



OPEN ACCESS

EDITED BY

Arfat Ahmad Khan,
Khon Kaen University, Thailand

REVIEWED BY

Priya E.,
Sri Sairam Engineering College, Chennai, India
Dharmesh Shah,
Indrashil University, India

*CORRESPONDENCE

Leticia Silva,
✉ araujos.leticia@gmail.com

RECEIVED 05 September 2025

REVISED 25 November 2025

ACCEPTED 30 December 2025

PUBLISHED 22 January 2026

CITATION

Silva L, Floriano A, Valadão C, Caldeira E,
Krishnan S and Bastos Filho T (2026) Comparing
compressive sensing and downsampling for
COVID-19 diagnosis from cough and speech
audio signals.

Front. Signal Process. 5:1700044.

doi: 10.3389/frsip.2025.1700044

COPYRIGHT

© 2026 Silva, Floriano, Valadão, Caldeira,
Krishnan and Bastos Filho. This is an open-
access article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Comparing compressive sensing and downsampling for COVID-19 diagnosis from cough and speech audio signals

Leticia Silva^{1*}, Alan Floriano², Carlos Valadão³, Eliete Caldeira¹,
Sridhar Krishnan⁴ and Teodiano Bastos Filho¹

¹Department of Electrical Engineering, Postgraduate Program in Electrical Engineering, Federal University of Espírito Santo, Vitória, Brazil, ²Federal Institute of Paraná (IFPR), Telêmaco Borba, Brazil, ³Federal Institute of Espírito Santo (IFES), Linhares, Brazil, ⁴Signal Analysis Research Group, Department of Electrical, Computer, and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON, Canada

Introduction: Since the onset of the COVID-19 pandemic, extensive research has focused on developing non-invasive diagnostic approaches of respiratory syndrome using biomedical signals, particularly cough and speech audio. Time-frequency representations combined with Machine Learning models have shown potential in identifying acoustic biomarkers associated with respiratory conditions. Although many existing approaches demonstrate high performance, their use may be limited in resource-constrained environments due to processing or implementation demands.

Methods: In this study, we propose an end-to-end approach for COVID-19 inference based on compressed time-domain audio signals. The method combines temporal signal compression strategies - Downsampling (DS) and Compressive Sensing (CS) - with a Convolutional Neural Network (CNN) trained directly on the waveforms. This design eliminates the need for handcrafted features or spectrograms, aiming to reduce computational complexity while preserving classification performance.

Results: To evaluate the proposed structure, we used data from two open-access datasets, one for coughing and one for speech. Experimental results, assessed using accuracy and F1-score metrics, indicate that CS outperformed DS in most scenarios, particularly under high compression rates (e.g., 200 Hz and 100 Hz).

Discussion: These findings support the use of compressed audio-based classification in real-world embedded and mobile health systems, where computational efficiency is essential.

KEYWORDS

COVID-19, compressive sensing, downsampling, CNN, cough, speech, embedded health systems

1 Introduction

COVID-19 (Coronavirus Disease 2019) is an infectious respiratory illness caused by the SARS-CoV-2 virus, first reported in late 2019. Due to its rapid global spread, the World Health Organization (WHO) officially classified it as a pandemic on 11 February 2020 (World Health Organization, 2020). The disease quickly became a major health challenge worldwide, significantly impacting both public health systems and economies. According to data from 2023 from Johns Hopkins University, COVID-19 has led to millions of infections

and deaths globally (Johns Hopkins University, 2022). Brazil alone has reported over 38 million confirmed cases and more than 700,000 deaths (Worldometer, 2024).

Clinically, COVID-19 presents a wide range of symptoms, including fever, fatigue, respiratory difficulties, cough, and voice changes. Dysphonia, characterized by changes in voice quality, has been reported in approximately 25% of patients with mild to moderate forms of COVID-19. This reflects the systemic nature of the virus and its impact on respiratory and vocal tract structures, such as the lungs, larynx, and vocal folds (Rai et al., 2021; Oliveira et al., 2020).

Given the pronounced respiratory and vocal involvement, audio signal analysis has emerged as a promising non-invasive tool for detecting and monitoring COVID-19. Previous studies have demonstrated that cough, breathing, and speech signals contain discriminative acoustic patterns correlated with respiratory conditions (Brown et al., 2020; Sharma et al., 2022; Pleva et al., 2022; Pahar et al., 2022). Such audio-based diagnostic methods are particularly suited to telemedicine and remote patient monitoring due to their simplicity, affordability, and ease of integration with mobile and portable platforms (Villa-Parra et al., 2022).

Traditionally, audio-based health diagnostics have relied on handcrafted acoustic features, such as Mel-Frequency Cepstral Coefficients (MFCC), Zero-Crossing Rate (ZCR), and Spectral Roll-off (Pramono et al., 2016; Verde et al., 2021). These features, typically extracted from spectrograms or raw waveforms, serve as input for classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and ensemble methods (Sharma et al., 2022; Casanova et al., 2021). More recently, deep learning architectures have attracted attention by enabling automatic learning of discriminative representations from raw or minimally processed audio signals (Pahar et al., 2022). Despite improved classification accuracy, these advanced approaches often require substantial computational resources, limiting their practical deployment in embedded and mobile devices. Real-time processing and classification in such resource-constrained environments remain significant challenges due to limitations in processing power, memory, and energy availability (Diab and Rodriguez-Villegas, 2022).

To address these challenges, we have developed the Integrated Portable Medical Assistant (IPMA), a multimodal platform designed to automatically acquire and analyze multiple physiological and acoustic biomarkers (Villa-Parra et al., 2022). The IPMA captures parameters such as cough, speech, forced breathing, oxygen saturation, blood pressure, heart rate, and body temperature, facilitating comprehensive patient screening in remote or resource-limited environments.

Although such platforms can benefit from cloud-based analytics and storage, the use of audio signals and physiological data raises important concerns regarding data privacy and security. Previous studies have emphasized the need for robust encryption and access control mechanisms to protect sensitive health information transmitted over networks (Deepika et al., 2021; Jayaram and Prabakaran, 2021). Ensuring secure and privacy-preserving computation is critical for the real-world applicability of these systems, especially in mobile and telemedicine contexts.

In resource-constrained scenarios, such as embedded or portable systems like the IPMA, it is essential to adopt strategies that balance diagnostic accuracy with low computational complexity. In this

context, signal compression methods serve as enabling technologies that facilitate the deployment of advanced diagnostic algorithms in resource-constrained environments, particularly for COVID-19 detection in remote or underserved areas. Among available approaches, Downsampling (DS) and Compressive Sensing (CS) emerge as compelling solutions that can potentially preserve diagnostic information while significantly reducing computational demands. DS directly reduces signal dimensionality by lowering the sampling frequency, potentially leading to loss of relevant acoustic information (Bent et al., 2021; Casaseca-de-la Higuera et al., 2015). Conversely, CS leverages the inherent sparsity of biomedical signals, enabling compact representations by projecting the data into a lower-dimensional space through random measurements (Casaseca-de-la Higuera et al., 2015; Prabhavathi et al., 2023; Wang et al., 2016). Although both methods have been extensively explored individually in audio signal processing, their comparative performance for COVID-19 audio-based detection remains largely unexplored (Casaseca-de-la Higuera et al., 2015), particularly regarding their ability to maintain diagnostic accuracy under extreme compression conditions.

Recent studies have proposed self-regulated diagnostic frameworks for diagnosing various conditions using Deep Neural Network (DNN) (Jo and Kwak, 2022; Kapoor et al., 2022; Patel et al., 2022). Among them, Convolutional Neural Networks (CNNs) have been applied to analyze respiratory audio signals—such as cough and breathing—to detect diseases like COVID-19 (Kapoor et al., 2022) and to classify complex audio patterns in areas including depression diagnosis (Jo and Kwak, 2022), and lung function prediction (Patel et al., 2022).

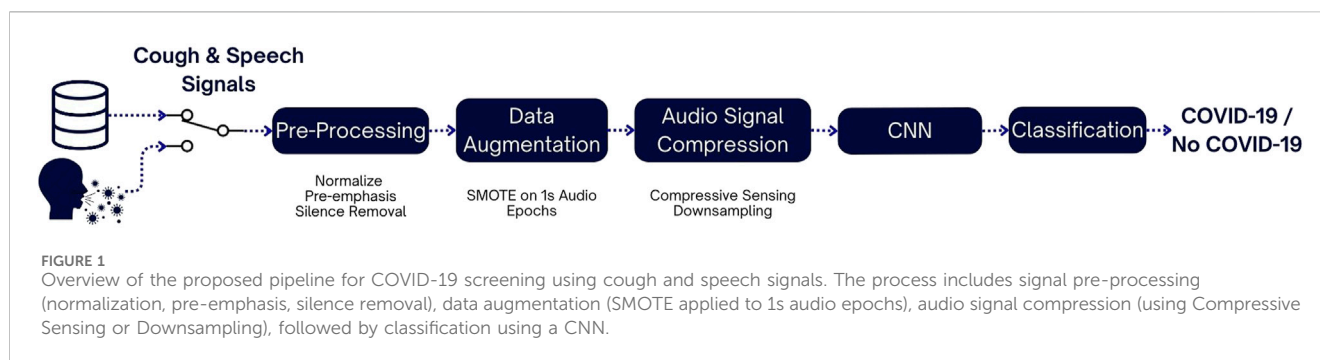
In this work, we propose an end-to-end methodology for COVID-19 inference from respiratory audio signals (cough and speech), combining temporal signal compression techniques—DS and CS—with a CNN architecture. Unlike many existing approaches that operate on spectrograms or require intermediate representations, our CNN processes compressed time-domain waveforms directly, maintaining end-to-end efficiency even under extreme compression conditions. This integration of signal compression with direct waveform processing creates a particularly streamlined pipeline suited for resource-constrained environments. Comparing CS with DS explicitly quantifies the performance advantage of a sparsity-aware approach over simple downsampling, supporting our hypothesis that CS can enable accurate classification of cough and speech signals while maintaining low computational complexity in resource-constrained systems. To validate this, we evaluate the methodology on two publicly available datasets, without performing cross-dataset inference.

The remainder of this paper is structured as follows. Section 2 describes the proposed system to infer COVID-19 from cough and speech signals, detailing the datasets and methods used. Section 3 presents experimental results, followed by their discussion and analyses in Section 4. Finally, concluding remarks are provided in Section 5.

