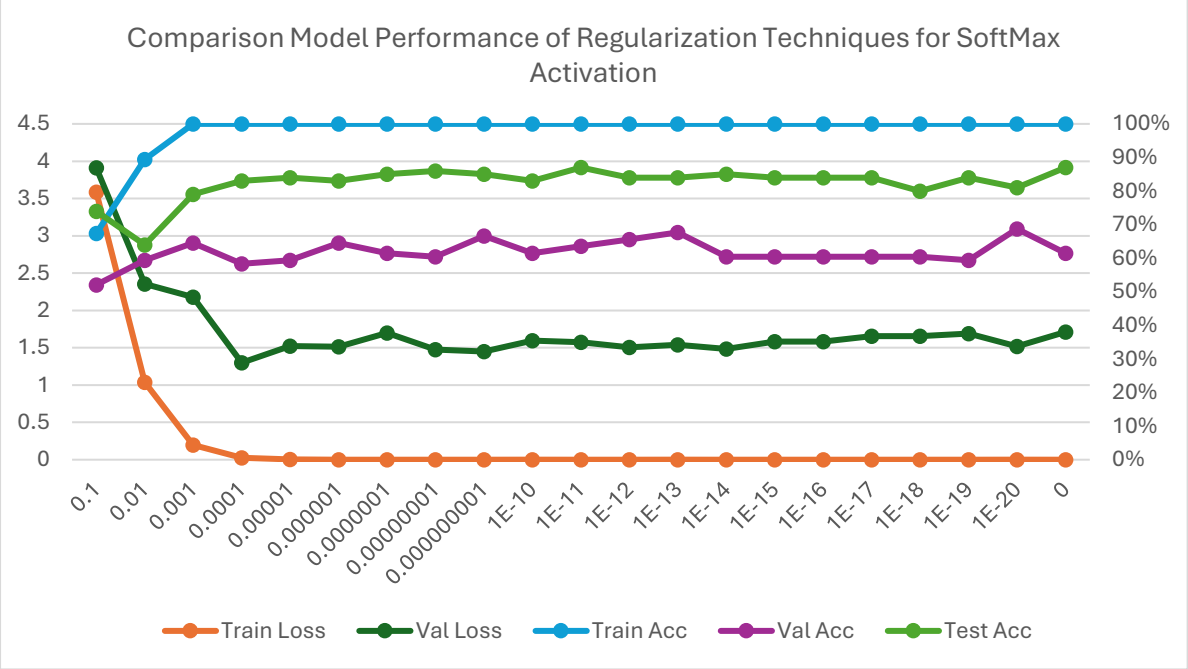
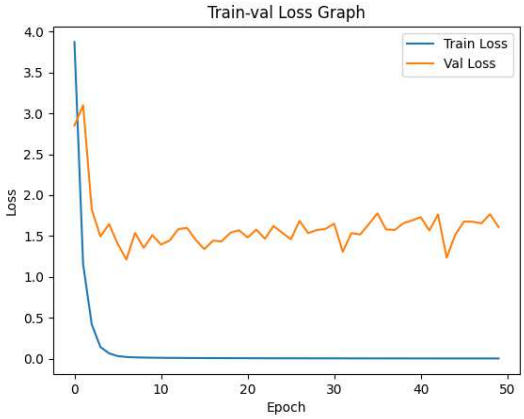
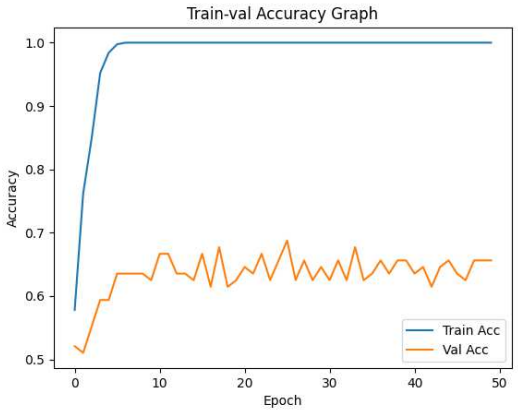


Figure. 5 Comparison of Regularization Techniques for SoftMax Activation (*Note: Mdl. (ResNet50V2), BS. (32), Ep. (50), Loss (Categorical), Opt. (Adam))

Comparing Figure 6 and 7, it proves that regularization parameter 1×10^{-11} is the best option, minimizing overfitting and achieving the highest testing accuracy of 87%. This parameter performs better and has stronger generalization capabilities.

