Harmonizing Chest Imaging and Cough Sound Analysis: A Multi-Modal Approach for Respiratory Disease Detection

May Rashid †, Hamada Nayel †,††, and Ahmed Taha†

†Computer Science Department, Faculty of Artificial Intelligence, Benha University, Benha 13518, Egypt
††Faculty of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Aldawasir, Saudi Arabia

Summary
In response to the challenges of the global respiratory illness specifically COVID-19 pandemic, this research presents a diagnostic model that integrates chest imaging and sound data for enhanced accuracy. Utilizing carefully curated datasets of chest X-rays, CT scans, and cough recordings, our model offers a comprehensive analysis of visual and auditory cues associated with the disease. The imaging dataset captures critical radiological nuances, enriched by detailed clinical features. Simultaneously, the sound dataset, comprising over 25,000 cough recordings, pioneers the integration of acoustic signatures into diagnostic methodologies. Our preprocessing pipeline employs advanced techniques, including image augmentation and Mel spectrogram transformations, ensuring model adaptability. The model architecture synergizes visual and auditory insights, culminating in a unified diagnostic capability that transcends individual modalities. This research contributes to global COVID-19 efforts by providing a nuanced and comprehensive diagnostic approach. By fusing visual and auditory insights, our model addresses the urgency and accuracy required in the face of the pandemic, offering a path towards more effective diagnostics.

Keywords:
Multi-modal COVID-19 detection, cough sound analysis, deep learning

1. Introduction

The relentless global march of the COVID-19 pandemic has underscored the need for innovative and efficient diagnostic approaches [1]. The traditional paradigms of diagnosis, while foundational, face challenges such as delayed results and the potential for false negatives. In response to these imperatives, our research charts a groundbreaking course by integrating chest imaging and sound data, forging a holistic diagnostic model capable of transcending the limitations of individual modalities [2],[3].

The diagnostic landscape for COVID-19 is characterized by its intricacy, demanding solutions that blend swiftness with accuracy. Conventional diagnostic methods, including clinical assessments and laboratory tests, play pivotal roles but have revealed limitations in addressing the urgent need for enhanced diagnostic capabilities [4], [5].

Chest imaging, particularly through X-rays and computed tomography (CT) scans, has emerged as a crucial tool in unraveling the visual complexities of COVID-19. The nuanced radiological manifestations, including ground-glass opacities and consolidations, are precisely captured by these imaging modalities. Our research stems from a profound understanding of these visual cues, leveraging a meticulously curated dataset from chest images to construct a robust foundation for our diagnostic model [6]–[8].

The imaging dataset, meticulously sourced from chest X-rays and CT scans of confirmed COVID-19 cases, encapsulates critical abnormalities that define the disease. From bilateral involvement to lobular consolidations and ground-glass opacities, each image contributes to a rich tapestry of insights [9], [10]. We draw upon detailed clinical features published in a seminal Chinese paper, enriching our dataset and facilitating reliable identification for timely intervention [11].

Beyond the visual realm, the acoustic dimension offers a unique perspective in the diagnosis of COVID-19. Cough, a prevalent symptom, serves as an acoustic signature that extends beyond traditional diagnostic boundaries. Our research pioneers the integration of sound data sourced from the COUGHVID crowdsourcing dataset, comprising over 25,000 cough recordings. This dataset, characterized by demographic diversity and expert-labeled recordings, forms a significant contribution to training models for widespread COVID-19 screening [12], [13].

The sound dataset is an outcome of a meticulous data collection process facilitated through a web application at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Participants, guided by safety instructions, recorded cough audio along with metadata through a streamlined process. This dataset undergoes a transformative preprocessing journey, constructing Mel spectrogram representations through advanced mathematical transformations, unlocking the nuanced details essential for robust machine learning models [14], [15].

Our proposed diagnostic model represents a harmonious fusion of visual and auditory pathways, aiming to provide a comprehensive understanding of COVID-19. By concurrently analyzing chest images and

Manuscript received January 5, 2025
Manuscript revised January 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.1.26

respiratory sounds, our approach seeks to surmount the limitations of isolated modalities, offering a nuanced and holistic diagnostic capability [16].

The model architecture unfolds in layers, both visual and auditory. The visual pathway meticulously dissects chest images through a pre-trained VGG19 architecture, while the auditory pathway navigates Mel spectrogram representations. The ultimate convergence of insights transpires through a unified model, where distinctive features from both modalities intertwine, creating a synergistic representation for enhanced diagnostic accuracy.

In embarking on this innovative intersection of medical imaging and acoustics, our research aspires to significantly contribute to the collective efforts to combat the global health crisis posed by COVID-19. As the paper unfolds, it unravels the intricacies of our proposed methodology, positioning our work as a pioneering endeavor in the pursuit of more effective and comprehensive diagnostics.

2. Related Work

The unprecedented global impact of the COVID-19 pandemic has prompted a surge in interdisciplinary research, leveraging advanced technologies to enhance diagnostic capabilities. This literature review delves into key domains, providing a comprehensive overview of the existing knowledge landscape, paving the way for the methodology detailed in the subsequent sections [17], [18].