2 Materials and methods

2.1 Overview of the system

Figure 1 presents the proposed pipeline for COVID-19 detection using cough and speech audio signals. The process begins with a pre-



processing stage that includes normalization, pre-emphasis filtering, silence removal using Root Mean Square (RMS) energy, and segmentation of the signal into fixed-length epochs.

Data augmentation is then applied to address class imbalance during training. The signal is subsequently compressed using either CS or DS to reduce data volume while preserving relevant information. The resulting representations are processed by a CNN that extracts temporal patterns for classification. Inference is then performed to classify the input as COVID-19 positive or negative. The following subsections detail each stage of the pipeline.

2.2 Dataset description

This study uses two subsets of the INTERSPEECH 2021 Computational Paralinguistics Challenge (ComParE), an open challenge based on speech signals: the ComParE COVID-19 Cough Sub-Challenge (CCS) and the ComParE COVID-19 Speech Sub-Challenge (CSS) (Schuller et al., 2021). These datasets were provided by Cambridge University under a mutual agreement for research purposes, and their use was approved by the Department of Computer Science and Technology at Cambridge University, following all the requisite sets by the ethics committee.

The CCS dataset contains cough audio recordings from both COVID-19 positive and negative subjects. In the same way, the CSS dataset includes speech recordings from individuals both infected by COVID-19 and those who are not. The audio data for both datasets were collected via the “COVID-19 Sounds App” available on multiple platforms (a webpage, an Android app, and an iOS app). Participants were asked to provide one to three forced coughs and to say, “I hope my data can help to manage the virus pandemic” one to three times. As described by Schuller et al. (2021), all audio files were manually checked, resampled, and converted to 16 kHz mono/16 bit format. In addition to the audio recordings, the dataset also includes demographic and clinical information provided by the participants. These features include variables such as sex, age group, presence of symptoms, and medical history (e.g., asthma, diabetes, valvular heart disease), allowing further analyses of potential associations between these variables and COVID-19 status.

For both the CCS and CSS datasets, the original data partitions were preserved, as proposed by Schuller et al. (2021), ensuring that our results may be compared with those reported in earlier studies using the same dataset. Specifically, the CCS dataset comprises 286 samples for training, 231 for validation, and 208 for testing.

Similarly, the CSS dataset includes 315 samples for training, 295 for validation, and 283 for testing.

2.3 Signal pre-processing

Pre-processing was applied uniformly to all audio samples to prepare them for analysis. Prior to any processing, amplitude normalization was performed on each audio recording to ensure consistent signal levels across all samples. Following, a first-order pre-emphasis filter ($\alpha = 0.97$) was applied to amplify high-frequency components, balancing the audio spectrum and enhancing features relevant for robust classification by CNN (Shi, 2025). Despite using an end-to-end time-domain approach, this pre-processing step improves the signal-to-noise ratio, making important audio characteristics more distinguishable (Shi, 2025). Preliminary experiments showed that removing pre-emphasis notably degraded classification performance, confirming its beneficial role in our methodology. Subsequently, silence removal was performed using a frame-based energy detection method, retaining only frames with normalized RMS energy above a fixed threshold of 0.02 (in the range [0,1], not in decibels). Each recording was segmented into overlapping frames of 25 ms with a 10 ms step. To standardize input length and enable epoch-based processing, zero-padding was applied to ensure the final signal length was a multiple of the sampling rate, allowing for segmentation into 1 s epochs. The 1-s epoch duration balances the time-frequency resolution trade-off: $\Delta f = 1/T$ ensures sufficient spectral resolution to preserve clinically relevant acoustic cues (Boashash, 2015).

2.4 Data augmentation

To address class imbalance in the training set, we employed the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002), which creates synthetic samples for the minority class by interpolating between existing instances and their nearest neighbors. This technique has shown effectiveness in high-dimensional datasets, including biomedical and audio-based classification tasks (Wang et al., 2023; Lee and Lee, 2023).

The original class distribution, detailed in Table 1, revealed a moderate imbalance (approximately 3:1) favoring the negative class. Although not severe, this imbalance could bias the model toward the majority class, negatively impacting the performance for the

TABLE 1 Class distribution: original dataset vs. SMOTE-augmented dataset.

		COVID-19 (+)			COVID-19 (−)		
		Train	Validation	Test	Train	Validation	Test
Original	CCS	71	48	39	215	183	169
	CSS	72	142	94	243	153	189
SMOTE	CCS	215	48	39	215	183	169
	CSS	243	142	94	243	153	189

minority class. To mitigate this, we applied SMOTE exclusively to the training set. Preliminary tests indicated that excluding SMOTE led to noticeable performance degradation, particularly for the minority class, confirming the relevance of this strategy (Joloudari et al., 2023; Lee and Lee, 2023).

Mathematically, the number of synthetic samples generated for each minority class is determined by the oversampling factor N_i , as defined in Equation 1 (Chawla et al., 2002):

$$N_i = \frac{n_{\text{majority}}}{n_i} - 1, \quad (1)$$

where n_i is the number of original samples in the minority class, and n_{majority} is the number of samples in the majority class. The integer part of N_i specifies how many synthetic samples are generated per original observation, whereas the fractional part is randomly distributed among a subset of observations to reach the exact target count. SMOTE generates each synthetic sample by interpolating between original samples and their $k = 5$ nearest neighbors.

Although SMOTE was originally developed for feature-space augmentation, recent studies have adapted it to time-series and waveform-based data (Iwana and Uchida, 2021). In our case, we treated each 1 s raw audio epoch as a high-dimensional vector and applied SMOTE directly in the time domain. While the resulting synthetic waveforms may sound unnatural, the CNN can learn discriminative patterns from these signals. The objective was not to produce perceptually realistic audio, but to improve class balance and model generalization. Experimental results confirmed that waveform-level SMOTE improved classification performance for the minority class.

2.5 Signal compression

Signal compression plays an important role in scenarios where transmission bandwidth, storage capacity, or computational power are constrained, as commonly observed in embedded biomedical systems (Casaseca-de-la Higuera et al., 2015; Santos, 2023). These systems are frequently deployed in real-world health applications such as portable respiratory monitors, wearable biosensors, and mobile diagnostic tools, where power supply is limited and data must be transmitted wirelessly in near real-time. Devices often rely on low-energy wireless protocols like Bluetooth Low Energy (BLE) (Santos, 2023), which impose strict constraints on transmission rates. Transmitting uncompressed, high-resolution biomedical audio—such as cough or speech signals—may cause significant

bottlenecks due to increased latency, buffer overflows, or excessive battery drain (Kaur and Singh, 2020). Furthermore, high sampling rates increase memory demands, making them unsuitable for microcontrollers with limited RAM and storage.

In this context, reducing the amount of data to be transmitted, stored, or processed—without compromising critical diagnostic information—is essential for real-time inference and robust system deployment. Compression techniques that preserve the discriminative structure of biomedical signals while reducing dimensionality offer a promising solution for embedded health applications.

2.5.1 Downsampling

DS is a temporal compression technique that reduces signal resolution by resampling the original waveform to a lower number of samples. While it is computationally efficient and simple to implement, DS inherently discards part of the original signal, leading to spectral degradation (Bent et al., 2021; Casaseca-de-la Higuera et al., 2015). According to the Nyquist–Shannon sampling theorem, reducing the sampling rate to f_s limits the maximum representable frequency to $f_s/2$, which effectively eliminates all spectral content above this threshold (Oppenheim and Schaffer, 1999).

This loss of high-frequency components is a known limitation of DS and may be considered when dealing with audio signals, as cough and speech contain diagnostically relevant information across a broad frequency range (Sharan, 2022). The effect becomes more pronounced at aggressive downsampling rates, where significant portions of mid- and high-frequency content may be lost.

In our experiments, each 1 s audio epoch was downsampled using linear interpolation to fixed lengths corresponding to target resolutions (1,000, 500, 200, and 100 Hz). This method was selected due to its simplicity and served as a baseline for comparison with Compressive Sensing.

2.5.2 Compressive Sensing

Compressive Sensing (CS) is an advanced signal compression technique that enables the acquisition and reconstruction of signals from fewer samples than traditional Nyquist-based methods. CS exploits signal sparsity or compressibility in specific domains (e.g., frequency or wavelet domains) to reconstruct signals from fewer measurements without substantial information loss (Candès and Wakin, 2008; Rivera-Flor et al., 2022).

The suitability of speech and cough signals for CS is rooted in their inherent sparsity within specific domains. CS leverages this property, where a signal can be reconstructed from fewer

measurements if it has a sparse representation (few significant coefficients) in a suitable basis (Kodraši and Bourlard, 2020; Candès and Wakin, 2008). Speech signals are known for their spectro-temporal sparsity, meaning their energy is concentrated in specific frequency bands and time instances (Kodraši and Bourlard, 2020). This characteristic, arising from phenomena like formant transitions and pauses, makes speech highly compressible and thus well-suited for CS. Cough signals, sharing physiological similarities with speech, also exhibit sparsity in frequency domains, a property exploited by techniques like Mel-frequency Cepstral Coefficients (MFCCs) for feature extraction (Sharan et al., 2018). This inherent sparsity in both signal types allows CS to enable efficient data acquisition and high-quality reconstruction, proving particularly advantageous for biomedical and health monitoring applications.

Mathematically, given an audio signal vector $\mathbf{x} \in \mathbb{R}^N$, where N represents the number of original time-domain samples in a 1 s recording ($N = 16,000$ at 16 kHz), CS compresses the original signal into a smaller representation. First, a random matrix $\mathbf{A} \in \mathbb{R}^{S \times N}$ is generated, where $S < N$ is the number of compressed measurements or compressed samples (Rivera-Flor et al., 2022). Subsequently, a single Gaussian random orthonormal measurement matrix $\Phi \in \mathbb{R}^{S \times N}$ is computed, as defined in Equation 2:

$$\Phi = \text{orth}(\mathbf{A}^T)^T \quad (2)$$

Using this matrix, the compressed signal representation $\mathbf{y} \in \mathbb{R}^{S \times 1}$ is obtained by projecting the original signal, as shown in Equation 3:

$$\mathbf{y} = \Phi \mathbf{x} \quad (3)$$

Here, $(\cdot)^T$ denotes the matrix transpose, and $\text{orth}(\cdot)$ indicates the orthonormalization of matrix rows. The variable S directly controls the compression ratio, with a smaller S corresponding to higher compression but potentially increased information loss (Rivera-Flor et al., 2022). The compressed signal representation facilitates signal transmission, storage, and processing (Casaseca-de-la Higuera et al., 2015), which is particularly advantageous in embedded systems for biomedical applications, such as automated cough or speech-based respiratory monitoring.