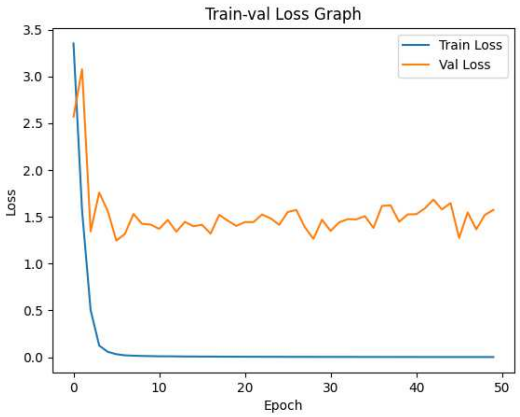


(a) Validation Loss Graph for "None"

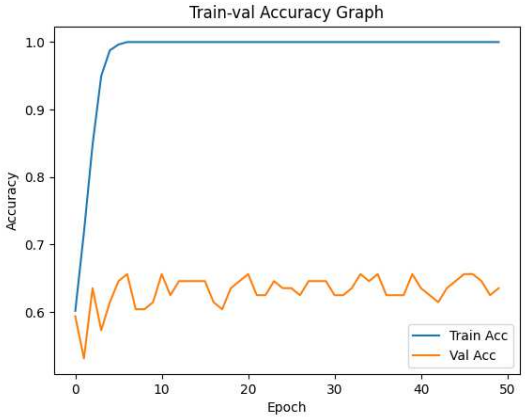


(b) Validation Accuracy for "None"

Figure. 6. (a) Validation Loss Graph for "None" (b) Validation Accuracy for "None" (*Note: Mdl. (ResNet50V2), BS. (32), Ep. (50), Loss (Categorical), Opt. (Adam), L1L2 regularization (None))



(a) Validation Loss Graph for "1×10^{-11}"



(b) Validation Accuracy for "1×10^{-11}"

Figure. 7. (a) Validation Loss Graph for "1×10^{-11}" (b) Validation Accuracy for "1×10^{-11}" (*Note: Mdl. (ResNet50V2), BS. (32), Ep. (50), Loss (Categorical), Opt. (Adam), L1 L2 regularization (1×10^{-11}))

The final optimized model with parameter 1×10^{-11} yielded a testing accuracy of 86.79%. Detailed performance evaluation revealed a precision of 86.96%, sensitivity of 83.33%, and specificity of 89.66%, resulting in a balanced F1-score of 85.11%. These results indicate a robust diagnostic capability that compares favorably with existing RT-PCR benchmarks in speed and accessibility. Overall, the findings confirm that a minimal-preprocessing approach, combined with feature fusion and optimized CNN architecture, provides an effective framework for non-invasive, cough-based COVID-19 screening.

Discussion

The results of this study provide consistent evidence that challenges the conventional assumption that complex preprocessing pipelines are necessary for robust audio-based diagnosis. Across multiple experimental settings, minimal preprocessing using Fast Fourier Transform (FFT) alone outperformed traditional filtering approaches, demonstrating that preserving the full spectral characteristics of cough signals leads to improved classification performance. These findings highlight the importance of signal preservation in deep learning-based diagnostic systems.

From a signal processing perspective, cough sounds are inherently broadband and contain diagnostically relevant acoustic energy extending beyond 4 kHz, particularly during the transient and turbulent phases of cough events. As demonstrated in Table 3, limiting the frequency range through lower cutoff frequencies resulted in a substantial

decline in performance, primarily due to the loss of high-frequency components associated with airflow irregularities and airway constriction. Although the 4th-order low-pass Butterworth filter with a 10 kHz cutoff preserved a wider frequency range and achieved the best performance among filtering approaches, it still did not surpass the FFT-only configuration.