The diagnostic landscape for COVID-19 has evolved significantly since the early days of the pandemic. Initial diagnostic approaches primarily relied on clinical symptoms, polymerase chain reaction (PCR) tests, and serological assays. While effective, these methods faced challenges, including delays in results, false negatives, and the need for specialized laboratory settings. As a response, researchers globally have explored innovative approaches to expedite and enhance diagnostic processes [19].

The utilization of medical imaging, particularly chest X-ray and computed tomography (CT), has emerged as a valuable tool in COVID-19 diagnosis. Radiological manifestations of the disease, such as ground-glass opacities and consolidations, are detectable through these imaging modalities [20], [21]. Notably, early studies from China, as referenced in our methodology, have provided foundational insights into the radiological features of COVID-19, forming the basis for dataset creation and model training.

The integration of radiomics, a field focused on extracting quantitative features from medical images, has further enriched the diagnostic landscape. Radiomic analyses of chest imaging have demonstrated the potential to discern subtle patterns indicative of COVID-19, facilitating more nuanced and accurate diagnoses. These advances underscore the importance of interdisciplinary collaboration between radiologists and data scientists [22].

Beyond visual diagnostics, the auditory domain has emerged as a novel frontier in COVID-19 research. Cough, a common symptom of respiratory illnesses, has gained attention for its potential as a diagnostic marker. The COUGHVID dataset, referenced in our methodology, represents a pioneering effort in leveraging sound data for large-scale cough analysis. This dataset, collected during the early months of the pandemic, provides a diverse and valuable resource for training machine learning models.

Research into the acoustic signatures of respiratory diseases predates the COVID-19 era. Studies have explored the unique characteristics of coughs, wheezes, and other respiratory sounds to differentiate between various illnesses. The integration of these findings into a COVID-19 diagnostic framework reflects a nuanced understanding of the diverse manifestations of respiratory diseases [23], [24].

The application of machine learning, particularly deep learning, has revolutionized medical image analysis. Pre-trained architectures, such as VGG19 as employed in our methodology, leverage extensive datasets to extract hierarchical features from medical images. The transfer learning approach, freezing layers to retain generic visual knowledge, showcases the adaptability of deep learning models to specific medical domains [25][26].

In the realm of sound data, the application of advanced signal processing techniques has become pivotal. Mel spectrograms, as detailed in our methodology, provide a robust foundation for capturing nuanced audio details. Techniques such as *pitch-shifting* and *SpecAugment* contribute to a diverse dataset, addressing challenges such as class imbalance and enhancing the resilience of machine learning models [27].

The convergence of visual and auditory data in COVID-19 diagnosis represents a paradigm shift in the pursuit of holistic understanding. By integrating image and sound models, researchers aim to capture a broader spectrum of disease manifestations. The architecture detailed in our methodology harmoniously blends the strengths of both modalities, fostering a comprehensive approach to classification [28], [29].

While the strides in COVID-19 diagnostics are promising, challenges persist. The interpretability of machine learning models, ethical considerations in data collection, and the need for real-time implementation are among the forefront challenges. Future directions in research may explore explainable AI, continuous dataset updates, and collaboration frameworks to address these challenges and enhance the robustness of diagnostic methodologies.

In conclusion, this literature review provides a detailed exploration of the multifaceted landscape

surrounding COVID-19 diagnostics. The integration of visual and auditory modalities, coupled with advancements in machine learning, reflects the dynamism of contemporary research in the fight against the pandemic. The methodology presented in this paper aligns with and contributes to this evolving body of knowledge, offering a promising avenue for further advancements in COVID-19 diagnosis.

3. Methodology

In this section, we provide an overview of the datasets used for training and testing in our COVID-19 classification. The datasets encompass both image and sound data, each contributing unique insights into the characteristics and diagnostic potential of COVID-19.

3.1 Image Dataset

The image dataset utilized in this research originates from chest X-ray and computed tomography (CT) images of patients with confirmed COVID-19 diagnoses. These images, capturing the unique features of the disease, offer valuable information for diagnosis and assessment. The dataset includes abnormalities observed in chest CT scans, with bilateral involvement, multiple lobular and subsegmental areas of consolidation, and ground-glass opacity identified as distinctive patterns. Notably, the dataset's clinical features were detailed in a paper published by a Chinese team in late January, providing critical insights into the radiological manifestations of COVID-19. The objective is to facilitate reliable identification of infected patients, enabling timely detection and the implementation of necessary supportive care [7].

3.2 Sound Dataset

The sound dataset, known as the COUGHVID crowdsourcing dataset, focuses on cough audio signal classification to advance the study of large-scale cough analysis algorithms. With over 25,000 crowdsourced cough recordings collected between April 1st, 2020, and December 1st, 2020, the dataset provides a diverse representation of participant demographics, including age, gender, geographic location, and COVID-19 status. The dataset is a significant contribution to the research community, featuring an open-sourced cough detection algorithm and expert-labeled cough recordings by four experienced physicians. This comprehensive dataset plays a pivotal role in training machine learning models for widespread COVID-19 screening [13].

