COVID-19 Detection from Cough Audio

Dora Eskridge
School of Data Science
University of Virginia
Charlottesville, USA
de2br@virginia.edu

Alexander Lilly
School of Data Science
University of Virginia
Charlottesville, USA
kzr3fb@virginia.edu

Matt Scheffel
School of Data Science
University of Virginia
Charlottesville, USA
mcs9ff@virginia.edu

***Abstract*—Through this analysis, spectrograms representing cough audio data from a wide and diverse population are analyzed in deep learning models to determine which models successfully classify individuals as healthy patients or as COVID patients based on the sound of their cough. Five pretrained convolutional neural networks of varying sizes were applied via transfer learning to images of spectrogram data. Hyper-parameters were optimized, the impact of data augmentation was assessed, and the benefits of a multi-modal approach were investigated. The authors found that data augmentation was not necessary to achieve high performance on validation data, and the incorporation of patient data in the model improved classifier performance immensely. The highest performing classifier using spectrogram data alone possessed an AUC of 0.74, whereas the highest performing classifier using both the spectrogram data and patient symptom data possessed an AUC of 0.91.**

***Index Terms*—COVID-19, audio, classification, deep learning, convolutional neural network.**

I. Motivation

Over the past several years, the novel coronavirus (COVID-19) has affected nearly every aspect of daily life all over the world. Working from home, wearing masks, and maintaining a safe distance from others became normalized behaviors as people grappled with the pervasiveness of the highly-contagious disease. COVID-19 is notorious for the ease with which it can be passed on from person to person, particularly when the infected person is symptomatic. The most obvious and transmissive symptom is a dry cough– a person infected with COVID-19 will often suffer from this symptom, and through their coughing they are able to spread the disease to a much higher magnitude than an asymptomatic sufferer of COVID-19. For years now, a cough has been a surefire sign that someone may have COVID-19, and simply being in the presence of someone with a cough has been enough to inspire anxiety and fear in a healthy person. However, not every person with a cough suffers from COVID-19.

A cough is a fairly common symptom of many prevalent illnesses. Some coughs are fairly harmless, due to asthma, allergies, or throat irritation unrelated to contagious disease. Other coughs signify a much more serious ailment, like pneumonia, tuberculosis, or whooping cough. Most infamously, COVID-19 is often accompanied by its trademark symptom, a dry cough. The ability to distinguish between diseases based on symptoms is something that doctors and physicians use constantly in their field of work. Often, the presence or lack thereof of a given symptom can help eliminate several diseases from the list of possible ailments for a given patient, and the list of symptoms is a key tool in pinpointing what afflicts a given patient and what treatment would best serve them. A symptom like a cough, however, is not extremely helpful in terms of determining type of illness. The ability to quickly distinguish between the illnesses that are accompanied by a cough would be extremely helpful in the medical world.

In application, a successful model would simplify the COVID testing process. Rather than queuing for PCR tests or trusting the results of unreliable rapid tests, people with a cough that may have contracted COVID-19 could simply process an audio recording of their symptom and get real time results. This model would also help doctors, both in hospitals and in the field, by allowing them to quickly diagnose patients and decide the appropriate treatment plan. Since COVID-19 first arose in the Wuhan province of China, the disease has been well-known for its infectiousness, and often the delay between testing for the virus and receiving results and guidance to isolate leads to even more healthy people being exposed to the virus. With a successful model for audio-diagnosis, this time window could be trimmed down, and highly-contagious patients would be less likely to infect other people.

Through our data analysis and modeling, we hope to construct a model that can identify whether or not a patient is infected with COVID-19 based on the sound of their cough. A successful model in this vein will have many implications. Rather than the time window of multiple days that is often required to get results from the reliable polymerase chain reaction (PCR) tests, or even the half hour needed to conduct a rapid test from a pharmacy, this model could analyze the sound of the patient's principal symptom and quickly diagnose their illness without any invasive nose-swabbing or saliva testing.

II. Literature Survey

When the world was hit by the COVID-19 pandemic in early 2020, the healthcare sector desperately needed rapid diagnostic methods to manage and curtail the spread of the virus. Traditional diagnostic methods, essential as they are, often come with the constraints of high cost, time consumption, and limited scope. This dilemma is highlighted by data from July 2020, which indicated a daily diagnostic range of 520,000 to 823,000 tests in the U.S., while the actual need

was assessed at five million tests. [1] A quick diagnosis of COVID-19 is essential for both isolating infected individuals to prevent further spread and initiating early treatment.