While CS reconstruction may be computationally intensive, our approach integrates CS directly with a CNN, avoiding explicit signal reconstruction. This end-to-end paradigm minimizes computational overhead, enabling a real-time processing on embedded devices (Xiao et al., 2019; Machidon and Pejović, 2023). Beyond effective data reduction, CS preserves critical information at sub-Nyquist rates (Candès and Wakin, 2008), beneficial for applications with limited bandwidth or storage capabilities. The random nature of CS measurements may provide a degree of privacy protection for sensitive biomedical data, although this should not be interpreted as a formal or quantified privacy guarantee (Djelouat et al., 2018).

In this work, the CS process was implemented in the time domain using a single orthonormal measurement matrix, generated once per experimental run. This matrix was applied uniformly to all 1 s signal epochs in the dataset. Each epoch was individually projected into a lower-dimensional space defined by the desired compression size, producing compressed representations with fixed length.

2.5.3 Visual analysis of compression strategies

To illustrate the impact of each compression method, Figures 2, 3 present waveform and spectrogram views of a COVID-19 positive cough signal from the CCS dataset at different compression levels.

In CS-based compression (Figure 2), the waveform changes progressively with increasing compression. However, core temporal structures remain present, and the spectrograms still exhibit coarse spectral patterns and low-frequency components. This suggests that, despite the loss of high-frequency detail, CS preserves enough discriminative structure to support classification—particularly for cough signals, which often contain relevant information in lower spectral bands.

In contrast, DS (Figure 3) exhibits progressive removal of high-frequency information as the sampling rate decreases. At 500 Hz, the spectrogram displays only the lowest spectral bands, as content above 250 Hz is inherently discarded due to the Nyquist limit. This spectral truncation, combined with reduced waveform resolution, results in a simplified temporal structure that may limit the classifier's ability to extract discriminative features.

These visual differences qualitatively support the motivation for comparing CS and DS as compression strategies, especially in scenarios where embedded audio processing demands both efficiency and preservation of discriminative information.

2.6 Experimental tests

Based on the datasets described above (CCS and CSS), we conducted a series of experimental tests to assess the impact of signal compression strategies on the classification of COVID-19 from biomedical audio. Cough and speech recordings were analyzed separately to assess the robustness of the proposed pipeline under different vocal conditions. All experiments were conducted separately for each dataset (CCS and CSS), and cross-dataset generalization was not assessed in this study.

Both CS and DS were applied to compress each 1 s signal epoch to predefined lengths. The resulting representations were used directly as input to the classification model, allowing for a systematic comparison between methods under same training conditions.

Four compression levels were tested: $S = \{1000, 500, 200, 100\}$, covering a range of temporal resolutions (Casaseca-de-la Higuera et al., 2015). The uncompressed baseline corresponds to $S = N = 16000$ samples (1 s at 16 kHz). The corresponding compression ratios are therefore $S/N = \{0.0625, 0.03125, 0.0125, 0.00625\}$. All other experimental parameters, including model architecture and training setup, were kept fixed to ensure fair and reproducible comparisons.

It is worth noting that the inclusion of 200 Hz and 100 Hz sampling rates, despite being close to the lower limit of human hearing, was an intentional choice to explore the feasibility of extreme compression levels for applications in resource-constrained environments. Previous studies have shown that important diagnostic information related to respiratory diseases is often concentrated in low-frequency bands (Ghrabli et al., 2024). Therefore, even under aggressive compression, relevant acoustic cues remain accessible to the classifier, supporting the viability of low-rate processing for audio-based health screening.

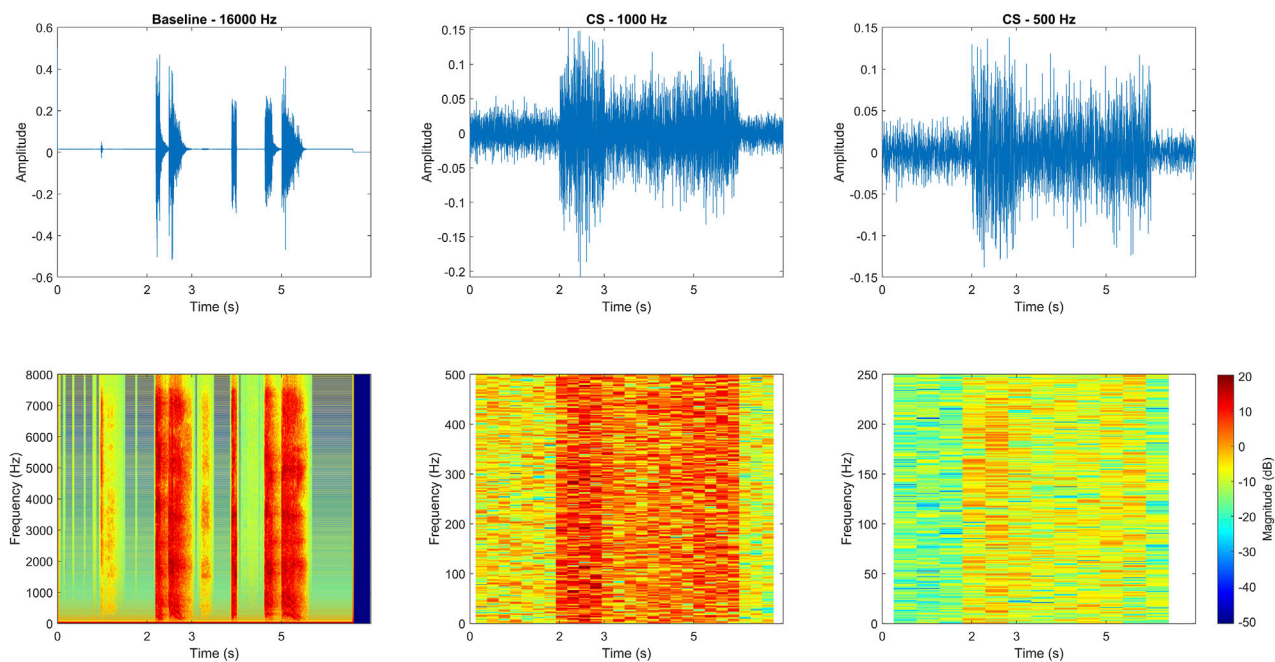


FIGURE 2
Example of a COVID-19 positive cough signal under different CS compression levels. Top: waveforms. Bottom: spectrograms. Despite reduced spectral detail at lower resolutions, coarse temporal and low-frequency patterns remain observable.