In contrast, alternative filtering strategies either reduced sensitivity due to excessive attenuation of diagnostically relevant signal components or increased false-positive rates by amplifying noise and non-cough artefacts. While the Butterworth filter benefits from a maximally flat passband that minimises amplitude distortion, even well-designed filters introduce some degree of spectral bias. The results therefore suggest that any form of signal manipulation, even when carefully optimised, may inadvertently remove subtle acoustic biomarkers essential for classification.

These observations lead to a key insight of this study: minimal preprocessing, rather than aggressive noise reduction, provides superior generalisation performance. As illustrated in Fig. 4, the FFT-only approach consistently outperformed filtered configurations across all major evaluation metrics. This indicates that preserving the raw spectral content of cough signals enables convolutional neural networks to learn more discriminative and robust features. Consequently, this work supports a paradigm shift from noise suppression toward information preservation in biomedical audio analysis.

From a representation learning perspective, the results in Table 5 further demonstrate that combining STFT and MFCC features yields superior performance compared to individual feature sets. STFT captures fine-grained temporal–spectral variations associated with cough dynamics, while MFCC provides a compact representation aligned with human auditory perception. Their integration produces a richer and more informative feature space, allowing CNN models to better capture subtle differences between COVID-19 and non-COVID-19 cough patterns.

In terms of model architecture, ResNet50V2 achieved the best balance between accuracy, computational efficiency, and generalisation. Although deeper architectures such as ResNet152V2 offer increased representational capacity, the marginal performance gains do not justify the additional computational cost. This balance is particularly important for real-world deployment scenarios, where lightweight and efficient models are required for real-time screening applications.

Overall, the findings of this study establish a clear relationship between signal preservation, feature richness, and model generalisation. By demonstrating that minimal preprocessing can outperform conventional filtering techniques, this work challenges established practices in audio-based disease detection and provides practical insights for the development of scalable, non-invasive diagnostic systems. In particular, the proposed framework is well-suited for deployment in real-world environments, where variability in recording devices, background noise, and user behaviour is unavoidable.

Conclusion

Despite these promising findings, several limitations should be acknowledged. First, the study relies on publicly available datasets, which may contain label noise, class imbalance, and variability in recording conditions. Second, the absence of cross-dataset validation limits the ability to fully assess generalisation across different populations and acquisition settings. Third, the proposed framework has not yet been evaluated in clinical environments and therefore should be considered as a pre-screening tool rather than a replacement for gold-standard diagnostic methods such as RT-PCR.

Future work should focus on cross-dataset evaluation, integration of multi-modal respiratory signals (e.g., cough, breath, and speech), and incorporation of explainable AI techniques to enhance interpretability and clinical trust. Additionally, lightweight model optimisation and on-device deployment should be explored to enable real-time, large-scale screening in resource-constrained settings.

In conclusion, this study demonstrates that minimal preprocessing can improve generalisation and robustness in cough-based COVID-19 detection, challenging conventional assumptions in biomedical signal processing. By emphasising signal preservation over aggressive noise reduction, the proposed framework provides a practical and scalable pathway toward non-invasive, AI-driven respiratory screening systems.

Data availability

All experimental data can be found at https://github.com/iiscleap/Coswara-Data (Project Coswara by the Indian Institute of Science (IISc) Bangalore).

Code availability

The Code available at GitHub https://github.com/cclim-research/Cough-audiobased-COVID

Funding

This research received no external funding.