In this context, deep learning, with a specific emphasis on Convolutional Neural Networks (CNN's), has become prominent in processing complex data types. Particularly, the realm of audio signals has witnessed a transformative shift. Historically, the analysis of such signals was anchored in Acoustic Event Detection (AED) that largely relied on handcrafted features and traditional machine learning classifiers. [5] The current wave of research underscores the impressive capability of CNN's to decipher the intricate temporal structures and patterns inherent in audio data. This is evident in the recent surge of interest in harnessing these networks to analyze cough sounds due to their status as a pervasive symptom of COVID-19. Such an approach not only offers a novel diagnostic avenue but also promotes a non-invasive alternative to existing diagnostic tools, offering hope for a more immediate and accessible means of diagnosis.

This literature review endeavors to critically examine and encapsulate the burgeoning research on the feasibility and efficacy of CNN's in diagnosing COVID-19 through cough sounds. Outside of COVID-19, this model could also someday be used to help diagnose a variety of other afflictions which present in the form of a cough, such as asthma, COPD, pneumonia, tuberculosis, pertussis, bronchitis, and other medical illnesses.

The evolving field of audio classification has seen significant advancements and innovations, notably in the application of deep learning architectures. Distinct architectures, including but not limited to AlexNet, VGG, Inception, and ResNet, have been rigorously tested for their efficacy in audio tasks. [5] While many researchers have developed custom models tailored to audio processing, a notable trend is the successful adaptation of models initially pre-trained on image datasets, like ImageNet. These models, when fine-tuned, have shown remarkable efficiency in audio classification tasks, exemplifying the flexibility of deep learning models and the efficacy of transfer learning. [1]

Despite the advancements, the field isn't without challenges. A prevalent issue is the weak labeling in sizable datasets, where many segments may lack definitive cues, making it challenging for models to generate accurate representations. [5] Additionally, the inherent intricacies of audio signals, where a single clip might encapsulate multiple labels of varied specificities, pose significant challenges. Another one of the paramount challenges in this domain is the lack of availability of diverse and extensive cough sound datasets, with the "small size of available and open-source datasets" cited as the "main challenge of the cough or breath-based COVID-19 diagnosis research direction." [2] A richer and broader dataset is a cornerstone for ensuring the model's generalizability across various populations and settings.

In response to these and other challenges, researchers have embraced innovative techniques, looking beyond traditional audio signals. One approach which offered "best feature representation for [the] particular problem" has been the utilization of log-mel spectrograms, converting multifaceted audio signals into more tangible image-like representations for processing. [6] The practice of transfer learning, especially leveraging weights from Keras ImageNet-pretrained models, has been shown to improve the performance of models as well. [1] Visualization techniques have provided insights into the models' processing, showcasing their focus on essential aspects of spectrograms, reminiscent of edge detection in image processing. [6]

In relation to the relevant task at hand — diagnosing COVID-19 from cough sounds — the researched methodologies largely hinge on the aforementioned practice of transforming audio data into spectrograms for CNN analysis, but take a variety of unique approaches:

- Research using various CNN architectures to classify the soundtracks of a dataset of 70M training videos (5.24 million hours) employed transfer learning on pre-trained models like VGG19 and ResNet50V2, converting voice signals into mel-spectrogram images and attaining an impressive 88.04% accuracy. While the incredibly large dataset in this model offers greater diversity, leading to potentially better performance than smaller sets, the practical constraints of training time and assumptions of possible overfitting in the smaller datasets (70K and 23K) suggest the need for regularization techniques to enhance performance. For example, the paper references a potential time frame of 23 weeks to run a single epoch. [5]
- Many models tend to complicate their design through the use of multiple models that take different inputs and aggregate the outputs to make predictions, leveraging features such as MFCC and Mel-Spectrogram. However, research showed that high quality performance could be achieved with simple mel-spectrograms. [6]
- Another research paper introduced a comprehensive five-module model encompassing sound extraction, sound feature extraction, cough detection, cough classification, and finally, COVID-19 diagnosis. A combination of CNN's, SVM's, KNN's, and RNN's was employed, and the model managed a relatively high accuracy of 81.25% on "completely unfamiliar cough samples." [3]
- In a streamlined approach, researchers introduced a two-fold approach with a low-complexity CNN model specifically for cough detection, coupled with another CNN model for diagnosing respiratory ailments. This methodology achieved over 89% accuracy in both identifying cough events and distinguishing between respiratory conditions while remaining "computationally efficient." Although successful, the authors noted that the performance is limited by the unbalanced nature of the training database and potential biases, such as environmental consistencies, necessitating the use of