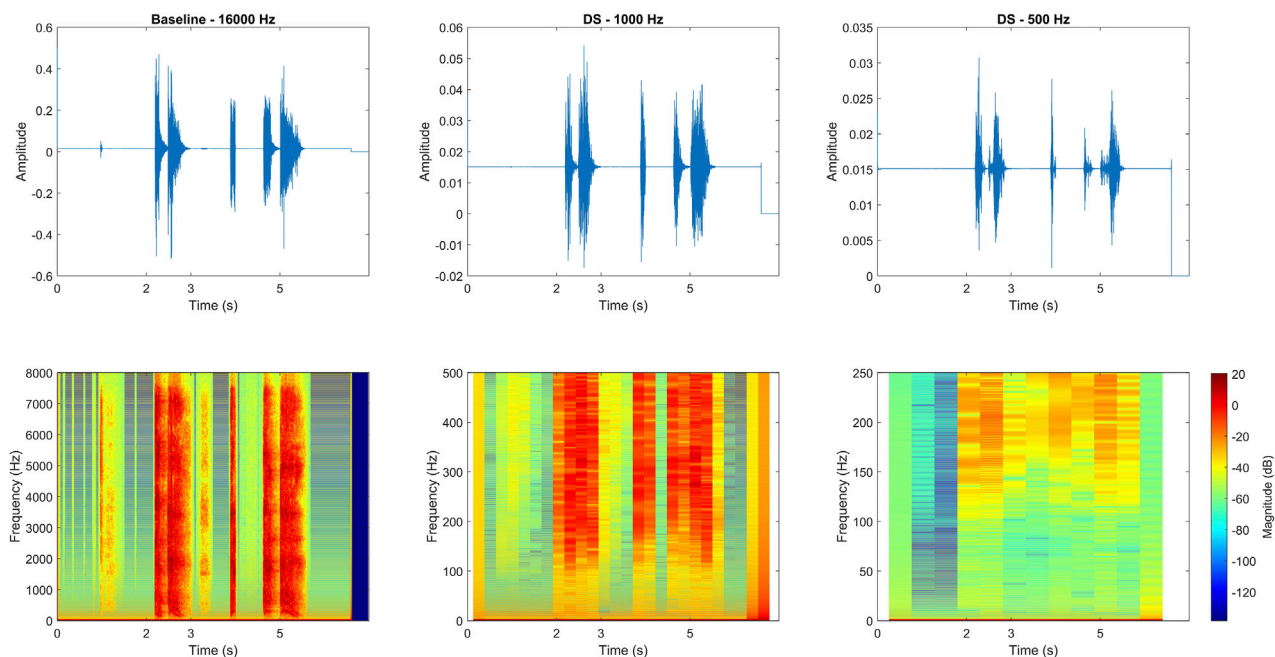


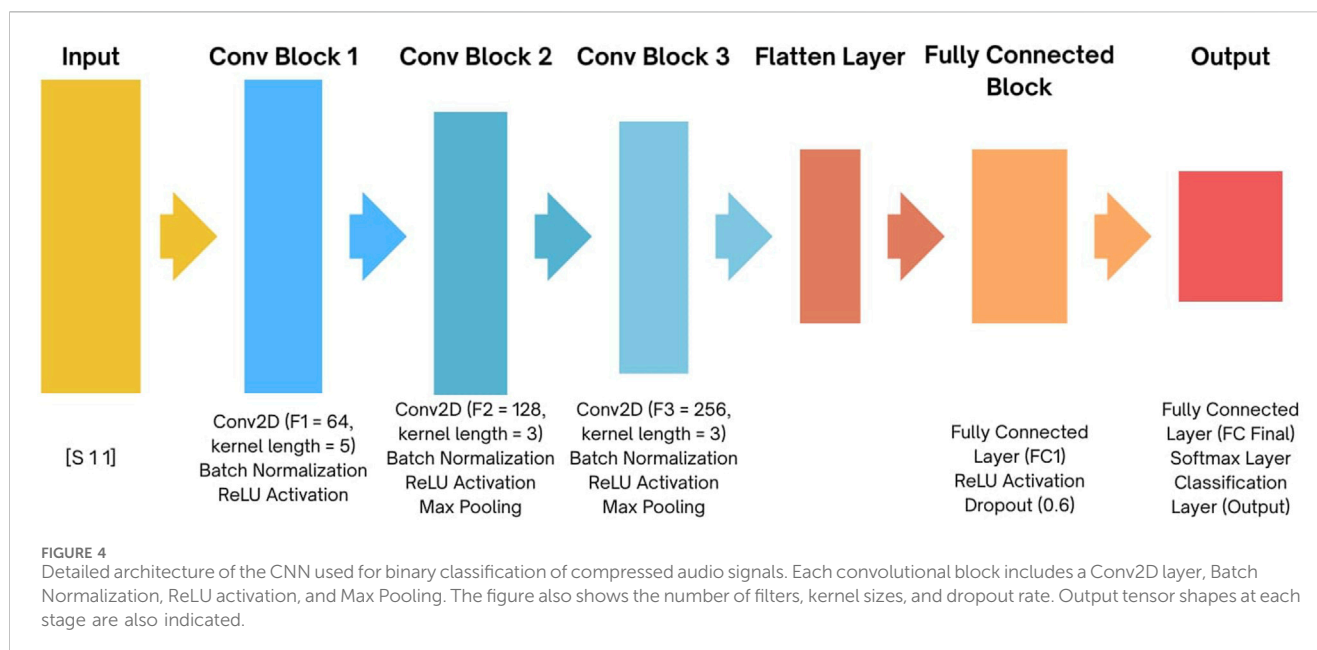
FIGURE 3
Example of a COVID-19 positive cough signal under different compression levels using DS. Top: waveforms. Bottom: spectrograms.

2.7 Classification and evaluation

The compressed audio signals were directly input into a compact time-domain CNN. The proposed architecture (Figure 4) is a compact CNN designed for classification of

compressed time-domain audio signals (Abdoli et al., 2019; Lee et al., 2022).

The CNN comprises three convolutional blocks. The first block used 64 filters with a kernel size of 5, followed by batch normalization, ReLU activation, and max pooling. The second



and third blocks employed 128 and 256 filters, respectively, both with kernel size 3, and included the same normalization, activation, and pooling steps. After flattening, the output passed through a fully connected layer with 64 units, ReLU activation, and dropout ($p = 0.6$) to reduce overfitting. The final layer used softmax for binary classification. A total of 40 training epochs were used, with the Adam optimizer (learning rate = 10^{-4} , batch size = 64), and L2 regularization ($\lambda = 0.001$) applied to improve generalization. The original dataset splits provided by the ComParE challenge were preserved across all experiments, ensuring consistency in evaluation.

Experiments were repeated 30 times with random shuffles to ensure robustness. Performance was evaluated using Accuracy (ACC) and weighted F1-score (F1-weighted).

2.8 Statistical analysis

To evaluate associations between qualitative variables, we employed the Chi-squared (χ^2) statistical test, as it provides a reliable measure to detect significant differences between observed and expected frequencies within categorical variables (McHugh, 2013). Differences in group means for quantitative variables were also assessed at a significance level of $p < 0.05$.

Before conducting further statistical tests, we verified the normality of data distributions using the Shapiro-Wilk test, since normality influences the selection of appropriate statistical methods (Mishra et al., 2019). Given that most of our datasets exhibited non-normal distributions, we applied non-parametric tests. Specifically, the Wilcoxon signed-rank test was used for comparisons involving two related samples, and the Friedman test was applied when multiple related groups required comparison (Demšar, 2006). When significant differences were identified by the Friedman test, *post hoc* pairwise comparisons were conducted using Dunn's test with Bonferroni correction to

adjust for multiple comparisons, thereby controlling the family-wise error rate. For all statistical analyses, a threshold of $p < 0.05$ was utilized to indicate statistical significance.

3 Results

3.1 CCS dataset

3.1.1 Demographic and clinical characteristics of participants in CCS

As mentioned, the CCS dataset comprises cough audio signals and clinical variables such as gender, age group, smoking habits, symptoms, and medical history. Using this dataset, we analyzed the association between these variables and COVID-19 infection through the Chi-squared test.

A total of 725 subjects were included (without data augmentation), of whom 158 (21.79%) tested positive for COVID-19 (+), and 567 (78.21%) tested negative (−), as shown in Table 2. Among the positive cases, 86 (54.43%) were female, 69 (43.67%) male, and 3 (1.90%) did not disclose their gender. Significant associations with COVID-19 positivity were observed for sex ($p = 0.0326$), age group ($p = 0.0034$), and the presence of symptoms ($p < 0.05$). In particular, higher proportions of COVID-19 positive cases were observed among females and younger adults aged 20–49 years. No significant associations were found for smoking habits ($p = 0.9815$) or medical history ($p = 0.0793$). These results suggest that symptomatic presentation, female sex, and age may be relevant factors in COVID-19 detection based on cough analysis.

3.1.2 Baseline performance in CCS

Figure 5 presents the classification results of the CNN model using uncompressed cough signals from the CCS dataset, sampled at 16,000 Hz. Boxplots summarize Accuracy (ACC) and weighted F1-score across 30 independent runs. The model achieved a mean ACC

TABLE 2 Association of sex, age, and clinical characteristics with COVID-19 infection using the CCS dataset (before data augmentation).

	COVID-19 (+)	COVID-19 (-)	p-value
	n = 158	n = 567	
Sex			0.0326 ^{*,a}
F	86	242	
M	69	322	
Unspecified	3	3	
Age			0.0034 ^{*,a}
0–19	4	10	
20–29	33	50	
30–39	37	131	
40–49	37	120	
50–59	33	106	
60–69	5	104	
70–79	8	42	
80–89	0	2	
>90	0	1	
Unspecified	1	1	
Smoking			0.9815 ^a
Never	90	352	
Ex	44	123	
1 to 10	13	50	
11 to 20	5	23	
21+	1	4	
Unspecified	5	15	
Medical history			0.0793 ^a
No	115	358	
Yes	43	209	
Symptoms			$\ll 0.05^{*,a}$
No	29	302	
Yes	129	265	

* Significant values $p < 0.05$.

^a Chi-squared test.

of approximately 78% and a weighted F1-score close to 0.75, with low variability. Outliers were rare, and the baseline configuration was thus adopted as a reference for subsequent comparisons.

3.1.3 Comparison between Compressive Sensing and downsampling strategies in CCS

Figure 6 illustrates the model's classification performance on compressed cough signals from the CCS dataset at sampling frequencies of 1,000, 500, 200, and 100 Hz. Two compression

strategies were evaluated: CS and DS. Results for ACC and weighted F1-score are shown across 30 independent runs.

Across all frequencies, CS yielded higher accuracy and F1-score values than DS. At 1,000 Hz, CS reached a median accuracy close to 80%, whereas DS remained around 75%. As the frequency decreased, performance declined for both methods, but the drop was substantially more pronounced in DS. CS maintained better classification outcomes across all conditions, with lower variability between runs.

To statistically validate these differences, the Shapiro–Wilk test was first applied to assess normality. Since the distributions were non-normal, the Wilcoxon signed-rank test was used to compare CS and DS at each frequency. In all cases, the null hypothesis was rejected ($p < 0.05$), confirming the superiority of CS. These findings highlight CS as a more robust strategy for signal compression in embedded systems, capable of preserving classification performance even under severe temporal reduction.

3.1.4 Statistical comparison between compressed signals and baseline performance in CCS

Figure 7 presents the statistical comparisons between the baseline condition (16,000 Hz) and compressed sampling frequencies for both CS and DS, using the Friedman test followed by *post hoc* pairwise analysis. The heatmaps show the p-values for each frequency pair; darker colors indicate lower p-values, and asterisks (*) denote statistically significant differences ($p < 0.05$), adjusted using the Bonferroni correction to account for multiple comparisons.

Results indicate that, for CS, performance at all tested frequencies was statistically equivalent to the baseline (16,000 Hz), suggesting that classification accuracy was preserved even under high compression levels. In contrast, DS exhibited statistically significant differences from the baseline in almost all evaluated frequencies. This finding suggests that DS fails to maintain classification performance, even at moderate compression rates, resulting in notable degradation in both accuracy and F1-score.

These findings reinforce the suitability of CS as a compression strategy for cough signals. Preservation of statistical equivalence with the uncompressed baseline, especially at moderate sampling rates, supports its integration into resource-constrained or embedded systems.

3.2 CSS dataset

3.2.1 Demographic and clinical characteristics of participants in CSS

As in Section 3.1, a demographic and clinical analysis was conducted on the CSS dataset, which contains speech audio recordings from 893 individuals labeled as COVID-19 positive (+) or negative (-). Among the participants, 308 (34.49%) tested positive, while 585 (65.51%) tested negative, as summarized in Table 3. These figures correspond to the original dataset, prior to any data augmentation.

The Chi-squared test was applied to evaluate associations between COVID-19 status and clinical variables. In the COVID-19 positive group, 134 (43.51%) were female, 173 (56.17%) male, and 1 (0.32%) unspecified. No statistically significant association was found between sex and infection status ($p = 0.9751$). In contrast, significant associations were observed for age group, medical history, and symptom presence (all $p < 0.05$).