Reference

1. *Scientific Papers of Silesian University of Technology Organization and Management Series* **2023**, (2023).
2. Banu, K. *et al.* Development and Validation of a Streamlined Multiplex RT-PCR Assay for Accurate SARS-CoV-2 Detection in Diverse Healthcare Environments. *Current Microbiology* **83**, (2026).
3. William, A. *et al.* Comparative evaluation of GeneXpert and reverse transcription polymerase chain reaction for SARS-CoV-2 diagnosis: a retrospective study. *Access Microbiology* **7**, (2025).
4. Akingbola, A., Adewole, O., Toluwalashe, S., Jombo, P. & Ojo, O. Using polymerase chain reaction cycle threshold (PCR Ct) values for predicting and managing cholera outbreaks: potential, challenges, and future directions. *Bulletin of the National Research Centre/Bulletin of the National Research Center* **48**, (2024).
5. Islam, R., Abdel-Raheem, E. & Tarique, M. A study of using cough sounds and deep neural networks for the early detection of Covid-19. *Biomedical Engineering Advances* **3**, 100025 (2022).
6. Sharma, N. K. *et al.* Coswara — A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis. in *Interspeech 2022* (2020). doi:10.21437/interspeech.2020-2768.
7. Qian, K., Schuller, B. W. & Yamamoto, Y. Recent Advances in Computer Audition for Diagnosing COVID-19: An Overview. *arXiv (Cornell University)* (2020) doi:10.48550/arxiv.2012.04650.
8. Mouawad, P., Dubnov, T. & Dubnov, S. Robust Detection of COVID-19 in Cough Sounds. *SN Computer Science* **2**, 34 (2021).
9. Rao, S., Narayanaswamy, V., Esposito, M., Thiagarajan, J. J. & Spanias, A. Deep Learning with hyper-parameter tuning for COVID-19 Cough Detection. 1 (2021) doi:10.1109/iisa52424.2021.9555564.
10. Manshouri, N. Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. *Cognitive Neurodynamics* **16**, 239 (2021).
11. Melek, M. Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound. *Neural Computing and Applications* **33**, 17621 (2021).
12. Xue, H. & Salim, F. D. Exploring Self-Supervised Representation Ensembles for COVID-19 Cough Classification. 1944 (2021) doi:10.1145/3447548.3467263.
13. Ashby, A. E., Meister, J. A., Soldar, G. & Nguyen, K. A. A novel cough audio segmentation framework for COVID-19 detection. *Electronic workshops in computing* (2022) doi:10.14236/ewic/odak22.1.
14. Campana, M. G., Rovati, A., Delmastro, F. & Pagani, E. L3-Net Deep Audio Embeddings to Improve COVID-19 Detection from Smartphone Data. 100 (2022) doi:10.1109/smartcomp55677.2022.00029.
15. Husssain, S. *et al.* Cough2COVID-19 detection using an enhanced multi layer ensemble deep learning framework and CoughFeatureRanker. *Scientific Reports* **14**, 25207 (2024).

16. Lin, Y. *et al.* SympCoughNet: symptom assisted audio-based COVID-19 detection. *Frontiers in Digital Health* **7**, 1551298 (2025).

17. Türkçetin, A. Ö., Koç, T. & Çilekar, Ş. COUGH SOUND ANALYSIS WITH DEEP LEARNING: THE IMPACT OF DATA AUGMENTATION ON RESPIRATORY DISEASE CLASSIFICATION. *DergiPark (Istanbul University)* (2025).

18. Altae, A. A., Rad, A. E. & Mohebbi, K. Advanced COVID-19 detection using cough signals with space reconstruction and 3D deep convolutional neural networks. *Scientific Reports* **15**, 45275 (2025).

19. Kumar, S. *et al.* Ensemble multimodal deep learning for early diagnosis and accurate classification of COVID-19. *Computers & Electrical Engineering* **103**, 108396 (2022).

20. Chowdhury, M. E. H. *et al.* QUCoughScope: An Artificially Intelligent Mobile Application to Detect Asymptomatic COVID-19 Patients using Cough and Breathing Sounds. *arXiv (Cornell University)* (2021) doi:10.48550/arxiv.2103.12063.

21. Pal, A. & Sankarasubbu, M. Pay Attention to the cough: Early Diagnosis of COVID-19 using Interpretable Symptoms Embeddings with Cough Sound Signal Processing. *arXiv (Cornell University)* (2020) doi:10.48550/arxiv.2010.02417.