The data collection process involved a web application deployed on a private server at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Participants recorded cough audio using a straightforward "one recording, one click" workflow, capturing audio for up to 10 seconds. A brief questionnaire collected metadata about the participant, including age, gender, and current condition, while geolocation information was optional. Safety instructions for coughing during a global pandemic were provided to ensure participant well-being [13].

3.3 Data Preprocessing

Image Data Augmentation

The augmentation strategy for the image dataset involves a meticulous orchestration of transformative techniques. Operations such as shifts in width and height, shearing, zooming, rotation, horizontal flipping, and brightness adjustments are applied to enrich the dataset. This ensemble of augmentations enhances the model's adaptability, fortifying it against potential overfitting. The training dataset becomes a diverse canvas, offering a spectrum of visual variations to ensure effective generalization and pattern recognition. In contrast, the validation dataset undergoes a simpler rescaling, preserving a clear demarcation between training and validation sets [30]–[32].

Sound Data Transformation

In our endeavor to unveil the intricate details of respiratory sounds for effective COVID-19 classification, we embark on a meticulous preprocessing journey for sound data using advanced mathematical transformations. Leveraging the COUGHVID dataset, our primary focus is on constructing Mel spectrogram representations, a process that adeptly captures nuanced audio details, forming the foundation for robust machine learning models [33].

- *Decoding the Spectrogram Mathematics*

Spectrograms and Mel-spectrograms, fundamental elements of our sound data preprocessing, are crafted using the Fast Fourier Transform (FFT). This mathematical framework executes the Discrete Fourier Transform (DFT), translating time domain signals into the frequency domain. The resulting Spectrum sets the stage for the subsequent mel-scaling transformation. The Mel-scale, a logarithmic representation introduced by Stevens, Volkmann, and Newmann, imparts a musical flavor to frequencies, ensuring pitch distances sound perceptually similar. The Mel-spectrogram represents a harmonious conversion of frequencies into the mel-scale [33], [34]. The Mel-scale conversion is mathematically expressed as:

$$Mel = 2595 \times log_{10}\left(1 + \frac{Hertz}{700}\right) \quad (1)$$

- *Harmony in Data Augmentation*

Our data augmentation strategy encompasses both the original audio signal and the computed Mel-spectrograms. Acknowledging the necessity of a fixed input size in deep learning, we harmonize the audio samples to a standard length of 156,027 (7.07 seconds). This resizing ensures uniformity, whether trimming lengthy samples or padding shorter ones with zeros [34].

- *Pitch-Shifting: A Crescendo for Audio Samples*

In the realm of audio data augmentation, we introduce a crescendo through Pitch-Shifting. This method, analogous to adjusting the pitch of a musical note, unfolds elegantly using the Librosa library. Specifically tailored for the Likely-COVID-19 class to address class imbalance, our implementation shifts audio samples down by four steps, with each step representing a semitone. This augmentation technique adds a unique dimension to our dataset while safeguarding the integrity of vocal features [15], [35].

- *Spectral Data Augmentation*

The symphony continues with *SpecAugment*, a technique fine-tuned for spectral data augmentation. Drawing inspiration from the DiCOVA challenge, where *SpecAugment* significantly improved accuracy, we apply a three-step augmentation method on mel-spectrograms. Time Warping, Frequency Masking, and Time Masking work in unison to create new mel-spectrograms, addressing class imbalance with meticulous precision. This symphonic augmentation approach solidifies the foundation for training our machine learning model [28], [33].

In this harmonious blend of mathematical transformations and meticulous augmentation, we craft a resilient and highly effective preprocessing pipeline for sound data. These intricacies set the stage for a symphony of classification, where the nuances of respiratory sounds contribute to the melody of COVID-19 detection.

4. Model Architecture

Our model architecture is a harmonious blend of visual and auditory insights, designed to comprehensively capture the nuanced characteristics of COVID-19. By integrating image and sound data, our approach seeks to provide a holistic understanding of the disease, leveraging both modalities for effective classification

4.1 Image-Based Model

The image-based model serves as a meticulous observer, scrutinizing the intricate visual manifestations of COVID-19 within medical images. Embarking on a transformative journey, the model meticulously dissects each image, seeking to extract salient features and patterns that signify the presence of the disease. As shown in figure 1, the proposed image-based model composites of the following layers:

Input Layer: At the forefront, a specialized input layer awaits images of dimensions (224, 224, 3), setting the stage for an in-depth exploration. These images, sourced from chest X-ray and CT scans, encapsulate the radiological nuances critical for COVID-19 diagnosis.

Base Model-VGG19: In our pursuit of robust feature extraction, we employ the VGG19 architecture, pre-trained on the expansive ImageNet dataset. This strategic choice leverages the wealth of visual knowledge acquired by VGG19, providing our model with a comprehensive understanding of generic visual patterns.

