High accuracy classification of COVID-19 coughs using Mel-frequency cepstral coefficients and a Convolutional Neural Network with a use case for smart home devices

Rob Dunne (✉ rob.dunne@manchester.ac.uk)

University of Manchester   https://orcid.org/0000-0002-4178-4468

Tim Morris

University of Manchester

Simon Harper

University of Manchester

Short Report

Abstract

Diagnosing COVID-19 early in domestic settings is possible through smart home devices that can classify audio input of coughs, and determine whether they are COVID-19. Research is currently sparse in this area and data is difficult to obtain. However, a few small data collection projects have enabled audio classification research into the application of different machine learning classification algorithms, including Logistic Regression (LR), Support Vector Machines (SVM), and Convolution Neural Networks (CNN). We show here that a CNN using audio converted to Mel-frequency cepstral coefficient spectrogram images as input can achieve high accuracy results; with classification of validation data scoring an accuracy of 97.5% correct classification of covid and not covid labelled audio. The work here provides a proof of concept that high accuracy can be achieved with a small dataset, which can have a significant impact in this area. The results are highly encouraging and provide further opportunities for research by the academic community on this important topic.

Introduction

The 2019 novel coronavirus, COVID-19, which became a pandemic in 2020, has been the largest global public health emergency in living memory[1]. The virus is a strain of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

[1] which afflicts the respiratory system and therefore causes symptoms such as coughing and breathing difficulties, fever, fatigue, as well as ageusia and anosmia [11]. Many research efforts are underway into various aspects of COVID-19 including vaccine research [12], anti-viral treatments like Remdesivir [5], large scale literature data mining like the COVID-19 Open Research Dataset Challenge (CORD-19)[2], and diagnostic tools for determining who has the virus at early stages [11]. In this study we investigate the feasibility of high accuracy audio classification of COVID-19 coughs as a potential diagnostic software application that would be available in-home or workplace through smart devices such as Amazon's Alexa, or Google Home.

[1]WHO COVID-19, http://archive.is/SUtHp

[2]CORD-19, https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge

1.1 Background

There are a small number of research projects that demonstrate the feasibility of some types of classifiers. Brown et al at the University of Cambridge have developed a mobile phone application for crowdsourcing cough and breathing samples from members of the public [3]. They position their research as one in a long history of using bodily noises to diagnose ailments, from the logical conclusion that physiological changes can alter the natural sounds that human bodies produce [14]. They classify

coughing and breathing audio using Logistic Regression (LR), Gradient Boosting Trees with Support Vector Machines (SVMs), achieving a best accuracy (AUC) of 82%. The data distribution of participants in this paper also shows that the dataset is skewed towards middle aged people, likely due to older participants (those more vulnerable to COVID-19) being less likely to engage with mobile phone crowdsourcing technology. However, the results show no difference between the age groups in classification.

Imran et al have also developed a mobile app which can classify a COVID-19 cough from a 2 second audio recording with accuracy in the region of 90%, with some data permutations producing accuracy of 96.76% [7]. The research used Mel-frequency cepstral coecients (MFCC), a type of spectrogram image, and a Convolutional Neural Network (CNN) for classification, which demonstrated better accuracy than the LR, and SVM, used by Brown et al.

Lastly, Sharma et al created Coswara, a database of coughing, breathing, and voice sounds (vowel sounds and counting) for COVID-19 diagnosis research. The data is collected through a web application which is used to collect, label, and quality control the dataset [10]. The data was then classified using a Random Forest (RF), achieving a mean accuracy of 70% for coughing.

1.2   Rationale

Research work is currently ongoing at the University of Manchester into audio classification in smart environments, towards a larger programme of work into human behaviour prediction. It seemed likely that the classifier developed as part of that work could easily be transferred with little modification to classify COVID-19 coughs. We were able to demonstrate a proof of concept that makes it possible for this diagnostic technology to be used at scale in smart home devices for early diagnosis of the virus before patients seek out clinical treatment, particularly in those cases where they might not seek help until the symptoms have significantly progressed.

1.3   Contribution to research

A demonstration of high accuracy classification of COVID-19 coughs using an MFCC CNN machine learning architecture.

1.4   Use case

In developing our classifier we proposed a use case scenario of a smart home device user (which could easily apply to   a workplace or other location) such as an Alexa or Home device. The device passively monitors for cough sounds, and upon a positive classification for a COVID-19 cough, prompts the user to seek a professional medical diagnosis, or even calls the relevant local services for the user. This use case is shown in Figure 1.