22. Xu, W. *et al.* A forced cough sound based pulmonary function assessment method by using machine learning. *Frontiers in Public Health* **10**, 1015876 (2022).

23. Andreu-Pérez, J. *et al.* A Generic Deep Learning Based Cough Analysis System from Clinically Validated Samples for Point-of-Need Covid-19 Test and Severity Levels. (2021) doi:10.31219/osf.io/tm2f7.

24. Pahar, M., Klopper, M., Warren, R. M. & Niesler, T. COVID-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine* **135**, 104572 (2021).

25. Wang, Y. *et al.* Automated Cough Analysis with Convolutional Recurrent Neural Network. *Bioengineering* **11**, 1105 (2024).

26. Ghourabi, M., Mourad-Chehade, F., Soubra, R. & Chkeir, A. Enhanced Cough Analysis Using 1-Dimensional CNN Features for Respiratory Health Diagnosis. 1 (2024) doi:10.1109/memea60663.2024.10596786.

27. Tan, K. P., Smith, J. A. & Gaydecki, P. Cough sound spectro-temporal analysis and automated detection using Vision Transformers. *Digital Health* **11**, (2025).

28. Aytekin, I. *et al.* COVID-19 Detection From Respiratory Sounds With Hierarchical Spectrogram Transformers. *IEEE Journal of Biomedical and Health Informatics* **28**, 1273 (2023).

29. Islam, R., Chowdhury, N. K. & Kabir, M. A. Robust COVID-19 Detection from Cough Sounds using Deep Neural Decision Tree and Forest: A Comprehensive Cross-Datasets Evaluation. *arXiv (Cornell University)* (2025) doi:10.48550/arxiv.2501.01117.

30. Bakırcı, M. Internet of Things-enabled unmanned aerial vehicles for real-time traffic mobility analysis in smart cities. *Computers & Electrical Engineering* **123**, 110313 (2025).

31. Orlandic, L., Teijeiro, T. & Atienza, D. The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms. *Infoscience (Ecole Polytechnique Fédérale de Lausanne)* https://doi.org/10.5281/zenodo.4048311 (2021).

32. Bhattacharya, D. *et al.* Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection. *Scientific Data* **10**, 397 (2023).

33. Muguli, A. *et al.* DiCOVA Challenge: Dataset, Task, and Baseline System for COVID-19 Diagnosis Using Acoustics. (2021) doi:10.21437/interspeech.2021-74.

34. Ganitidis, T., Αθανασίου, M., Mitsis, K., Zarkogianni, K. & Nikita, K. S. A Comprehensive Drift-Adaptive Framework for Sustaining Model Performance in COVID-19 Detection From Dynamic Cough Audio Data: Model Development and Validation. *Journal of Medical Internet Research* **27**, (2025).

35. Islam, R., Chowdhury, N. K. & Kabir, M. A. Robust COVID-19 detection from cough sounds using deep neural decision tree and forest: A comprehensive cross-datasets evaluation. *Expert Systems with Applications* **310**, 131235 (2026).

36. Orlandic, L., Teijeiro, T. & Atienza, D. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Scientific Data* **8**, 156 (2021).

37. Harvill, J. *et al.* Classification of COVID-19 from Cough Using Autoregressive Predictive Coding Pretraining and Spectral Data Augmentation. 926 (2021) doi:10.21437/interspeech.2021-799.

38. Aquino-García, A. M., Perez-Espinosa, H., Andreu-Pérez, J. & Rodríguez-González, A. Y. Audio-Based Screening of Respiratory Diseases Using Machine Learning: A Methodological Framework Evaluated on a Clinically Validated COVID-19 Cough Dataset. *Machine Learning and Knowledge Extraction* **8**, 80 (2026).

39. Kho, S. J. *et al.* Malaysian cough sound analysis and COVID-19 classification with deep learning. *Intelligence-Based Medicine* **9**, 100129 (2023).

40. Atmaja, B. T., Zanjabila, Suyanto, S. & Sasou, A. Comparing hysteresis comparator and RMS threshold methods for automatic single cough segmentations. *International Journal of Information Technology* **16**, 5 (2023).