TABLE I
DISTRIBUTION OF CLASS LABELS

Status	Count	Percentage
Healthy	12,479	45.3%
Symptomatic	2,590	9.4%
COVID-19	1,155	4.2%
Unlabeled	11,327	41.1%

TABLE II
CANDIDATE PRE-TRAINED MODELS FOR EXPERIMENTATION

Model	Size (MB)	Top-5 Accuracy	Parameters	Depth
MobileNetV2	14	90.10%	3.5M	105
EfficientNetB2	36	94.90%	9.2M	186
InceptionV3	92	93.70%	23.9M	189
EfficientNetB7	256	97.00%	66.7M	438
VGG19	549	90.00%	143.7M	19

multiple performance metrics and a larger, more varied dataset in order to ultimately achieve improved accuracy. [4]

Deep learning for audio classification is undoubtedly promising, yet it's not devoid of critiques. Emphasizing the concerns raised in the research summary, there's a general consensus on the need for expansive and diverse data collection. Enhanced data collection would not only refine the existing models but also potentially unveil novel patterns, leading to heightened model efficacy. Specifically, while the utility of vast datasets (such as the YouTube-100M example) is recognized, there's a consensus about the limitations imposed by the unavailability of diverse and comprehensive datasets specific to coughing audio. The effectiveness of models across different datasets also remains a contested matter. Another important issue to consider moving forward is benchmarking and comparative analysis. For robustness, it would be suggested that upcoming research efforts compare and contrast the performance of proposed models with existing ones, helping to establish a standard in the process.

The real-world impact of these models hinges on their seamless integration into prevailing healthcare systems or applications. The future should see more collaborations between technologists and healthcare professionals to achieve this. As deep learning and audio recognition technology continues to evolve, there is a promising future field that goes beyond rudimentary binary classifications. Our research sources hint at the potential of models that can diagnose a spectrum of respiratory conditions from cough sounds, offering a more holistic and useful diagnostic tool.

III. METHOD

To construct our model, we will be using the COVID-19 Cough Audio Classification dataset from Kaggle. The cough audio dataset, also known as CoughVid, contains over 25,000 audio recordings of different people coughing. The population recorded spans across different ages, genders, locations, and illnesses. Of these tens of thousands of cough recordings, 2,800 were reviewed by a group of doctors and given a diagnosis. These professionally diagnosed and labeled cough recordings can be used to build out a successful audio classification model. Table 1 shows the distribution of class membership.

There are many different ways to analyze sound data. The majority of them are based in three key features of a given piece of sound data: time period, amplitude, and frequency. In the world of data analysis, observations of sound data are often converted into visual representations prior to data processing and analysis. In the wake of this transformation, researchers are able to pull specific information from each audio observation with ease, and they also gain the ability to compare observations based on several key features.

To prepare the data for the model-building process, we will convert the audio recordings into spectrograms, which serve as visual representations of each audio observation. By using spectrogram data, we can build a model that processes image data and classifies each unique cough according to its respective spectrogram. The proposed work aims to accomplish three things:

1) to identify the smallest model which is effective at detecting COVID-19 from cough audio.
2) to identify whether the incorporation of patient symptom information will improve model performance.
3) to identify data augmentation techniques that will improve model performance.

The target application of this model is in edge devices that could be used in the medical field, most likely during rapid and high throughput screening of patients for COVID-19. Five pre-trained models available with the Keras package will be fine tuned to this application. These models will vary in size. See Table 1 for the list of models to be considered and relevant model properties, according to Keras.