Freezing Learned Features: To preserve the wealth of knowledge embedded in the pre-trained VGG19, we judiciously freeze its layers. This decision ensures that the model retains the ability to recognize fundamental visual features while adapting its subsequent layers to the specific intricacies of COVID-19-related abnormalities.

Abstraction through Additional Layers: Building upon the VGG19 foundation, we introduce additional layers designed to abstract visual patterns effectively. These layers act as specialized filters, discerning hierarchical features that contribute to the unique radiological signature's indicative of COVID-19.

Enhancing Adaptability: Global average pooling is strategically employed to distill the extracted features into a more compact representation, promoting adaptability and preventing the model from fixating on irrelevant details. This step serves as a critical dimensionality reduction, paving the way for efficient processing in subsequent layers.

Adopting Dropout Mechanisms: To fortify our model against the risk of overfitting, dropout mechanisms are implemented judiciously. These mechanisms randomly deactivate a fraction of neurons during training, encouraging the model to develop robust and generalized representations, essential for accurate COVID-19 classification. unavoidable.

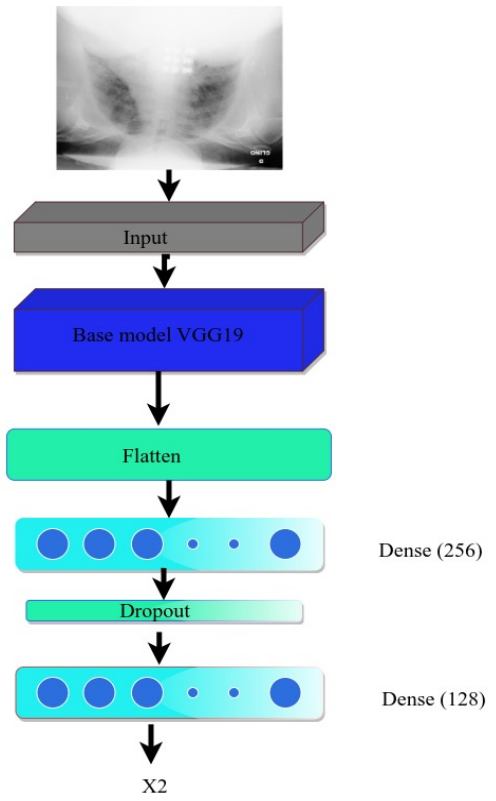

Fig. 1 Architecture of the image-based model.

4.2 Sound-Based Model

In the realm of audio analysis, the proposed sound-based model as shown in figure 2 embarks on a nuanced exploration of COVID-19's acoustic manifestations. The model processes sound data encapsulated as Mel spectrogram images of dimensions (156,27,3), delving into the intricate details of respiratory sounds. This auditory pathway employs a sophisticated combination of convolutional layers, spatial dropout mechanisms, Long-Short Term Memory (LSTM) layers, and an attention mechanism to discern subtle audio patterns indicative of the disease.

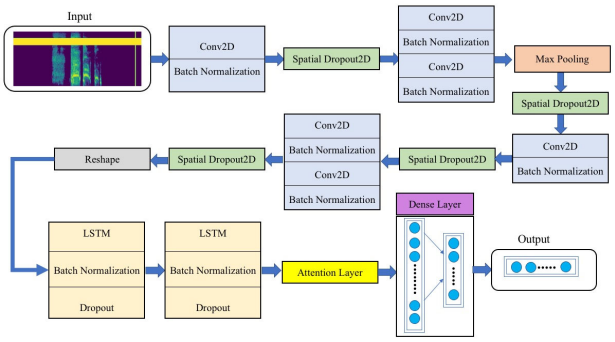

Fig. 2 A visual representation of the sequential flow of layers in the proposed sound-base model

Convolutional Layers: The journey unfolds with four meticulously crafted convolutional layers, each with an escalating number of filters (16, 32, 64, 128). The kernel size of (2×2) acts as a receptive window, capturing local acoustic features that contribute to the overall audio profile. Rectified Linear Unit (ReLU) activation functions, Batch Normalization, and Spatial Dropout are seamlessly interwoven to enhance non-linearity, normalize activations, and prevent overfitting.

Max Pooling: Strategic Max Pooling layers, interspersed with convolutional blocks, serve as windows that reduce the complexity of the network. These layers link feature maps to fixed-size windows, enabling efficient processing and dimensionality reduction.

Reshaping and LSTM Layers: Following convolution, the feature maps undergo a reshaping process to streamline their representation. The reshaped data is then fed into two LSTM layers (512 and 256 units), unraveling temporal dependencies within the audio sequences. The LSTM architecture, with its input, forget, and output gates, adeptly captures sequential information and retains contextual understanding across varying time frames.

Attention Mechanism: At the heart of our sound model lies an attention mechanism, a pivotal component for focusing on salient audio features. This mechanism operates in three phases: Scores Alignment, Weights, and Context Vector. Scores Alignment, expressed as a hyperbolic tangent (tanh) function, aligns hidden states with trainable weights. Subsequent application of the *softmax* function computes attention weights, indicating the relevance of each hidden state. The final stage culminates in the computation of a context vector, a weighted sum of hidden states. This attention vector encapsulates the essence of critical audio features and propels them forward for further analysis.