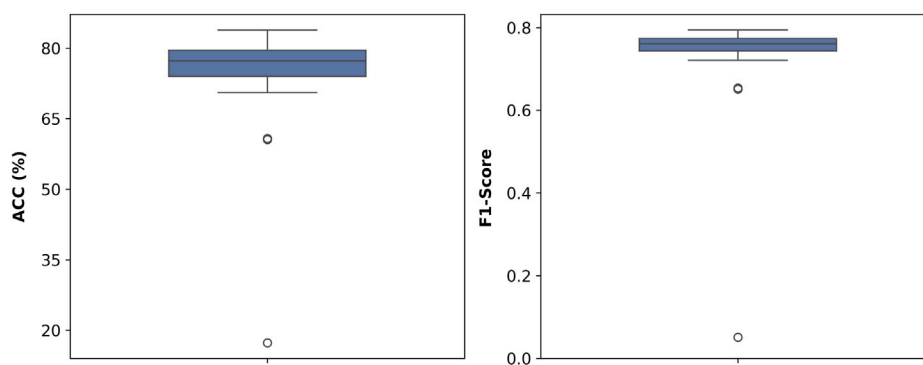


FIGURE 5
Baseline performance of the CNN model using uncompressed cough signals (16,000 Hz, CCS dataset).

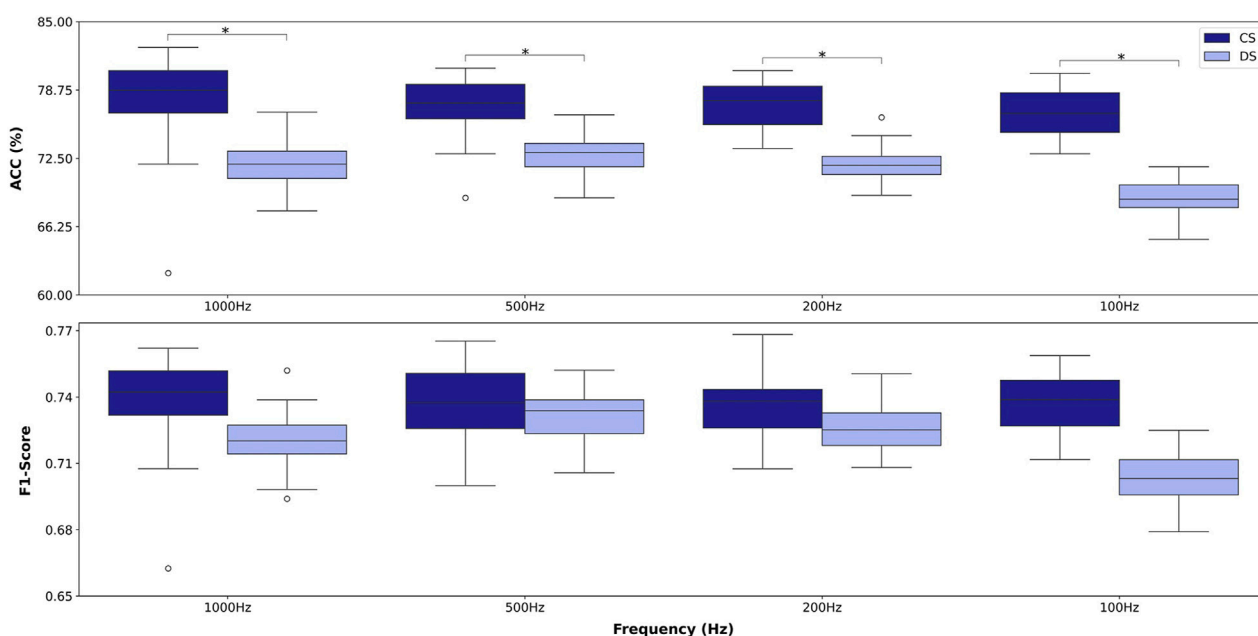


FIGURE 6
Performance comparison between Compressive Sensing (CS, dark blue) and Downsampling (DS, light blue) across multiple sampling frequencies. Boxplots show accuracy (top) and F1-score (bottom) of the CNN model on the CCS dataset. Asterisks (*) indicate statistically significant differences ($p < 0.05$).

Consistent with findings from the CCS dataset, the presence of symptoms remained a strong discriminator of COVID-19 positivity, reinforcing its diagnostic relevance across both cough and speech modalities.

3.2.2 Baseline performance in CSS

Figure 8 shows the baseline performance of the CNN model using uncompressed speech signals from the CSS dataset, sampled at 16,000 Hz. The results were obtained from 30 independent executions using the original data partitions and balanced training set.

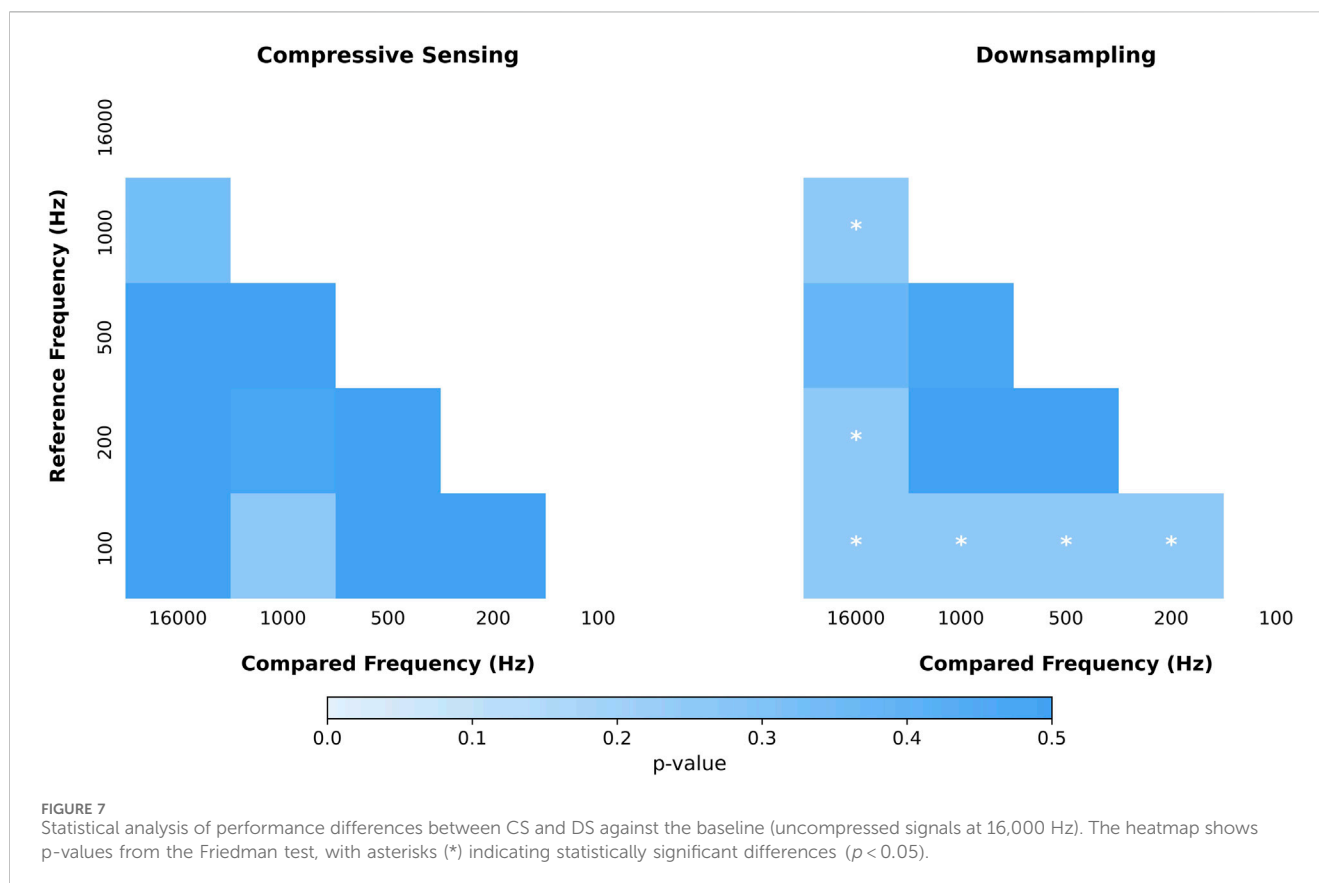
The model reached a median accuracy of approximately 69% and a weighted F1-score close to 0.66, with low variability across runs. Most results are concentrated in a narrow range, indicating

that the model was able to generalize consistently under this configuration. A few outliers were observed, but without strong impact on overall performance metrics.

Compared to the CCS dataset, the baseline performance for speech signals was slightly lower, which may be explained by the more complex and heterogeneous nature of speech compared to cough sounds. These results serve as a reference for evaluating the effects of signal compression in the next sections.

3.2.3 Comparison between Compressive Sensing and downsampling strategies in CSS

Figure 9 presents the classification performance of CS and DS at four compression levels (1,000, 500, 200, and 100 Hz) on the CSS dataset. Each configuration was evaluated over 30 independent runs,



and the distributions of accuracy and weighted F1-score are shown as boxplots. As in the CCS dataset, the Shapiro–Wilk test indicated non-normal distributions, and the Wilcoxon signed-rank test was used to compare CS and DS at each frequency.

At 1,000 Hz, no statistically significant difference was observed between the methods. From 500 Hz and below, CS achieved superior results in both metrics, with statistically significant differences ($p < 0.05$). While performance declined for both strategies at lower frequencies, CS maintained higher median values in accuracy and F1-score at 500, 200, and 100 Hz. At 1,000 Hz, DS produced a slightly higher median F1-score, whereas CS achieved better accuracy.

Compared to the CCS dataset, overall accuracy and F1-score values were slightly lower in CSS across all configurations, possibly reflecting the acoustic variability of speech signals. Still, the comparative advantage of CS over DS remained consistent, particularly under high compression.

3.2.4 Statistical comparison between compressed signals and baseline performance in CSS

Figure 10 shows the pairwise statistical comparisons between the baseline (16,000 Hz) and the compressed versions using Compressive Sensing and Downsampling on the CSS dataset. The Friedman test followed by *post hoc* analysis was applied, and the heatmaps display the p-values between each pair of sampling frequencies. Asterisks indicate significant differences ($p < 0.05$), and darker shades represent lower p-values.

For CS, only the 100 and 1,000 Hz configurations showed statistically significant differences from the baseline. This indicates that performance was preserved at 500 and 200 Hz, even under compression. On the other hand, DS showed significant differences at most part of tested frequencies, including 500 Hz, suggesting that even moderate downsampling led to performance degradation in speech classification.

Overall, the results reinforce the robustness of CS in speech-based classification, particularly under higher compression rates.