IV. PRELIMINARY EXPERIMENTS

Before running preliminary experiments and testing models, the audio data must be converted into spectrogram data that can be ingested by the relevant models. Using the FFmeg application, each data observation was converted from a webm audio file to a .wav spectrogram image file. These files were then saved as TensorFlow Tensors within a TFRecords dataset, along with the relevant classification as a COVID or non-COVID observation. Regardless of the length of the original audio clip, the spectrogram images were standardized in size to be a consistent 512 by 512 pixel representation throughout the dataset. In terms of data augmentation, a contrast filter was applied to increase contrast between the audio information and the background of the spectrogram.

Once the data was prepared, the five models specified in Table 1 were run using the same general architecture: an initial layer of the specified pretrained model, then a convolutional layer, then a pooling layer, followed by a flatten layer, a 50% dropout layer, and finally a dense layer. All models were run with softmax activation and ImageNet weights, as well as a batch size of 16, a learning rate of 0.0001, and a training

TABLE III
PRELIMINARY MODEL RESULTS

Model	AUC	Validation AUC	Accuracy	Validation Accuracy
MobileNetV2	0.5927	0.5768	0.8771	0.9225
EfficientNetB2	0.5791	0.5689	0.8780	0.9144
InceptionV3	0.5721	0.5557	0.8998	0.9228
EfficientNetB7	0.5613	0.5603	0.8681	0.9215
VGG19	0.5996	0.5763	0.8742	0.9170

TABLE IV
FURTHER MODEL RESULTS

Model	Training AUC	Validation AUC
MobileNetV2	0.6655	0.6908
EfficientNetB2	0.6323	0.6600
InceptionV3	0.6852	0.6940
EfficientNetB7	0.6457	0.6599
VGG19	0.7058	0.7369

period of 5 epochs. See Table 3 for preliminary results from these experiments, focusing on the mean AUC and accuracy of the training and validation data for each model.

Based on these results, these models are highly promising in predicting the COVID status of a patient based on the spectrogram data of the patient's cough. These preliminary results are useful, but there are several next steps that can be taken to improve model performance and help determine if one specific model architecture can begin to outperform its peers in terms of spectrogram classification.

V. CONTINUED EXPERIMENTS

Continued testing of the preliminary models is necessary to identify the model and parameters best suited to classifying the COVID-19 audio spectrogram files. Each model was tested across a number of epochs and with varying combinations of batch sizes and learning rates. For each distinct combination of model, number of training epochs, batch size, and learning rate, key values including accuracy, loss, and AUC were recorded to identify the most successful combination of parameters. Shown in Figure 1 is the model architecture used.

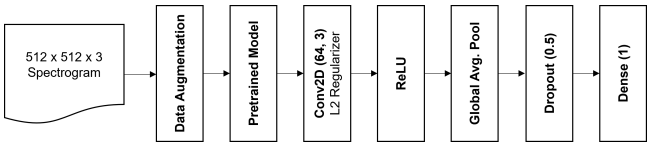

Fig. 1. Model Architecture

The effects of the spectrogram data augmentation were also investigated. Each distinct preliminary model was trained twice, once using data augmentation and once not using data augmentation. Models trained with data augmentation had applied a random horizontal flip, random change in contrast of 50%, and random change in brightness by 10%. The results of each iteration of the model were then compared across 30 epochs. For each model, it was observed that data augmentation did not positively impact the validation AUC. This is likely due to the fact that the dropout and L2 regularization were robust enough at minimizing overfitting that data augmentation was no longer beneficial.

Based on the results from retraining each model at the optimal learning rate and batch size settings out to 30 epochs, VGG19 was identified as the most successful model for predicting COVID-19 status from spectrogram data. (See Table IV). The VGG19 model was then used to generate scores representing the likelihood of an audio file corresponding to the cough of a COVID-19 positive patient for each audio file in the dataset for development of the multi-modal model.