Fully Connected Neural Network: The attention vector converges into a fully connected neural network, comprising a layer with 100 units activated by the Rectified Linear Unit (ReLU). This transformative step further refines the representation, infusing a rich hierarchical understanding of audio patterns. To prevent overfitting, a Dropout layer with a rate of 0.5 is thoughtfully introduced.

4.3 Merging Modality for Holistic Classification

The ultimate culmination of our research journey resides in the final model, where visual and auditory pathways converge, creating a harmonious symphony of insights for holistic COVID-19 classification. This model seamlessly blends the unique perspectives garnered from image and sound data, allowing the fusion of distinctive features crucial for an all-encompassing understanding of the disease.

Unified Pathways: At the heart of this architectural masterpiece are two distinct pathways—visual and auditory. The visual pathway begins with a dedicated input layer for images (224, 224, 3) and traverses through a VGG19 architecture pretrained on ImageNet. The auditory pathway processes sound data in the form of Mel spectrogram images (156,027, 3), unraveling the intricate acoustic signatures using convolutional layers, spatial dropout, LSTM layers, and an attention mechanism. These pathways, each a specialist in its domain, extract valuable insights from diverse modalities.

Concatenation of Modalities: The culmination of insights occurs at the concatenation layer, where the extracted features from the visual (x_1) and auditory (x_2) pathways intertwine. This symbolic handshake between visual and auditory cues creates a unified representation, setting the stage for synergy in the classification journey.

Dense Fusion Layers: Additional dense layers act as maestros orchestrating the fusion of modalities. These layers, akin to a skilled conductor leading a symphony, refine the interplay of visual and auditory features. The introduction of dense layers enhances the model's capacity to comprehend the nuances presented by both modalities.

Sigmoid Activation for Classification: The grand crescendo of the model's architecture occurs in the final layer, where a sigmoid activation function serves as the ultimate arbiter of COVID-19 classification. This binary activation encapsulates the model's decision-making process, distinguishing between infected and non-infected cases.

Holistic Synergy: Figure 3 visually encapsulates the intricate dance of layers within our final model. The synergy achieved through the fusion of visual and auditory modalities defines the model's ability to harmoniously interpret COVID-19 manifestations. This holistic approach not only boosts the model's accuracy but also ensures a comprehensive understanding of the diverse manifestations presented in our multi-modal dataset

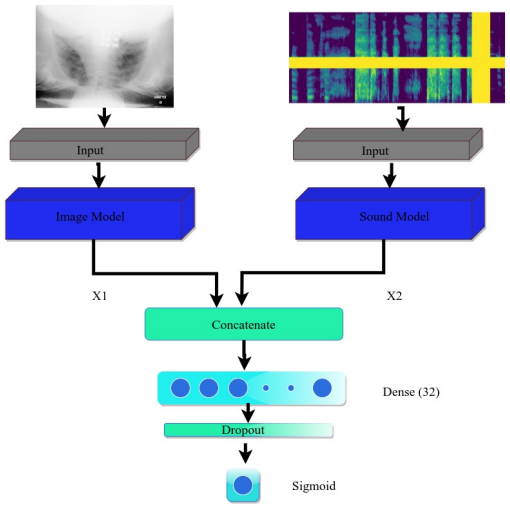

Fig. 3 Graphical representation of the overall model architecture.

5. Training Parameters and Experimental Setup

The ensemble model underwent a meticulously planned training process, guided by thoughtfully chosen parameters and strategic monitoring techniques. These configurations played a pivotal role in shaping the model's convergence and enhancing its ability to generalize.

- Epochs: 200
- Batch Size: 16
- Learning Rate: 0.001
- Optimizer: Adamax

The choice of 200 epochs aimed to provide the model with ample iterations to discern intricate patterns within the dataset.

A moderate batch size of 16 struck a balance between computational efficiency and gradient accuracy during optimization. The learning rate of 0.001, coupled with the Adamax optimizer, facilitated adaptive optimization, allowing the model to navigate the complex parameter space effectively.

Model Checkpoint and training Time: A crucial aspect of the training process was the implementation of the *ModelCheckpoint* callback. This monitoring strategy continuously observed the validation accuracy and saved the model's weights whenever an improvement was detected. The saved weights served as a snapshot of the model at its optimal state, mitigating the risk of overfitting.

The entire training process took approximately 4295.55 seconds, indicating the computational demand of the model. This duration encompassed the execution of all epochs, including forward and backward passes, weight updates, and validation assessments.

Learning Curve Visualization: To gain insights into the model's learning dynamics, Figure 4 presents the learning curves. Figure 4 illustrates the evolution of training and validation accuracy across epochs, offering a visual narrative of the model's ability to learn from the training data while gauging its performance on unseen validation data. Figure 5 provides insights into the corresponding loss dynamics, offering a comprehensive view of the model's convergence.

The learning curves serve as valuable tools for understanding the model's behavior throughout training. The close alignment of training and validation curves suggests effective learning without pronounced overfitting.