4 Discussion

Since the onset of COVID-19, several studies have investigated non-invasive screening methods based on audio biomarkers, particularly cough and speech signals. Traditional acoustic features such as MFCC, Zero-Crossing Rates (ZCR), and Spectral Entropy have been widely employed, often in combination with classical Machine Learning (ML) or Deep Learning (DL) models (Brown et al., 2020; Sharma et al., 2022; Pahar et al., 2022; Villa-Parra et al., 2022). For example (Sharma et al., 2022), reported high classification accuracy using textural features, while (Pahar et al., 2022) demonstrated that could achieve Receiver Operating Characteristic Area Under the Curve (AUC-ROC) scores above 0.90 in COVID-19 detection tasks using cough and speech.

Beyond the specific context of COVID-19, audio biomarkers have gained increasing relevance in biomedical signal processing. Speech, in particular, has been investigated for diagnosing various

TABLE 3 Association of sex, age, and clinical characteristics with COVID-19 infection using the CSS dataset (before data augmentation).

	COVID-19 (+)	COVID-19 (-)	p-value
	n = 308	n = 585	
Sex			0.9751 ^a
F	134	251	
M	173	330	
Unspecified	1	4	
Age			≪ 0.05 ^{*,a}
0–19	8	9	
20–29	35	50	
30–39	53	137	
40–49	114	124	
50–59	80	112	
60–69	10	105	
70–79	8	43	
80–89	0	3	
>90	0	1	
Unspecified	0	1	
Smoking			0.2680 ^a
Never	219	367	
Ex	65	126	
1 to 10	12	51	
11 to 20	7	21	
21+	1	4	
Unspecified	4	16	
Medical history			≪ 0.05 ^{*,a}
No	257	373	
Yes	51	212	
Symptoms			≪ 0.05 ^{*,a}
No	64	312	
Yes	244	273	

* Significant values $p < 0.05$.
^a Chi-squared test.

diseases, including Parkinson’s, depression, and respiratory syndromes (Botelho et al., 2024). These findings highlight the growing clinical potential of audio signals and motivate further exploration of efficient signal processing techniques for healthcare applications (Cauzinille et al., 2024).

In this study, we compared two temporal compression strategies—DS and CS—applied directly to raw audio signals. Instead of relying on handcrafted features, we used an end-to-end Convolutional Neural Network (CNN) architecture that processes the compressed waveforms as input. This approach

simplifies the processing pipeline and reduces computational load, which is beneficial for deployment in embedded or portable systems (Kaur and Singh, 2020; Santos, 2023).

The proposed end-to-end methodology, processing compressed time-domain audio signals directly with a CNN, offers advantages for deployment in resource-constrained environments compared to approaches requiring extensive signal reconstruction or feature engineering. Operating directly on compressed measurements reduces computational overhead, a critical benefit for embedded systems demanding high computational efficiency and low latency (Pietrolaj and Blok, 2024). Furthermore, integrating signal compression techniques, specifically CS, not only facilitates effective data reduction but also preserves essential discriminative information. This characteristic is particularly beneficial for applications constrained by bandwidth or storage limitations (Saeed et al., 2025), thereby enhancing the practical applicability and accessibility of our diagnostic approach.

CS demonstrated superior performance compared to DS, especially under aggressive compression levels (e.g., 200 Hz and 100 Hz). These results are consistent with previous findings showing that CS is effective at preserving relevant information in compressed biomedical signals (Casaseca-de-la Higuera et al., 2015; Prabhavathi et al., 2023; Wang et al., 2016). By exploiting the underlying sparsity of audio signals, CS retains key temporal patterns that are essential for robust classification.

When comparing these results to prior studies, it is clear that several models in the literature reported a better performance. For instance (Shati et al., 2023; Aytekin et al., 2023), achieved AUC values above 0.80 using cough and speech signals. These studies rely on spectrogram-based inputs and sophisticated architectures such as Hierarchical Spectrogram Transformers (HST) or use high-dimensional handcrafted features combined with classical ML classifiers. Similarly (Sharma et al., 2022), achieved ACC of 98.9% in a binary task and 72.2% in a five-class task by employing textural features like Local Binary Patterns (LBP) and Haralick on spectrograms. Likewise (Pahar et al., 2022), reached AUC-ROC values of 0.98, 0.94, and 0.92 for cough, breath, and speech, respectively, leveraging deep learning models. In contrast, our pipeline intentionally adopts a waveform-based end-to-end approach with aggressive temporal compression, aiming to prioritize computational efficiency and simplicity over classification performance. This naturally imposes a different performance ceiling. Our models exhibited lower AUC, sensitivity, and specificity but this outcome aligns with the study’s primary objective: to investigate whether compressed waveform representations can retain sufficient discriminative information for lightweight, embedded applications. Additionally, most state-of-the-art methods rely on curated datasets with manual validation of cough segments, complex segmentation, or high-resolution spectral inputs, which are not feasible in real-time or constrained environments (Schuller et al., 2021; Casanova et al., 2021; Aytekin et al., 2023; Pahar et al., 2022). Our pipeline avoids such steps, focusing on robustness, simplicity, and operational viability for embedded systems.

These comparisons reinforce the trade-off embedded in our design choice: while spectrogram-based models with sophisticated architectures yield higher absolute performance, our work demonstrates the potential viability of a simplified waveform-

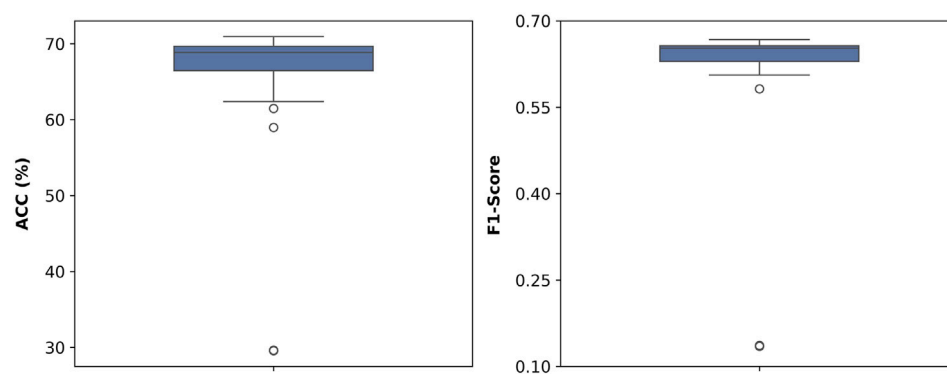


FIGURE 8
Baseline performance of the CNN model using uncompressed speech signals (16,000 Hz, CSS dataset).