41. Suyanto, S. Performance Improvement of Covid-19 Cough Detection Based on Deep Learning with Segmentation Methods. *Journal of Applied Data Sciences* **5**, 520 (2024).

42. Cai, Y. & Xu, W. Recognition and Extraction of Cough Sound from Audio Signals. in *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)* 696 (2021). doi:10.1109/icaice54393.2021.00136.

43. Monge-Álvarez, J., Hoyos-Barceló, C., San-José-Revuelta, L. M. & Casaseca-de-la-Higuera, P. A Machine Hearing System for Robust Cough Detection Based on a High-Level Representation of Band-Specific Audio Features. *IEEE Transactions on Biomedical Engineering* **66**, 2319 (2018).

44. Zhu, Y., Shaik, M. H. & Falk, T. H. On the Importance of Different Cough Phases for COVID-19 Detection. **5**, 1 (2023).

45. Sharan, R. V. Cough sound detection from raw waveform using SincNet and bidirectional GRU. *Biomedical Signal Processing and Control* **82**, 104580 (2023).

46. Zhu, Y., Shaik, M. & Falk, T. H. On the importance of different cough phases for COVID-19 detection. (2022) doi:10.36227/techrxiv.21382176.

47. Dağılma, Y. & Özbay, E. Processing and Classification of Covid-19 Audio Data with Optimization-Based Learning Techniques. 1 (2024) doi:10.1109/idap64064.2024.10710887.

48. Vinod, A., Mohan, N., Kumar, S. S. & Soman, K. P. Covid Cough Identification using Machine Learning and Deep Learning Networks. 1 (2023) doi:10.1109/conit59222.2023.10205895.

49. Puja, W. D., Sultana, S. & Hossain, A. B. M. A. COVID-19 Detection from Cough Sound Signal using Random Forest Learning of Deep Spectral Features. 1 (2024) doi:10.1109/spicscon64195.2024.10940971.

50. Shih, Y.-T. *et al.* Design of cough sound classifiers based on attention mechanism and contrastive loss for detection of coronavirus disease 2019. *Engineering Applications of Artificial Intelligence* **163**, 112949 (2025).

51. Lee, G.-T. *et al.* Deep learning based cough detection camera using enhanced features. *Expert Systems with Applications* **206**, 117811 (2022).

52. Zhang, X. *et al.* Robust Cough Feature Extraction and Classification Method for COVID-19 Cough Detection Based on Vocalization Characteristics. in *Interspeech 2022* 2168 (2022). doi:10.21437/interspeech.2022-10401.

53. Gupta, R., Krishna, T. A. & Adeeb, M. Cough-based COVID-19 Detection with Multi-head Deep Neural Network. 1 (2022) doi:10.1109/icdsaai55433.2022.10028889.

54. Çelik, G. CovidCoughNet: A new method based on convolutional neural networks and deep feature extraction using pitch-shifting data augmentation for covid-19 detection from cough, breath, and voice signals. *Computers in Biology and Medicine* **163**, 107153 (2023).

55. Pahar, M., Klopper, M., Warren, R. M. & Niesler, T. COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features. *Computers in Biology and Medicine* **141**, 105153 (2021).

56. Lella, K. K. & Alphonse, P. J. A. COVID-19 disease diagnosis with light-weight CNN using modified MFCC and enhanced GFCC from human respiratory sounds. *The European Physical Journal Special Topics* **231**, 3329 (2022).

57. Ávila, F. R. *et al.* Investigating Feature Selection and Explainability for COVID-19 Diagnostics from Cough Sounds. 951 (2021) doi:10.21437/interspeech.2021-2197.

58. Ghourabi, M., Mourad-Chehade, F. & Chkeir, A. Machine Learning in COVID-19 Diagnosis: Innovation Approaches for Cough Sound Classification. 1 (2024) doi:10.1109/iccitx61791.2024.11070548.