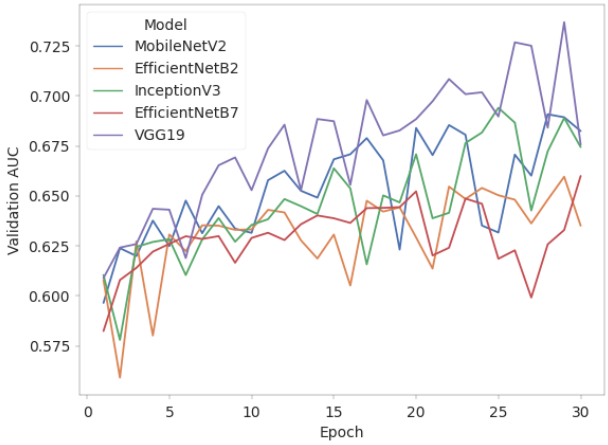

Fig. 2. Validation AUC vs. Epoch

VI. RESULTS

To finalize the multimodal model approach, a final step of model-building was applied to the newly joined dataset. After processing the spectrogram data through the deep learning model and appending the prediction scores to the full dataset, the secondary modeling process takes the form of a classification problem.

First, the dataset was prepared for analysis. Features not useful for modeling, including index columns and file names, were immediately removed from the dataframe to simplify the modeling process. Any missing values were replaced using backfill, and any infinite values were replaced with a numeric value higher than the maximum number of their respective feature. Any categorical variables were broken down into dummy variables. The dataset was then split into training data and testing data, with one third of the data being set aside for testing while the remainder was used for model building.

Four potential models were tested for the final classification task: a logistic regression model, a Gaussian Naive Bayes classifier, linear discriminant analysis, and quadratic discriminant analysis. Each model was trained on the training data split and evaluated on the holdout test set. The models were evaluated using key metrics like accuracy and F1 score, as well as a deeper look into metrics like precision and recall and the confusion matrix. (See Table V)

The results of the extensive model-building are clear. In the first phase of model-building, several deep learning models

TABLE V
CLASSIFICATION MODEL RESULTS

Model	Accuracy	F1 Score	AUC
Logistic Regression	0.9206	0.9384	0.8158
Gaussian Naive Bayes	0.6517	0.5750	0.7484
Linear Discriminant Analysis	0.9135	0.9243	0.8075
Quadratic Discriminant Analysis	0.9336	0.9384	0.9144

were investigated. Different combinations of hyperparameters were evaluated, as well as the effects of data augmentation on the success of the relevant models. After extensive evaluation and testing, the VGG19 model was found to be the most successful model for predicting COVID-19 diagnosis based off of cough audio spectrogram data. The spectrogram data was processed through this first model, where VGG19 served as the base layer for a full deep learning model architecture.

For the second phase of the mulitmodal approach, the prediction scores from the VGG19 deep learning model were appended to the existing COVID-19 patient dataset. After a bit of data preparation, four distinct classification models were run on the newly joined dataset. These models were evaluated on several key metrics, and the quadratic discriminant model was found to be the most successful classifier for the COVID-19 data.

With an accuracy level of 0.9336 and an F1 score of 0.9384, the quadratic discriminant analysis model is highly accurate and demonstrates a high level of success when it comes to classifying patients as healthy or sick with COVID-19. Compared to the spectrogram-only model AUC of 0.69, the model using a QDA architecture and patient data has an AUC of 0.9144.

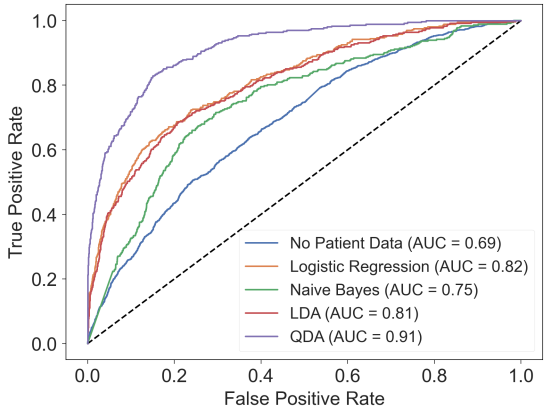

Fig. 3. ROC Curves for Multi-modal Models

VII. CONCLUSION

The implications of this model are extensive and highly promising, not only for the people of Virginia but for the general population. With an AUC of 0.9144 and an accuracy measure of 0.9336, the final model is extremely efficient and can effectively distinguish between COVID-19 patients and healthy patients with a high level of accuracy. This capability is invaluable in the context of today's society, particularly as we deal with the continuing effects of the COVID-19 pandemic.