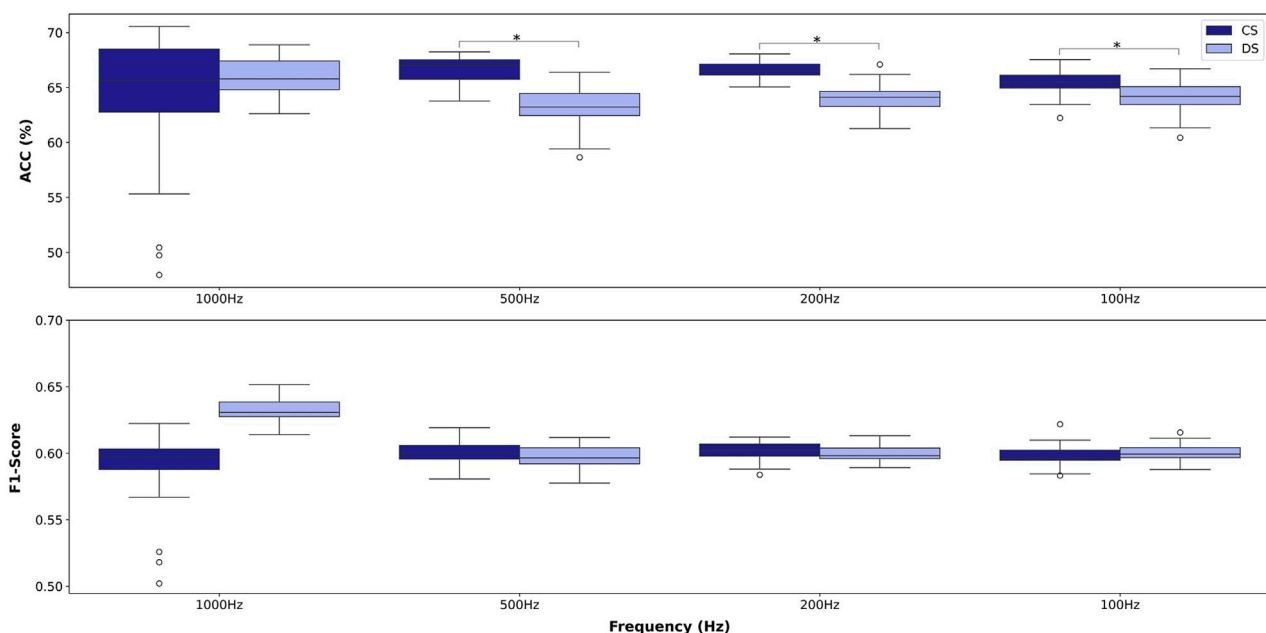


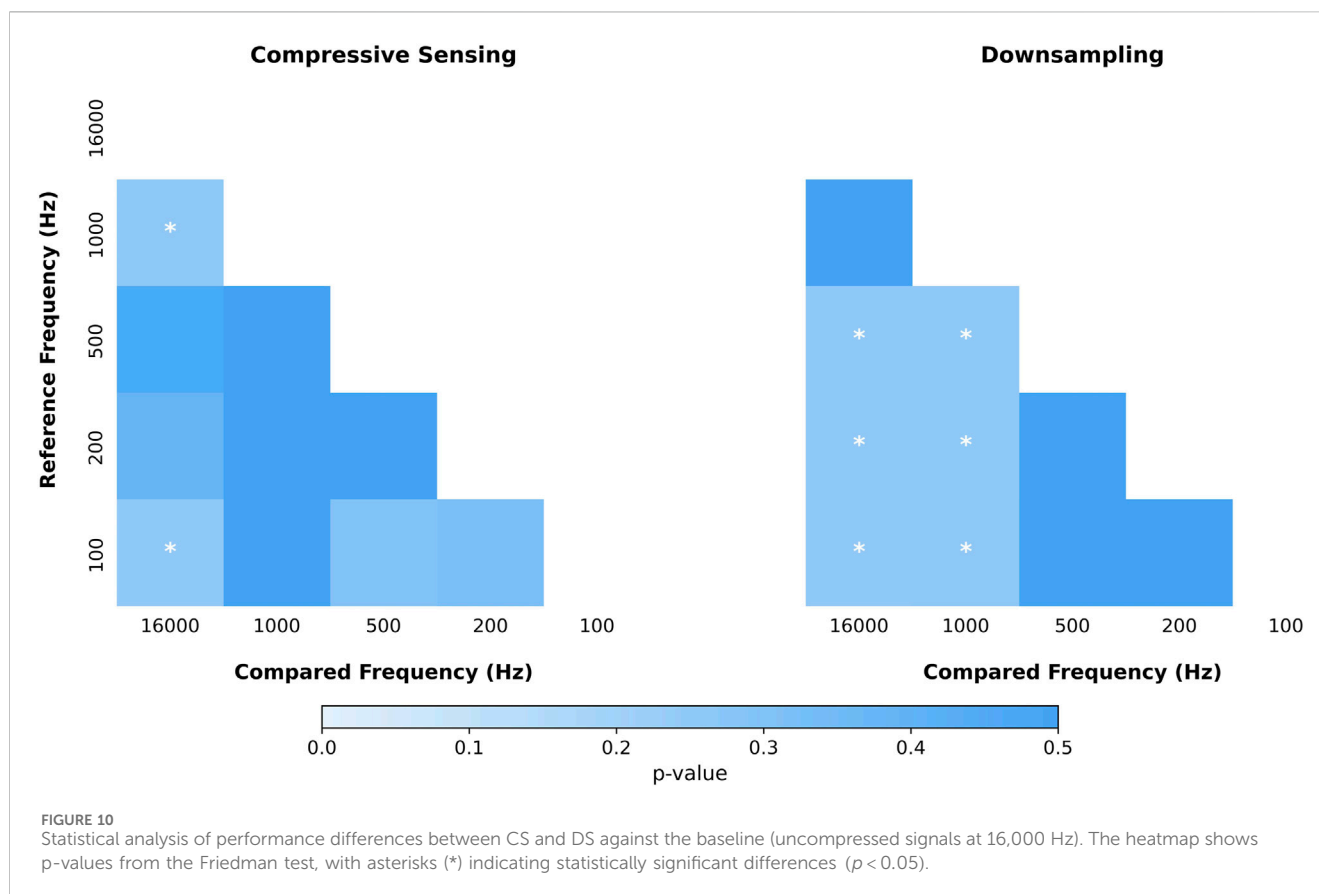
FIGURE 9
Performance comparison between Compressive Sensing (CS, dark blue) and Downsampling (DS, light blue) across multiple sampling frequencies. Boxplots show accuracy (top) and F1-score (bottom) of the CNN model on the CSS dataset. Asterisks (*) indicate statistically significant differences ($p < 0.05$).

based pipeline operating directly on aggressively compressed signals—a relevant contribution for deployment scenarios constrained by bandwidth, memory, or processing power.

Notably, our approach remained promissory even at sampling rates as low as 200 Hz and 100 Hz. Although these values are close to the lower limit of human hearing, prior studies have shown that cough signals contain diagnostically relevant information within these low-frequency bands. Sharan (2022) observed that coughs present spectral components starting around 80 Hz, while (Ghrabli et al., 2024) identified specific low-frequency patterns linked to respiratory pathologies. These findings reinforce the feasibility of low-rate audio analysis, with our results indicating that the core discriminative information

needed for classification is preserved despite aggressive compression—thereby validating this approach for bandwidth-constrained applications.

It is important to highlight that cough and speech signals responded differently to compression. Cough signals maintained higher classification performance even at low sampling rates, which may be attributed to their broadband spectral characteristics that are more easily preserved under compression. In contrast, speech signals—characterized by more complex and fine-grained temporal dynamics—were more sensitive to compression-induced distortion, particularly under DS. Similar behavior was observed by (Shen et al., 2024), which also reported better performance when using compressed cough signals for COVID-19 detection.



Recent studies, however, have raised important concerns regarding the clinical robustness of cough-based COVID-19 detection. For instance (Kim et al., 2024), demonstrated that model performance may deteriorate significantly across viral variants, with AUC values dropping from 0.93 for Alpha to 0.55 for Omicron. Similarly (Coppock et al., 2024), reported that when controlling for confounding factors, audio-based classifiers provided limited diagnostic value beyond simple symptom questionnaires, with AUC decreasing from 0.85 to 0.62. It is important to note that the dataset used in our study was collected in 2020, before the emergence of variants of concern such as Alpha, Delta, or Omicron. Therefore, the samples primarily correspond to the original Wuhan strain (He et al., 2023). Our results indicate that cough-based detection should not be interpreted as a definitive clinical diagnostic tool, but rather as a computational approach that may offer value in specific scenarios such as rapid screening, self-assessment, population-level monitoring, or resource-limited deployments where access to laboratory diagnostics is limited. Accordingly, the findings presented in our study should be interpreted within a methodological scope, with the feasibility of compressed waveform analysis being emphasized rather than the clinical conclusiveness of cough-based detection.

Although the experiments were conducted offline, the results reinforce the feasibility of using compressed audio representations in real-time health monitoring. Processing directly in the time domain, without relying on handcrafted features, may reduce system complexity and enable deployment on constrained

platforms such as embedded or mobile devices (Kaur and Singh, 2020; Santos, 2023). Furthermore, the consistent performance of CS under aggressive compression conditions highlights its practical relevance for real-time health monitoring applications with limited computational resources.

5 Conclusion

This study investigated the effects of temporal signal compression techniques—Downsampling and Compressive Sensing—on the performance of a CNN for COVID-19 detection from cough and speech audio signals. The proposed end-to-end approach operates directly on compressed waveforms, enabling a classification pipeline that may reduce computational and memory demands.

Our findings show that Compressive Sensing consistently outperformed Downsampling under higher compression levels (200 Hz and 100 Hz), particularly for cough signals. This suggests that CS better preserves essential discriminative information and that cough may be a more suitable modality for audio-based screening.

The proposed strategy simplifies signal processing by avoiding explicit feature engineering and may reduce computational and memory requirements, favoring implementation in embedded systems. As a limitation, this study was conducted offline using fixed-length segments and dataset-specific audio, which may limit generalizability due to speaker variability, viral evolution, and the

lack of real-time evaluation. While our experiments have focused on COVID-19 detection, future work will investigate the method's applicability to other respiratory conditions with similar acoustic signatures, and evaluate performance across emerging SARS-CoV-2 variants. Future work will also focus on evaluating the method in real-time scenarios to assess performance and generalization in practical applications, including deployment on microcontroller-based platforms. Additional directions include adaptive audio segmentation, integration with mobile devices, and broader validation across diverse populations and environments.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Only credentialed users who sign the data transfer agreement can access the files. Requests to access these datasets should be directed to COVID-19 Sounds (<https://www.covid-19-sounds.org/>).

Author contributions

LS: Writing – original draft, Software. AF: Writing – review and editing. CV: Writing – review and editing. EC: Writing – review and editing. SK: Supervision, Writing – review and editing. TB: Supervision, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brazil): 012/2020.

References

- Abdoli, S., Cardinal, P., and Koerich, A. L. (2019). End-to-end environmental sound classification using a 1d convolutional neural network. *Expert Syst. Appl.* 136, 252–263. doi:10.1016/j.eswa.2019.06.040
- Aytenkin, I., Dalmaz, O., Gonc, K., Ankishan, H., Saritas, E. U., Bagci, U., et al. (2023). Covid-19 detection from respiratory sounds with hierarchical spectrogram transformers. *IEEE J. Biomed. Health Informatics* 28, 1273–1284. doi:10.1109/JBHI.2023.3339700
- Bent, B., Lu, B., Kim, J., and Dunn, J. P. (2021). Biosignal compression toolbox for digital biomarker discovery. *Sensors* 21, 516. doi:10.3390/s21020516
- Boashash, B. (2015). *Time-frequency signal analysis and processing: a comprehensive reference*. Academic Press.
- Botelho, C., Abad, A., Schultz, T., and Trancoso, I. (2024). Speech as a biomarker for disease detection. *IEEE Access* 12, 184487–184508. doi:10.1109/access.2024.3506433
- Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., et al. (2020). "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining*, 3474–3484.
- Candès, E. J., and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine* 25, 21–30. doi:10.1109/msp.2007.914731
- Casanova, E., Candido, A., Fernandes, R. C., Finger, M., Gris, L. R. S., Ponti, M. A., et al. (2021). "Transfer learning and data augmentation techniques to the covid-19 identification tasks in compare 2021," in *22nd annual conference of the international speech communication association, INTERSPEECH 2021*, 4301–4305.
- Casaseca-de-la Higuera, P., Lesso, P., McKinstry, B., Pinnock, H., Rabinovich, R., McCloughan, L., et al. (2015). "Effect of downsampling and compressive sensing on audio-based continuous cough monitoring," in *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (IEEE), 6231–6235.
- Cauzinille, J., Favre, B., Marxer, R., and Rey, A. (2024). Applying machine learning to primate bioacoustics: review and perspectives. *Am. J. Primatology* 86, e23666. doi:10.1002/ajp.23666
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artificial Intelligence Research* 16, 321–357. doi:10.1613/jair.953
- Coppock, H., Nicholson, G., Kiskin, I., Koutra, V., Baker, K., Budd, J., et al. (2024). Audio-based ai classifiers show no evidence of improved covid-19 screening over simple symptoms checkers. *Nat. Mach. Intell.* 6, 229–242. doi:10.1038/s42256-023-00773-8
- Deepika, J., Rajan, C., and Senthil, T. (2021). Security and privacy of cloud-and-iot-based medical image diagnosis using fuzzy convolutional neural network. *Comput. Intell. Neurosci.* 2021, 6615411. doi:10.1155/2021/6615411
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learning Research* 7, 1–30.
- Diab, M. S., and Rodriguez-Villegas, E. (2022). Embedded machine learning using microcontrollers in wearable and ambulatory systems for health and care applications: a review. *IEEE Access* 10, 98450–98474. doi:10.1109/access.2022.3206782
- Djelouat, H., Amira, A., and Bensaali, F. (2018). Compressive sensing-based iot applications: a review. *J. Sens. Actuator Netw.* 7, 45. doi:10.3390/jsan7040045

