Automated COVID-19 Detection Using Machine Learning

Team Members:

Rodrigo Alarcon - ralarcon2019@my.fit.edu
Emma Conti - econti2020@my.fit.edu
Lamine Deen - ldeen2016@my.fit.edu
Audrey Eley - aeley2020@my.fit.edu

Faculty Advisor:

Dr. Zahra Nematzadeh, znematzadeh@fit.edu

Client:

Dr. Zahra Nematzadeh, znematzadeh@fit.edu,
Florida Institute of Technology College of Engineering and Science: Department of Electrical Engineering and Computer Science

Date of Meeting:

01/23 - 3:30 PM

Goal and Motivation:

The COVID-19 pandemic has underscored the need for effective and innovative diagnostic tools. A web-based application that analyzes cough audio to predict COVID-19 infection can offer a convenient and non-invasive screening method, potentially aiding early detection and reducing healthcare burdens. Although at-home testing for COVID-19 is effective, it can be expensive and inconvenient. Therefore, a convenient COVID-19 screening tool to assess the need for testing can help people make an informed decision.

Approach (Key Features of the System):

The user can record their coughs and receive predictions on their COVID-19 infection status. This feature not only provides real-time feedback but also aids in maintaining a history of the user's infection status, making it a non-invasive and cost-effective tool for early screening. By tracking this data over time, the user can monitor their health status without the immediate need for a healthcare provider.

The web app design prioritizes ease of use, ensuring anyone can navigate it effortlessly and check their COVID-19 status at any time. Other aspects of the website will include details about the research and development of the ML model

The user can view a week-long progress chart, making it easier to visualize changes in their infection status over time. This feature helps determine when a user is recovering and no longer symptomatic, or still infected. The user-friendly layout ensures that navigating the web app is straightforward.

Users can access their data at any time, providing continuous access to their COVID-19 status history. This eliminates the need to wait for a healthcare provider for early detection and offers users a convenient and effective way to monitor their health

Algorithms and tools (Potentially useful algorithms and software tools):

Convolutional Neural Network (CNN)

- A neural network architecture with the potential for good performance for this classification task. This intakes cough audio as input and outputs a COVID-19 infection prediction.

ResNet50

- A 50 layer deep convolutional neural network used to classify images. For our purposes, this will be used as a benchmark model.

Django

- A web development framework. This handles the back end of the website and data transference.

TensorFlow, Numpy, Pandas

- Common python frameworks and tools that help in manipulating data and creating the neural network model.

Librosa

- Python package used to manipulate audio samples of coughing.

Novel Features/Functionalities:

The web app not only predicts COVID-19 infections based on cough recordings but can also track the progression of the user's condition over time. By analyzing daily recordings, the user can observe trends in your symptoms, making it easier to determine when medical intervention is necessary. This continuous monitoring feature offers a more personalized health-tracking experience.

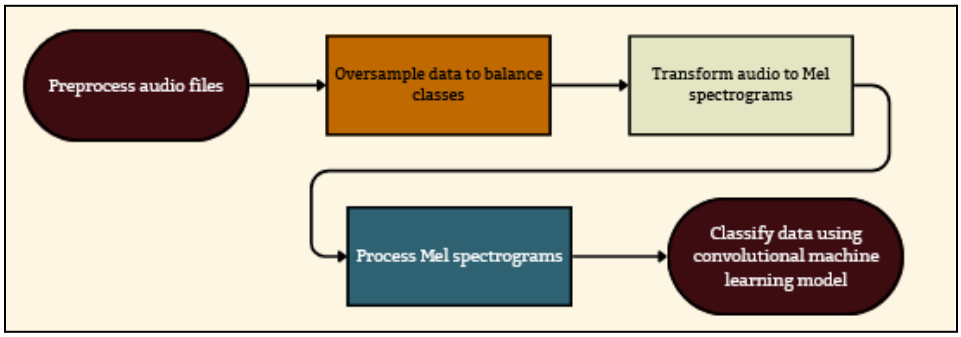
Technical Challenges:

All team members have limited experience in website application development, which is a necessary component of the final product.

The dataset should be analyzed by a convolutional neural network (CNN). Research must be done continuously to ensure that the CNN is developed appropriately, and is as accurate as possible.

Frameworks/architectures selected for this project that team members were not previously familiar with must continue to be researched to ensure their appropriate and effective usage.

Design:



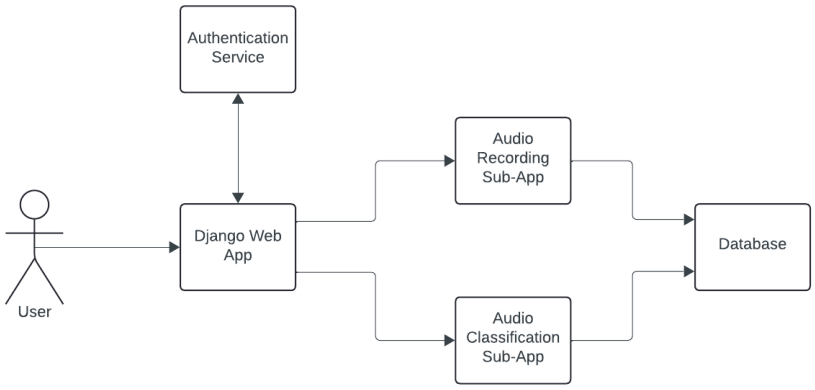
Current ML workflow

WebApp will be integrated with the CNN in order to provide results based on a trained CNN model used to identify whether or not a cough shows signs of COVID-19.