59. Södergren, I., Nodeh, M. P., Chhipa, P. C., Nikolaidou, K. & Kovács, G. Detecting COVID-19 from Audio Recording of Coughs Using Random Forests and Support Vector Machines. 916 (2021) doi:10.21437/interspeech.2021-2191.

60. Shati, A., Hassan, G. M. & Datta, A. COVID-19 Detection System: A Comparative Analysis of System Performance Based on Acoustic Features of Cough Audio Signals. *arXiv (Cornell University)* (2023) doi:10.48550/arxiv.2309.04505.

61. Chehili, I. & Hamdi, S. Precision Diagnosis in Pandemics: Machine Learning and Cough Sound Analysis for COVID-19. 1 (2024) doi:10.1109/ict-dm62768.2024.10798940.

62. Wang, H. "SOS Signal" in Breathing Sound - Rapid COVID-19 Diagnosis Based on Machine Learning. **710**, 522 (2022).

63. Kaushik, P., Bhutani, M. & Bhagat, K. M. Cough Signal Processing for Classification and Early Detection of COVID-19 Using Deep Learning Integrated Approach. in *Lecture notes in networks and systems* 202 (Springer International Publishing, 2025). doi:10.1007/978-3-032-02949-2_18.

64. Cai, R. *et al.* Predicting and Visualizing Covid-19 Identification by a Hybrid Machine Learning and Pre-trained Model. 5936 (2024) doi:10.1109/bibm62325.2024.10821829.

65. Hamdi, S., Moussaoui, A., Oussalah, M. & Saidi, M. Autoencoders and Ensemble-Based Solution for COVID-19 Diagnosis from Cough Sound. in *Lecture notes in networks and systems* 279 (Springer International Publishing, 2022). doi:10.1007/978-3-031-18516-8_20.

66. Raj, A. & Bhukya, R. K. Vocal Biomarker Based COVID-19 Detection Using DNN and Transfer Learning ResNet50. in *2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)* vol. 20 1 (2022).

67. Nguyen-Trong, K. & Nguyen-Hoang, K. Multi-modal approach for COVID-19 detection using coughs and self-reported symptoms. *Journal of Intelligent & Fuzzy Systems* **44**, 3501 (2022).

68. Rayan, A. *et al.* Utilizing CNN-LSTM techniques for the enhancement of medical systems. *Alexandria Engineering Journal* **72**, 323 (2023).

69. Islam, R., Abdel-Raheem, E. & Tarique, M. Early Detection of COVID-19 Patients using Chromagram Features of Cough Sound Recordings with Machine Learning Algorithms. 82 (2021) doi:10.1109/icm52667.2021.9664931.

70. Căliman-Sturdza, O. A., Filip, F., Baitan, M. T. & Dimian, M. Deep Learning Network Selection and Optimized Information Fusion for Enhanced COVID-19 Detection: A Literature Review. *Diagnostics* **15**, 1830 (2025).

71. Esposito, M., Rao, S., Narayanaswamy, V. & Spanias, A. COVID-19 Detection using Audio Spectral Features and Machine Learning. in *2021 55th Asilomar Conference on Signals, Systems, and Computers* 1146 (2021). doi:10.1109/ieeeconf53345.2021.9723323.

72. Truong, T., Lenga, M., Serrurier, A. & Mohammadi, S. Fused Audio Instance and Representation for Respiratory Disease Detection. *Sensors* **24**, 6176 (2024).

73. Bhosale, S., Tiwari, U., Chakraborty, R. & Kopparapu, S. K. Contrastive Learning of Cough Descriptors for Automatic COVID-19 Preliminary Diagnosis. 946 (2021) doi:10.21437/interspeech.2021-1249.

74. Hussain, S., Amran, G. A., Alabrah, A., Alkhalil, L. & AL-Bakhrani, A. A. C19-MLE: A Multi-Layer Ensemble Deep Learning Approach for COVID-19 Detection Using Cough Sounds and X-Ray Imaging. *IEEE Access* **12**, 197151 (2024).

75. 7Effati, M. & Nejat, G. Deep Learning-Based Multi-modal COVID-19 Screening by Socially Assistive Robots Using Cough and Breathing Symptoms. in *Lecture notes in computer science* 217 (Springer Science+Business Media, 2022). doi:10.1007/978-3-031-24670-8_20.