The judicious selection of training parameters, coupled with robust monitoring mechanisms, contributes to the reliability and generalization of the ensemble model. The learning curves, presented in Figure 4, provide a visual narrative of the model's training journey, offering transparency into its strengths and potential areas for refinement.

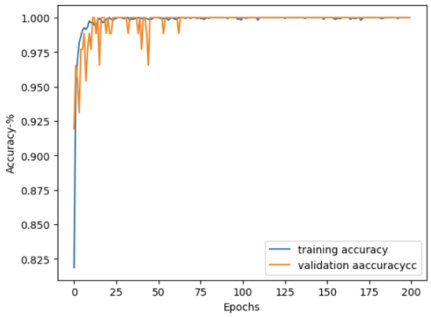

Fig. 4 Learning curve for accuracy rate.

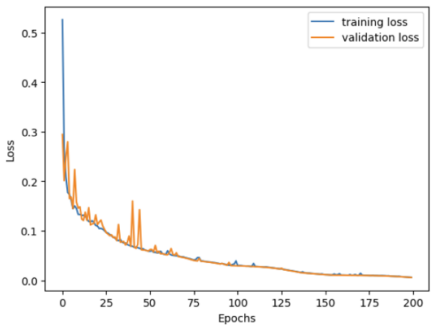

Fig. 5 Learning curve for loss rate.

6. Results and Discussion

The ensemble model underwent a thorough evaluation, with a focus on various metrics to provide a comprehensive view of its COVID-19 classification performance. Training and testing metrics are shown in Table 1.

Table 1: Training and Testing results

	Metric	**Value**
Training	Loss	0.122
	Accurate	99.77%
Testing	Loss	0.126
	Accurate	99.54%

Discrepancy Analysis

The small discrepancy of 3.50% between training and testing accuracy emphasizes the model's robustness and generalization ability.

Class-wise Accuracy

The ensemble model demonstrated exceptional accuracy in classifying both classes. The negative and positive classes reported 99.68% and 99.14% respectively. Class-wise accuracy highlights the model's proficiency in discriminating between COVID-19 and non-COVID-19 instances, showcasing high accuracy for both classes.

Confusion Matrix

The confusion matrix in figure 6 represents the model's performance in distinguishing between classes.

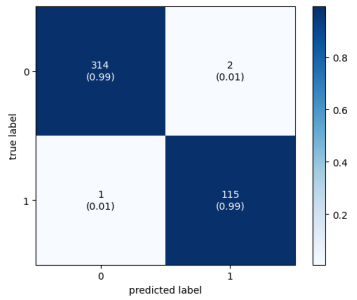

Fig. 6 Confusion Matrix.

In Table 2, detailed results are reported.

Table 2: Training and Testing results

Metric	Value
Sensitivity	9968
Specificity	9829
Precision	9937
Accuracy	9931
F1-score	9952

The ensemble model exhibits outstanding performance in COVID-19 classification, with minimal training-testing discrepancies

.

6. Conclusion

In conclusion, our groundbreaking research endeavors to revolutionize COVID-19 detection by amalgamating medical imaging and sound analysis. The strategic integration of diverse modalities, including chest X-ray, CT images, and cough recordings, imparts a nuanced and comprehensive perspective on the disease. Our preliminary results showcase a promising level of accuracy, underlining the potential efficacy of our model in identifying intricate COVID-19-related abnormalities.

The incorporation of a VGG19-based image model and a sophisticated sound model utilizing Mel spectrograms, convolutional layers, LSTM layers, and an attention mechanism forms the backbone of our approach. The meticulous orchestration of data augmentation techniques, including image transformations and spectral data augmentation, contributes to the robustness of our machine learning model.

During the training phase, our model learns to discern radiological patterns indicative of COVID-19 in medical images and subtle audio features present in cough recordings. The harmonious fusion of these modalities in the final model, as depicted in Figure 3, showcases a synergistic interplay that enhances our model's ability to interpret the diverse manifestations of the disease.

While our research has shown promising results, ongoing efforts are dedicated to refining and validating the model on larger and more diverse datasets. The ultimate goal is to ensure the reliability and generalizability of our approach across varied demographic and clinical scenarios. The proposed work holds significant promise for advancing the field of COVID-19 diagnostics, with potential implications for improving public health outcomes through timely and accurate identification of infected cases.

References

[1] M. Isgut, L. Gloster, K. Choi, J. Venugopalan, and M. D. Wang, “Systematic Review of Advanced AI Methods for Improving Healthcare Data Quality in Post COVID-19 Era,” IEEE Reviews in Biomedical Engineering, vol. 16. Institute of Electrical and Electronics Engineers Inc., pp. 53–69, 2023. doi: 10.1109/RBME.2022.3216531.

[2] Zhaoshan Liu and Lei Shen, “CECT: Controllable ensemble CNN and transformer for COVID-19 image classification,” Computers in Biology and Medicine, Volume 173, 2024, doi:10.1016/j.compbiomed.2024.108388.