Methods

In development of our classifier we chose to employ an MFCC CNN architecture which has proven, high accuracy results, for audio classification [2][6][9][13]. The datasets available contained several different types of audio including coughs, breathing, and speech (vowel sounds). For initial investigation we chose to focus on one of these sounds, which we hypothesised would boost the CNN performance by training against binary class options rather than attempting a multi-class CNN with different types of audio which could negatively bias our results.

2.1   Dataset

Our classifier is trained on data from three datasets. We used Google's AudioSet [4] for the baseline not covid labelled coughs. This audio is taken from multiple YouTube video clips, and the cough class has been given a 100% quality rating[3] by human verification. The COVID-19 cough audio data is from two other sources, the Corswara project [10] which is available on Github[4] and is gathered through the Corswara web application[5] which handles data collection and quality control. This is combined with data from the Stanford University led Virufy mobile app[6] which collects data and also makes it available on Github[7].

The Virufy and Corswara datasets were combined due to the low number of available COVID-19 cough audio samples; with a combined number of 17 audio samples of COVID-19 coughs in total. One of these audio samples was dropped from the dataset due to issues with it being read by our code, leaving a total of 16 COVID-19 audio samples for training and validation.

The data was split to create a balanced train/test dataset of 28 audio samples; 14 each for the covid and non covid labels. Along with a validation set of 40 audio samples, two of which are covid labels, with the remaining 38 non covid labels.

The audio data is loaded and processed into MFCCs, as shown in Figure 2, using the Librosa Python library [8]. The MFCC in the form of a $120 \times 431 \times 1$ tensor along with the corresponding label is passed to the model for training. Our audio data and labels have been made available online for further research .

[3]AudioSet cough class quality rating, http://archive.is/MZMRJ [4]Corswara dataset, https://github.com/iiscleap/Coswara-Data [5]Corswara web app, http://archive.is/6MABO

[6]Virufy, http://archive.is/hbrfE

[7]Virufy data, https://github.com/virufy/covid

[8]Audio data and labels, https://www.dropbox.com/sh/mjyrspykfx116Io/AADb3-7_SpUpF90LcPkbRV3Ga?dl=0

2.2   MFCC CNN classification

MFCC CNN architectures are very effective for classifying audio data. Audio is usually represented as a flat 1 dimensional vector of amplitude values over time, however this can be difficult to classify due to the sparse features. Converting the audio to a type of spectrogram, an MFCC, which is a 2D image as shown in Figure 2, we can use an image classifier

- already a highly developed area of machine learning - for classification of the audio as an image. We used a deep convolutional neural network, shown in Figure 7, with multiple hidden layers and a binary dense output layer for label classification.

Our model was trained using the combined balanced dataset of 28 audio samples, 14 of each label, on the University of Manchester's Research IT Computational Shared Facility (CSF) computing cluster using a NVIDIA Tesla V100 GPU. Once trained the model was passed the validation dataset of 40 audio samples (2 with the covid label) via a testing API. The results data was then logged to an SQL database. The test framework is shown in Figure 6.

Results

Training for the CNN model yielded a 100% accuracy with a small dataset, shown in the confusion matrix in Figure 3 and the plots in Figures 4 and 5. Testing with the validation dataset, which is data the model has not seen before, yielded an accuracy of 97.5%, with a single false positive result for a not covid label. Both of the covid labels in the validation dataset were correctly classified. The validation dataset results can be seen in Table 2, and a comparison of our results against similar research work can be seen in Table 1.

Discussion

Our CNN classifier proof of concept demonstrates that high accuracy audio classification of COVID-19 is possible, and could be used as a software application in a multitude of ubiquitous smart devices such as mobile phones and smart speakers. Deploying this technology to existing devices that passively monitor for trigger sounds could rapidly improve COVID-19 early detection rates in technologically developed countries.

Our results, and also that of Imran et al, confirm that an MFCC CNN approach produces superior classification results compared to other types of classifiers such as Logistic Regression, Support Vector Machine, and Random Forest.

Conclusion

Our classifier demonstrated a high accuracy of 97.5% compared to the other studies, marginally outperforming the Imran et al CNN model which achieved 96.76% accuracy, however we were able to train

our model using a much smaller dataset. We also showed that an MFCC CNN architecture significantly outperforms the Brown et al classifier algorithms of Logistic Regression, and Gradient Boosting Trees with Support Vector Machines.