The model has useful applications at the expected level– it was trained to classify patients and give a trustworthy COVID diagnosis based on the sound of their cough. When COVID-19 originally broke out, there were a number of issues that arose in the realm of testing. In many regions, there was a consistent shortage of available tests, so infected people were not able to confirm their status as COVID-19 patients. In areas where testing was available, PCR test results often took several days to confirm a patient's COVID-19 status. People testing often did not quarantine until they received their results, increasing the number of overall COVID-19 exposures and infections. Today, as COVID-19 continues to affect everyday life around the world, people still often struggle to find tests when they feel sick. If this model were made widely available, the need for PCR and rapid tests would diminish, as would the demand for the tests. Patients would also be able to minimize their test wait time, decreasing the number of overall exposures and infections by permitting near immediate quarantine for COVID-19 patients.

The model also has broader applications in the healthcare sphere. If this model can predict COVID-19 infections with a high degree of accuracy, audio data can feasibly be utilized to detect a multitude of illnesses, limiting the number of invasive tests necessary and the potential wait time required for such tests. Cough is a common symptom for many illnesses, from diseases requiring intense treatment like bronchitis and pneumonia to ailments as common as seasonal allergies or the common cold. Extending the findings of this analysis, a powerful model could take the input of a patient's cough audio and diagnose them with a high level of accuracy in a short period of time. Not only would this model simplify the process for patients and make their lives easier, it would also streamline things on the side of the medical professionals. Doctors and nurses would spend less time conducting tests and waiting for results, opening up their availability to administer aid and help patients. The classification model could even be administered remotely, limiting the public's exposure to the patient's illness and further streamlining the process on the healthcare provider's side.

Overall, the analysis confirms the hypothesized success of the model as a powerful patient classifier. The multimodal model combining the deep learning spectrogram classification with the patient data analysis generates highly accurate predictions of COVID-19 diagnosis, opening the door for continued diagnoses on new patient data. The implications for the people of Virginia as well as the general public are wide-ranging and highly promising, as this model would limit the need for tests and increase the speed of patient diagnosis. In future work, models like this could be expanded, moving beyond the binary classification of COVID-19 or healthy to an advanced classification of a multitude of illnesses.

VIII. Members' Contribution

In terms of each member's contribution to the project thus far, Dora Eskridge managed the second phase of model-building and the write-up portion of the project, Alexander Lilly processed the data and managed the first phase of model-building, and Matt Scheffel conducted the literature review and prepared the final presentation.

References

[1] AKGÜN, D; KABAKUŞ, A; ŞENTÜRK, Z; ŞENTÜRK, A; and KÜÇÜKKÜLAHLI, E (2021). "A transfer learning-based deep learning approach for automated COVID-19 diagnosis with audio data," Turkish Journal of Electrical Engineering and Computer Sciences: Vol. 29: No. 8, Article 15.

[2] Hemdan, E.ED.; El-Shafai, W; & Sayed, A. "CR19: a framework for preliminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications." J Ambient Intell Human Comput 14, 11715–11727 (2023). https://doi.org/10.1007/s12652-022-03732-0

[3] Feng, K; Fengyu, H; Steinmann, J; and Demirkiran, I (2021). "Deep-learning Based Approach to Identify Covid-19." SoutheastCon 2021.

[4] Bales, C; Nabeel, M; John, C; Masood, U; Qureshi, H; Farooq, H; Posokhova, I; and Imran, A. (2020) "Can Machine Learning Be Used to Recognize and Diagnose Coughs?" The 8th IEEE International Conference on E-Health and Bioengineering - EHB 2020. Grigore T. Popa University of Medicine and Pharmacy, Web Conference, Romania, October 29-30, 2020

[5] Hershey, S; Chaudhuri, S; Ellis, d; Gemmeke, J; Jansen, A; Moore, R; Plakal, M; Platt, D; Saurous, R; Seybold, B; Slaney, M; Weiss, R; and Wilson, K. "CNN ARCHITECTURES FOR LARGE-SCALE AUDIO CLASSIFICATION." Google, Inc., New York, NY, and Mountain View, CA, USA. 2017.

[6] Palanisamy, K; Singhania, D; and Yao, A. "Rethinking CNN Models for Audio Classification." Department of Instrumentation and Control Engineering, National Institute of Technology, Tiruchirappalli, India. November 2020.