[3] Li, Y., Shi, J., Xia, J., Duan, J., Chen, L., Yu, X., Lan, W., Ma, Q., Wu, X., Yuan, Y. and Gong, L., “Asymptomatic and Symptomatic Patients with Non-severe Coronavirus Disease (COVID-19) Have Similar Clinical Features and Virological Courses: A Retrospective Single Center Study,” Frontiers in Microbiology, vol. 11, Jun. 2020, doi: 10.3389/fmicb.2020.01570.

[4] K. Ren, G. Hong, X. Chen, and Z. Wang, “A COVID-19 medical image classification algorithm based on Transformer,” Sci Rep, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-32462-2.

[5] Toda, Ryo, Hayato Itoh, Masahiro Oda, Yuichiro Hayashi, Yoshito Otake, Masahiro Hashimoto, Toshiaki Akashi, Shigeki Aoki, and Kensaku Mori. "Identifying Suspicious Regions of Covid-19 by Abnormality-Sensitive Activation Mapping." *arXiv preprint arXiv:2303.14901*, 2023.

[6] Song, Ying, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen. "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images." *IEEE/ACM transactions on computational biology and bioinformatics* 18, no. 6: 2775-2780, (2021).

[7] Y. Jiang, H. Chen, H. Ko, and D. K. Han, “Few-shot learning for CT scan based covid-19 diagnosis,” in International Conference on Acoustics, Speech and Signal Processing – Proceedings (ICASSP), Institute of Electrical and Electronics Engineers (IEEE) Inc., pp. 1045–1049, 2021 doi:10.1109/ICASSP39728.2021.9413443Erw

[8] Silva, Leticia, Carlos Valadão, Lucas Lampier, Denis Delisle-Rodríguez, Eliete Caldeira, Teodiano Bastos-Filho, and Sridhar Krishnan. "COVID-19 respiratory sound analysis and classification using audio textures." *Frontiers in Signal Processing,* vol(2): 986293, 2022, doi: 10.3389/frsip.2022.986293.

[9] Ghosh, Susmita, and Abhiroop Chatterjee. "Automated COVID-19 CT Image Classification using Multi-head Channel Attention in Deep CNN." *arXiv preprint arXiv:2308.00715*, 2023.

[10] U. Bhattacharjya, K. K. Sarma, J. P. Medhi, B. K. Choudhury, and G. Barman, “Automated diagnosis of COVID-19 using radiological modalities and Artificial Intelligence functionalities: A retrospective study based on chest HRCT database,” Biomedical Signal Processing and Control, vol. 80, Feb. 2023, doi: 10.1016/j.bspc.2022.104297.

[11] Antonios Makris, Ioannis Kontopoulos, and Konstantinos Tserpes. 2020. COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks. In 11th Hellenic Conference on Artificial Intelligence (SETN 2020). Association for Computing Machinery, New York, NY, USA, 60–66. doi:10.1145/3411408.3411416

[12] Sharma, Neeraj, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Prasanta Kumar Ghosh, and Sriram Ganapathy. "Coswara--a database of breathing, cough, and voice sounds for COVID-19 diagnosis." arXiv preprint arXiv:2005.10548, 2020.

[13] Orlandic, L., Teijeiro, T. & Atienza, D. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. Scientific Data 8, no. 1, 156, 2021. doi:10.1038/s41597-021-00937-4

[14] N. Yella and B. Rajan, "Data Augmentation using GAN for Sound based COVID 19 Diagnosis," In proceedings of the 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2021, Cracow, Poland, pp. 606-609, doi: 10.1109/IDAACS53288.2021.9660990.

[15] S. Hamdi, M. Oussalah, A. Moussaoui, and M. Saidi, "Attention-based hybrid CNN-LSTM and spectral data augmentation for COVID-19 diagnosis from cough sound," Journal of Intelligent Information System, 2022, doi: 10.1007/s10844-022-00707-7.

[16] A. O. Popadina, A. M. Salah, and K. Jalal, "Voice Analysis Framework for Asthma-COVID-19 Early Diagnosis and Prediction: AI-based Mobile Cloud Computing Application," in Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2021, IEEE Inc., pp. 1803–1807. doi: 10.1109/ElConRus51938.2021.9396367.

[17] Ai, Tao, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases." Radiology 296, no. 2 (2020): E32-E40.

[18] Wu, Yu-Huan, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Rong-Guo Zhang, and Ming-Ming Cheng. "JCS: An explainable covid-19 diagnosis system by joint classification and segmentation." IEEE Transactions on Image Processing 30 (2021): 3113-3126. doi: 10.1109/TIP.2021.3058783.

[19] Rehman, Arshia, Saeeda Naz, Ahmed Khan, Ahmad Zaib, and Imran Razzak. "Improving coronavirus (COVID-19) diagnosis using deep transfer learning." In Proceedings of International Conference on Information Technology and Applications: ICITA 2021, pp. 23-37. Springer Nature Singapore, 2022. Doi: 10.1007/978-981-16-7618-5_3

[20] A. Y. A. Saeed and A. E. Ba Alawi, "Covid-19 Diagnosis Model Using Deep Learning with Focal Loss Technique," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), IEEE, 2021, pp. 1-4, doi: 10.1109/ICOTEN52080.2021.9493477.