Further work

There are many more opportunities for further research in this area, particularly training a classifier with a larger COVID-19 dataset, perhaps from the Brown et al project at the University of Cambridge who are working on open sourcing their crowd sourced data. Or from one of the ongoing clinical studies such as the NIH Audio Data Collection for Identification and Classification of Coughing[9] which includes COVID-19 data collection. These projects may be able to provide a larger training dataset that has been through a much more rigorous collection process.

Declarations

Acknowledgments

Statement of funding

References

1. Kristian G Andersen, Andrew Rambaut, W. Ian Lipkin, Edward C Holmes, and Robert F The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4):450–452, apr 2020.

2. Aweem Ashar, Muhammad Shahid Bhatti, and Usama Speaker Identification Using a Hybrid CNN-MFCC Approach. *2020 International Conference on Emerging Trends in Smart Technologies, ICETST 2020*, 2020.

3. Chlo¨e Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. 2020.

4. Jort Gemmeke, Daniel P.W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. Audio Set: An ontology and human-labeled dataset for audio events. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 776–780, 2017.

5. Jonathan Grein, Norio Ohmagari, Daniel Shin, George Diaz, Erika Asperges, Antonella Castagna, Torsten Feldt, Gary Green, Margaret L. Green, Fran¸cois-Xavier Lescure, Emanuele Nicastri, Rentaro Oda, Kikuo Yo, Eugenia Quiros-Roldan, Alex Studemeister, John Redinski, Seema Ahmed, Jorge Bernett, Daniel Chelliah, Danny Chen, Shingo Chihara, Stuart H. Cohen, Jennifer Cunningham, Antonella D'Arminio Monforte, Saad Ismail, Hideaki Kato, Giuseppe Lapadula, Erwan L'Her, Toshitaka Maeno, Sumit Majumder, Marco Massari, Marta Mora-Rillo, Yoshikazu Mutoh, Duc Nguyen, Ewa Verweij, Alexander Zoufaly, Anu O. Osinusi, Adam DeZure, Yang Zhao, Lijie Zhong, Anand Chokkalingam, Emon Elboudwarej, Laura Telep, Leighann Timbs, Ilana Henne, Scott Sellers, Huyen Cao, Susanna K. Tan, Lucinda Winterbourne, Polly Desai, Robertino Mera, Anuj Gaggar, Robert Myers, Diana M. Brainard, Richard Childs, and Timothy Flanigan. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *New England Journal of Medicine*, 382(24):2327–2336, jun 2020.

6. Shawn Hershey, Sourish Chaudhuri, Daniel W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin Wilson. CNN architectures for large-scale audio classification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 131–135, 2017.

7. Ali Imran, Iryna Posokhova, Haneya N. Qureshi, Usama Masood, Sajid Riaz, Kamran Ali, Charles N. John, Muham- mad Nabeel, and Iftikhar Hussain. AI4COVID-19: AI Enabled Preliminary Diagnosis for COVID-19 from Cough Samples via an App. *Informatics in Medicine Unlocked*, page 100378, apr

8. Brian Mcfee, Colin Raffel, Dawen Liang, Daniel P W Ellis, Matt Mcvicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and Music Signal Analysis in Python. Technical report, 2015.[9]NIH audio collection, http://archive.is/2an9c

9. Kuniaki Noda, Yuki Yamaguchi, Kazuhiro Nakadai, Hiroshi G. Okuno, and Tetsuya Audio-visual speech recognition using deep learning. *Applied Intelligence*, 42(4):722–737, 2015.

10. Neeraj Sharma, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Nirmala R., Pras- anta Kumar Ghosh, and Sriram Coswara – A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis. may 2020.

11. Claire J Steves and Tim D Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*.

12. Tung Thanh Le, Zacharias Andreadakis, Arun Kumar, Rau´l G´omez Rom´an, Stig Tollefsen, Melanie Saville, and Stephen Mayhew. The COVID-19 vaccine development landscape. *Nature Reviews Drug Discovery*, 19(5):305–306, may

13. Wyse. Audio Spectrogram Representations for Processing with Convolutional Neural Networks. 1(1):37–41, 2017.

14. Renard Xaviero, Adhi Pramono, Syed Anas Imtiaz, and Esther Rodriguez-Villegas. A Cough-Based Algorithm for Automatic Diagnosis of Pertussis.