76. Das, R. K., Madhavi, M. C. & Li, H. Diagnosis of COVID-19 Using Auditory Acoustic Cues. 921 (2021) doi:10.21437/interspeech.2021-497.

77. Zimmer, A. *et al.* Cough acoustic analysis using artificial intelligence for COVID-19 detection: A comparative study of patient cohorts from Lima, Peru and Montreal, Canada. *Annals of Epidemiology* **118**, 110076 (2026).

78. Shen, J. *et al.* Novel audio characteristic-dependent feature extraction and data augmentation methods for cough-based respiratory disease classification. *Computers in Biology and Medicine* **179**, 108843 (2024).

79. Henna, S., Alcaraz, J. M. L., Rathnayake, U. & Amjath, M. An interpretable deep learning framework for medical diagnosis using spectrogram analysis. *Healthcare Analytics* **8**, 100408 (2025).

80. A, A. M. Javitha. & Z., M. Black Box to Trust Box: Building Confidence in AI-Powered Diagnosis Through Explainable Models for COVID-19. 949 (2025) doi:10.1109/iciss63372.2025.11076338.

81. Wullenweber, A., Akman, A. & Schuller, B. W. CoughLIME: Sonified Explanations for the Predictions of COVID-19 Cough Classifiers. in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* vol. 2022 1342 (2022).

82. Sultana, S., Hossain, A. B. M. A. & Alam, J. COVID-19 detection from optimized features of breathing audio signals using explainable ensemble machine learning. *Results in Control and Optimization* **18**, 100538 (2025).

83. Grant, D., McLane, I. & West, J. E. Rapid and Scalable COVID-19 Screening using Speech, Breath, and Cough Recordings. 1 (2021) doi:10.1109/bhi50953.2021.9508482.

84. Chen, S. *et al.* Deep Learning and Machine Learning -- Python Data Structures and Mathematics Fundamental: From Theory to Practice. *arXiv (Cornell University)* (2024) doi:10.48550/arxiv.2410.19849.

85. Freire, F. D. & Soares, I. G. Conservation of the passband signal amplitude using a filter based on the Fast Fourier Transform algorithm. *arXiv (Cornell University)* (2024) doi:10.48550/arxiv.2407.13414.

86. Morice, A. H. *et al.* Cough frequency has a high daily variation in patients with chronic cough. *ERJ Open Research* **11**, 670 (2024).

87. Pingale, T. H. & Patil, H. T. Analysis of Cough Sound for Pneumonia Detection Using Wavelet Transform and Statistical Parameters. 1 (2017) doi:10.1109/iccubea.2017.8463900.

88. Ferrari, S., Silva, M. M. da, Guarino, M. & Berckmans, D. Analysis of Cough Sounds for Diagnosis of Respiratory Infections in Intensive Pig Farming. *Transactions of the ASABE* **51**, 1051 (2008).

89. Piralideh, N. O., Gournay, P. & Lefebvre, R. Zero-Bit Transmission of Adaptive Pre- and De-emphasis Filters for Speech and Audio Coding. *arXiv (Cornell University)* (2024) doi:10.48550/arxiv.2407.02672.

90. Hadjaidji, E., Korba, M. C. A. & Khelil, K. COVID-19 Detection from Cough Sounds Using XGBoost and LSTM Networks. *Traitement du signal* **42**, 939 (2024).

91. Krishnan, S. Advanced analysis of biomedical signals. in *Elsevier eBooks* 157 (Elsevier BV, 2021). doi:10.1016/b978-0-12-813086-5.00003-7.

Acknowledgements

The authors received no specific funding for this work.

Author contributions

C.C.L., L.B.T, and J.T.L.Y. contributed to design and conceptualization of the study. M.S.A.M.A., C.C.L. and J.T.L.Y. contributed to data analysis and interpretation. All authors contributed to drafting and revision of the manuscript and approved the final manuscript. All authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.C. L.