Users shall be able to access the web app to view their previous recordings and log their symptoms. These results will be graphed and shown in a usable format for the user to view their current health status.

The WebApp shall have a recording feature implemented to allow users to record their own coughs for immediate testing.

The WebApp shall have an audio recording uploading feature implemented to allow users to record their own coughs for retroactive testing.



Evaluation:

Speed: The dataset has been pre-processed to include an equal amount of cases for each type of cough that will be present during testing.

Accuracy: The level of accuracy achieved in current CNN testing is not ideal, but should improve now that the dataset has been cleaned.

Reliability: The WebApp runs reliably, and as the CNN continues to be implemented

User Survey: Users are currently able to access our website and are able to see up to date information on the research we have done and will continue to do throughout the project. Users are able to create an account that will direct them to the screen that

Progress Summary:

Module/feature	Completion %	To do
GUI	50%	allowing drop down menus, mouse-over help messages...
WebApp Home Screen	90%	Still finalizing design. Content has been completed.
WebApp Login Screen	90%	Looking for a more secure way to store usernames and passwords, otherwise the login and account creation is completely functional
WebApp sickness tracking	30%	Deciding on formatting for data to be displayed - Current status - Current recommendations - Previously logged symptoms
CNN	40%	Results are skewed due to issues with dataset, testing set to continue once dataset is cleaned
ResNet50 CNN model	40%	Results are skewed due to issues with dataset, testing set to continue once dataset is cleaned
WebApp recording feature		

Milestone 4 Tasks (Feb 24):

- Finish cleaning data
- ML testing
- Continued refinement of ML workflow
- Web testing
- Continued integration of WebApp and CNN

Milestone 5 Tasks (Mar 26):

- ML testing
- Refinement of ML workflow
- Continued integration of WebApp and CNN

Milestone 6 Tasks (Apr 21):

- Finalize ML workflow
- Complete final ML tests
- Fully integrate WebApp with CNN

Task Matrix for Milestone 4:

Task	***Rodrigo***	***Emma***	***Lamine***	***Audrey***
1. Finish cleaning data	25%	25%	25%	25%
2. ML testing	0%	50%	50%	0%
3. Continued refinement of ML workflow	10%	40%	40%	10%
4. Web testing	50%	0%	0%	50%
5. Continued integration of WebApp and CNN	40%	10%	10%	40%

Description of Planned Tasks for Milestone 4:

- Task 1: In this task, the database of audio samples must be manually reviewed to ensure the quality of data. Each audio sample is listened to by a team member. If an audio sample has more than three seconds of talking without coughing, the microphone is blown out, coughing is inaudible, or is otherwise deemed unusable for model training (e.g. too short, extremely poor audio quality), the sample is removed from the database.

- Task 2: To refine the machine learning workflow for the COVID-19 cough detection project, the team will continue enhancing the ML model by systematically implementing improvement strategies informed by comprehensive testing results. Initially, the dataset will undergo meticulous cleaning, which involves manually reviewing and listening to each audio file to eliminate noisy or irrelevant samples, thereby ensuring high data quality and integrity. Following this, the team will integrate and stack the results from experiments that demonstrated increases in training or validation accuracy, leveraging these successful approaches to build a more robust foundation for the model. This strategic selection of enhancement techniques will be pivotal in addressing any residual class imbalances and optimizing data representation.

  Furthermore, the project will advance by developing three more complex neural network architectures designed to extract intricate patterns and features from cough audio samples. These architectures will incorporate deeper layers, advanced activation functions, and sophisticated regularization methods to maximize the model's learning capacity. In parallel, ResNet50 will be employed as a benchmark to compare its performance against the custom-designed models, providing valuable insights into the strengths and potential areas for improvement of each approach. Additionally, extensive fine-tuning will be conducted, including hyperparameter optimization and the application of advanced data augmentation techniques, to further enhance model accuracy and generalization. This comprehensive refinement process aims to create a highly accurate and reliable COVID-19 cough detection system, building upon the successes of previous milestones and setting the stage for ongoing innovation.

- Task 3: As research continues and the model develops, the website will continue to grow. As pages are added, the web app will become more robust, and it will become more user-friendly. As integration continues, allowing for test cases may become possible to ensure that the web app is working as needed.

- Task 4: For this task, the machine learning and web app components need to be fully integrated. At the conclusion of Milestone 3, a placeholder model was stored within the web application, but there were issues with the model being able to accept and classify new audio samples. In this milestone, we will work to resolve this issue, and replace the placeholder model with our current CNN. As the model continues to develop, it is important to continue to integrate the most recent model with the website to determine whether it can be accommodated on the web app.

Approval from Faculty Advisor

"I have discussed with the team and approved this project plan. I will evaluate the progress and assign a grade for each of the three milestones."

Signature _____

Date _____