Tables

Table 1: Comparative classifier accuracies for our CNN and other recent COVID-19 audio classification research projects

Classifier	Accuracy	Source
Logistic Regression	65 %	[3]
Random forest	70 %	[10]
SVM	82 %	[3]
CNN	96.76 %	[7]
Our CNN	97.5 %	-

Table 2: Validation dataset results

| Input label	Classified label	Classified label weight
not covid	not covid	0.968189775943756
not covid	not covid	0.867058038711548
not covid	not covid	0.972648620605469
not covid	not covid	0.980828821659088
not covid	not covid	0.979108273983002
not covid	not covid	0.969455122947693
not covid	not covid	0.981869757175446
not covid	not covid	0.976808428764343
not covid	not covid	0.974170088768005
not covid	not covid	0.9797745347023011
not covid	not covid	0.969428598880768
not covid	covid	0.95296049118042
not covid	not covid	0.971169650554657
covid	covid	1
not covid	not covid	0.648595631122589
not covid	not covid	0.971447110176086
not covid	not covid	0.978846549987793
not covid	not covid	0.979544579982758
not covid	not covid	0.965152859687805
not covid	not covid	0.976934015750885
not covid	not covid	0.962848961353302
not covid	not covid	0.974863409996033
not covid	not covid	0.975609183311462
not covid	not covid	0.97742623090744
not covid	not covid	0.97981196641922
not covid	not covid	0.976785659790039
not covid	not covid	0.95566463470459
covid	covid	0.999932527542114
not covid	not covid	0.97165721654892
not covid	not covid	0.976053178310394
not covid	not covid	0.968549609184265
not covid	not covid	0.974917888641357
not covid	not covid	0.971332132816315
not covid	not covid	0.973721325397491
not covid	not covid	0.891939580440521
not covid	not covid	0.972109258174896
not covid	not covid	0.9789959192276
not covid	not covid	0.978078663349152
not covid	not covid	0.966878831386566
not covid	not covid	0.979725897312164

Figures
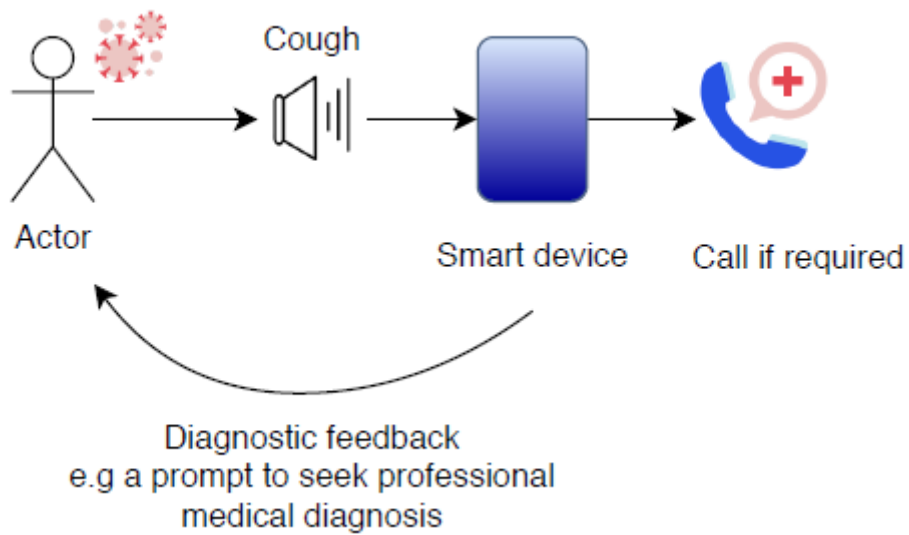
Figure 1
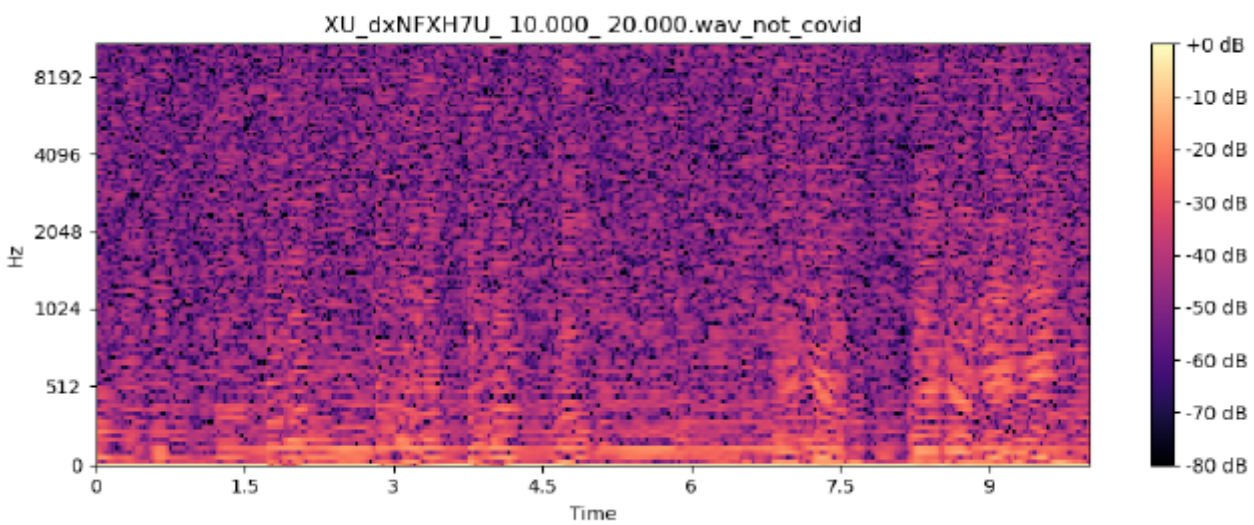
Use case diagram