Acknowledgements

The authors acknowledge the financial support from Global Affairs Canada, CAPES/Brazil, and FACITEC/Brazil.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author SK declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Ghrabli, S., Elgendi, M., and Menon, C. (2024). Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments. *Sci. Rep.* 14, 593. doi:10.1038/s41598-023-50371-2
- He, D., Cowling, B. J., Ali, S. T., and Stone, L. (2023). Rapid global spread of variants of concern of sars-cov-2. *IJID Reg.* 7, 63–65. doi:10.1016/j.ijregi.2022.12.005
- Iwana, B. K., and Uchida, S. (2021). An empirical survey of data augmentation for time series classification with neural networks. *Plos One* 16, e0254841. doi:10.1371/journal.pone.0254841
- Jayaram, R., and Prabakaran, S. (2021). Onboard disease prediction and rehabilitation monitoring on secure edge-cloud integrated privacy preserving healthcare system. *Egypt. Inf. J.* 22, 401–410. doi:10.1016/j.eij.2020.12.003
- Jo, A.-H., and Kwak, K.-C. (2022). Diagnosis of depression based on four-stream model of bi-lstm and cnn from audio and text information. *IEEE Access* 10, 134113–134135. doi:10.1109/access.2022.3231884
- Johns Hopkins University (2022). Covid-19 dashboard. Available online at: <https://coronavirus.jhu.edu/map.html> (Accessed July 1, 2022).
- Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., and Hussain, S. (2023). Effective class-imbalance learning based on smote and convolutional neural networks. *Appl. Sci.* 13, 4006. doi:10.3390/app13064006
- Kapoor, T., Pandhi, T., and Gupta, B. (2022). Cough audio analysis for covid-19 diagnosis. *SN Comput. Sci.* 4, 125. doi:10.1007/s42979-022-01522-1
- Kaur, N., and Singh, M. (2020). A lossless compression and encryption mechanism for remote ecg applications. *Future Gener. Comput. Syst.* 108, 1012–1026. doi:10.1016/j.future.2019.12.045
- Kim, J., Choi, Y. S., Lee, Y. J., Yeo, S. G., Kim, K. W., Kim, M. S., et al. (2024). Limitations of the cough sound-based covid-19 diagnosis artificial intelligence model and its future direction: longitudinal observation study. *J. Med. Internet Res.* 26, e51640. doi:10.2196/51640
- Kodrasi, I., and Boulard, H. (2020). Spectro-temporal sparsity characterization for dysarthric speech detection. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 28, 1210–1222. doi:10.1109/taslp.2020.2985066
- Lee, J.-N., and Lee, J.-Y. (2023). An efficient smote-based deep learning model for voice pathology detection. *Appl. Sci.* 13, 3571. doi:10.3390/app13063571
- Lee, G.-T., Nam, H., Kim, S.-H., Choi, S.-M., Kim, Y., and Park, Y.-H. (2022). Deep learning based cough detection camera using enhanced features. *Expert Syst. Appl.* 206, 117811. doi:10.1016/j.eswa.2022.117811
- Machidon, A. L., and Pejović, V. (2023). Deep learning for compressive sensing: a ubiquitous systems perspective. *Artif. Intell. Rev.* 56, 3619–3658. doi:10.1007/s10462-022-10259-5
- McHugh, M. L. (2013). The chi-square test of independence. *Biochem. Medica* 23, 143–149. doi:10.11613/bm.2013.018
- Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., and Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Ann. Cardiac Anaesthesia* 22, 67–72. doi:10.4103/aca.ACA_157_18
- Oliveira, B. A., Oliveira, L. C. d., Sabino, E. C., and Okay, T. S. (2020). Sars-cov-2 and the covid-19 disease: a mini review on diagnostic methods. *Rev. Inst. Med. Trop. Sao Paulo*. 62, e44. doi:10.1590/S1678-9946202062044
- Oppenheim, A. V., and Schaffer, R. W. (1999). *Discrete-time signal processing*. second edn. Englewood Cliffs, NJ: Prentice Hall.
- Pahar, M., Klopfer, M., Warren, R., and Niesler, T. (2022). Covid-19 detection in cough, breath and speech using deep transfer learning and bottleneck features. *Comput. Biol. Med.* 141, 105153. doi:10.1016/j.combiomed.2021.105153
- Patel, A., Degadwala, S., and Vyas, D. (2022). “Lung respiratory audio prediction using transfer learning models,” in *2022 sixth international conference on I-SMAC (IoT in social, Mobile, analytics and Cloud)(I-SMAC)* (IEEE), 1107–1114.
- Pietrolaj, M., and Blok, M. (2024). Resource constrained neural network training. *Sci. Rep.* 14, 2421. doi:10.1038/s41598-024-52356-1
- Pleva, M., Martens, E., and Juhar, J. (2022). “Automated covid-19 respiratory symptoms analysis from speech and cough,” in *2022 IEEE 20th jubilee world symposium on applied machine intelligence and informatics (SAMI)* (IEEE). 000127–000132.
- Prabhavathi, C., Kashyap, G. G., and Gagana, N. (2023). “Compressive sensing and its application to speech signal processing,” in *2023 international conference on network, multimedia and information technology (NMITCON)* (IEEE), 1–5.
- Pramono, R. X. A., Imtiaz, S. A., and Rodriguez-Villegas, E. (2016). A cough-based algorithm for automatic diagnosis of pertussis. *PloS One* 11, e0162128. doi:10.1371/journal.pone.0162128
- Rai, P., Kumar, B. K., Deekshit, V. K., Karunasagar, I., and Karunasagar, I. (2021). Detection technologies and recent developments in the diagnosis of covid-19 infection. *Appl. Microbiol. Biotechnol.* 105, 441–455. doi:10.1007/s00253-020-11061-5
- Rivera-Flor, H., Gurve, D., Floriano, A., Delisle-Rodriguez, D., Mello, R., and Bastos-Filho, T. (2022). Cca-based compressive sensing for ssvep-based brain-computer interfaces to command a robotic wheelchair. *IEEE Trans. Instrum. Meas.* 71, 1–10. doi:10.1109/tim.2022.3218102
- Saeed, A., A. Khan, M., Akram, U., Obidallah, J., Jawed, S., and Ahmad, A. (2025). Deep learning based approaches for intelligent industrial machinery health management and fault diagnosis in resource-constrained environments. *Sci. Rep.* 15, 1114. doi:10.1038/s41598-024-79151-2
- Santos, D. e. a. (2023). Enhancing connected health ecosystems through iot-enabled monitoring technologies: a case study of the monit4healthy system. *Sensors* 25, 2292. doi:10.3390/s25072292
- Schuller, B. W., Batliner, A., Bergler, C., Mascolo, C., Han, J., Lefter, I., et al. (2021). The interspeech 2021 computational paralinguistics challenge: Covid-19 cough, covid-19 speech, escalation and primates. *arXiv preprint arXiv:2102.13468*
- Sharan, P. (2022). Automated discrimination of cough in audio recordings: a scoping review. *Front. Signal Process.* 2, 759684. doi:10.3389/frsip.2022.759684
- Sharan, R. V., Abeyratne, U. R., Swarnkar, V. R., and Porter, P. (2018). Automatic group diagnosis using cough sound recognition. *IEEE Trans. Biomed. Eng.* 66, 485–495. doi:10.1109/TBME.2018.2849502
- Sharma, G., Umapathy, K., and Krishnan, S. (2022). Audio texture analysis of covid-19 cough, breath, and speech sounds. *Biomed. Signal Process. Control* 76, 103703. doi:10.1016/j.bspc.2022.103703
- Shati, A., Hassan, G. M., and Datta, A. (2023). “Covid-19 detection system: a comparative analysis of system performance based on acoustic features of cough audio signals,” in *2023 IEEE 22nd international conference on trust, security and privacy in computing and communications (TrustCom)* (IEEE), 2706–2713.
- Shen, J., Zhang, X., Zhang, P., Yan, Y., Zhao, Q., Li, T., et al. (2024). “One-epoch training with single test sample in test time for better generalization of cough-based covid-19 detection model,” in *ICASSP 2024-2024 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (IEEE), 931–935.
- Shi, Y. (2025). A cnn-based approach for classical music recognition and style emotion classification. *IEEE Access* 13, 20647–20666. doi:10.1109/access.2025.3535411
- Verde, L., Pietro, G. D., Ghoneim, A., Alrashoud, M., Al-Mutib, K. N., and Sannino, G. (2021). Exploring the use of artificial intelligence techniques to detect the presence of coronavirus covid-19 through speech and voice analysis. *IEEE Access* 9, 65750–65757. doi:10.1109/ACCESS.2021.3075571
- Villa-Parra, A. C., Criollo, I., Valadão, C., Silva, L., Coelho, Y., Lampier, L., et al. (2022). Towards multimodal equipment to help in the diagnosis of covid-19 using machine learning algorithms. *Sensors* 22, 4341. doi:10.3390/s22124341
- Wang, J.-C., Lee, Y.-S., Lin, C.-H., Wang, S.-F., Shih, C.-H., and Wu, C.-H. (2016). Compressive sensing-based speech enhancement. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 24, 2122–2131. doi:10.1109/taslp.2016.2598306
- Wang, X., Ren, H., Ren, J., Song, W., Qiao, Y., Ren, Z., et al. (2023). Machine learning-enabled risk prediction of chronic obstructive pulmonary disease with unbalanced data. *Comput. Methods Programs Biomed.* 230, 107340. doi:10.1016/j.cmpb.2023.107340
- World Health Organization (2020). Who director-general’s opening remarks at the media briefing on covid-19 - 11 march 2020. Available online at: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020> (Accessed June 30, 2022).
- Worldometer (2024). Brazil covid - coronavirus statistics. Available online at: https://www.worldometers.info/coronavirus/country/brazil/#google_vignette (Accessed April 17, 2025).
- Xiao, J., Hu, F., Shao, Q., and Li, S. (2019). A low-complexity compressed sensing reconstruction method for heart signal biometric recognition. *Sensors* 19, 5330. doi:10.3390/s19235330