[21] Y. Chang, Z. Ren, and B. W. Schuller, "Transformer-based CNNs: Mining Temporal Context Information for Multi-sound COVID-19 Diagnosis," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, IEEE Inc., 2021, pp. 2335–2338. doi: 10.1109/EMBC46164.2021.9629552.

[22] Amin Khodaei, Parvaneh Shams, Hadi Sharifi, Behzad Mozaffari-Tazehkand, "Identification and classification of coronavirus genomic signals based on linear predictive coding and machine learning methods," Biomedical Signal Processing and Control, 80(1), 2023, doi: 10.1016/j.bspc.2022.104192. r.

[23] Deng, Cheng, Jiaxin Ding, Luoyi Fu, Weinan Zhang, Xinbing Wang, and Chenghu Zhou. "Covidia: COVID-19 Interdisciplinary Academic Knowledge Graph." arXiv preprint arXiv:2304.07242 (2023).

[24] Mejía-Mejía, Elisa, and Panicos A. Kyriacou. "Effects of noise and filtering strategies on the extraction of pulse rate variability from photoplethysmograms." Biomedical Signal Processing and Control 80 (2023), doi: 10.1016/j.bspc.2022.104291.

[25] Aytekn, Idil, Onat Dalmaz, Kaan Gonc, Haydar Ankishan, Emine U. Saritas, Ulas Bagci, Haydar Celik, and Tolga Çukur. "COVID-19 Detection from Respiratory Sounds with Hierarchical Spectrogram Transformers," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1273-1284, March 2024, doi: 10.1109/JBHI.2023.3339700. s

[26] Lin Li, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, Yi Lu, Zhenghan Fang, Qi Song, Kunlin Cao, Daliang Liu, Guisheng Wang, Qizhong Xu, Xisheng Fang, Shiqin Zhang, Juan Xia, and Jun Xia. " Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT." Radiology 296, no. 2 (2020): E65-E71. doi: 10.1148/radiol.2020200905.

[27] Yida Mu, Ye Jiang, Freddy Heppell, Iknoor Singh, Carolina Scarton, Kalina Bontcheva, Xingyi Song. "A large-scale comparative study of accurate COVID-19 information versus misinformation," In: TrueHealth 2023: Workshop on Combating Health Misinformation for Social Wellbeing (2023). doi: 10.36190/2023.45

[28] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 cough classification using machine learning and global smartphone recordings," Computers in Biology and Medicine, vol. 135, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104572.

[29] Chih-Chung Hsu, Chih-Yu Jian, Chia-Ming Lee, Chi-Han Tsai, and Sheng-Chieh Dai. Strong baseline and bag of tricks for covid-19 detection of CT scans," arXiv preprint, Mar 2023, arXiv:2303.08490.

[30] Ghazal Bargshady, Xujuan Zhou, Prabal Datta Barua, Raj Gururajan, Yuefeng Li, U. Rajendra Acharya, "Application of CycleGAN and transfer learning techniques for automated detection of COVID-19 using X-ray images," Pattern Recognition Letters, vol. 153, pp. 67–74, 2022, doi: 10.1016/j.patrec.2021.11.020

[31] Sohaib Asif, Yi Wenhui, Hou Jin and Si Jinhai, "Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network,", 2020, IEEE 6th International Conference on Computer and Communications (ICCC), China, pp. 426-433, doi: 10.1109/ICCC51575.2020.9344870.

[32] K. Gupta and V. Bajaj, "Deep learning models-based CT-scan image classification for automated screening of COVID-19," Biomedical Signal Processing and Control, vol. 80, Feb. 2023, doi: 10.1016/j.bspc.2022.104268.

[33] Park, Daniel S., William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, and Quoc V. Le. "SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition." Interspeech 2019 (April), doi:10.21437/Interspeech.2019-2680.

[34] M. Unser and T. Blu, "Wavelet theory demystified," IEEE Transactions on Signal Processing, vol. 51, issue 2, pp. 470-483, 2003, doi: 10.1109/TSP.2002.807000.

[35] P. Bagad, A. Dalmia, J. Doshi, A. Nagrani, P. Bhamare, A. Mahale, et al., "Cough against COVID: Evidence of COVID-19 signature in cough sounds", arXiv:2009.08790, 2020.

Mai Rashid is an MSc. student at the Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt. She can be contacted at email: mai.rasheed@fci.bu.edu.eg.

Hamada Nayel is an Assistant Professor at the Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt. In 2019, he received his Ph.D. from Mangalore University, India. His research interests include Arabic NLP, and Data Science.
He can be contacted at email: hamada.ali@fci.bu.edu.eg

Ahmed Taha is an Associate Professor at the Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt. In 2019, he received his Ph.D. from Ain Shams University, Egypt. His research interests include Digital Forensics, Image forgery detection, and social media analysis.
He can be contacted at email: ahmed.taha@fci.bu.edu.eg.