Figure 2

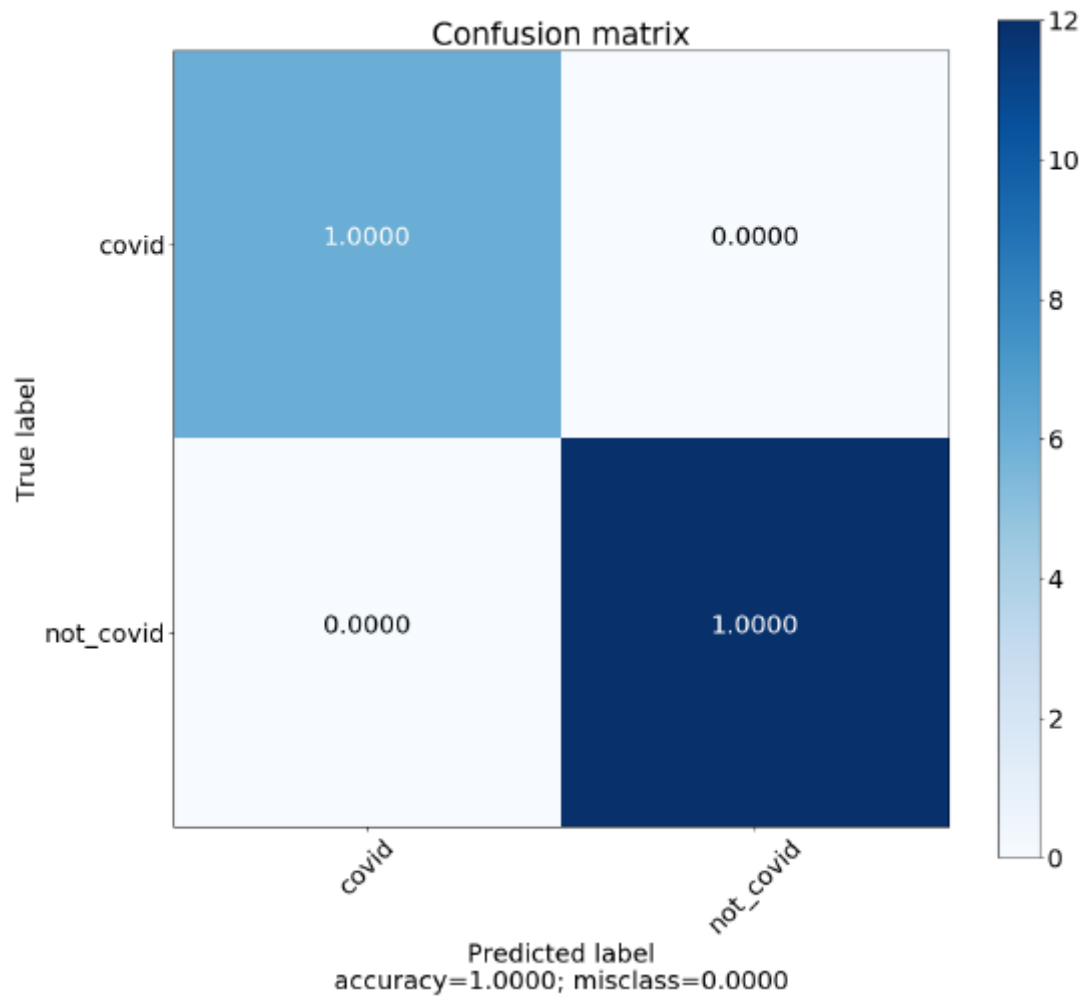MFCC for XU dxNFXH7U 10.000 20.000.wav audio

Figure 3

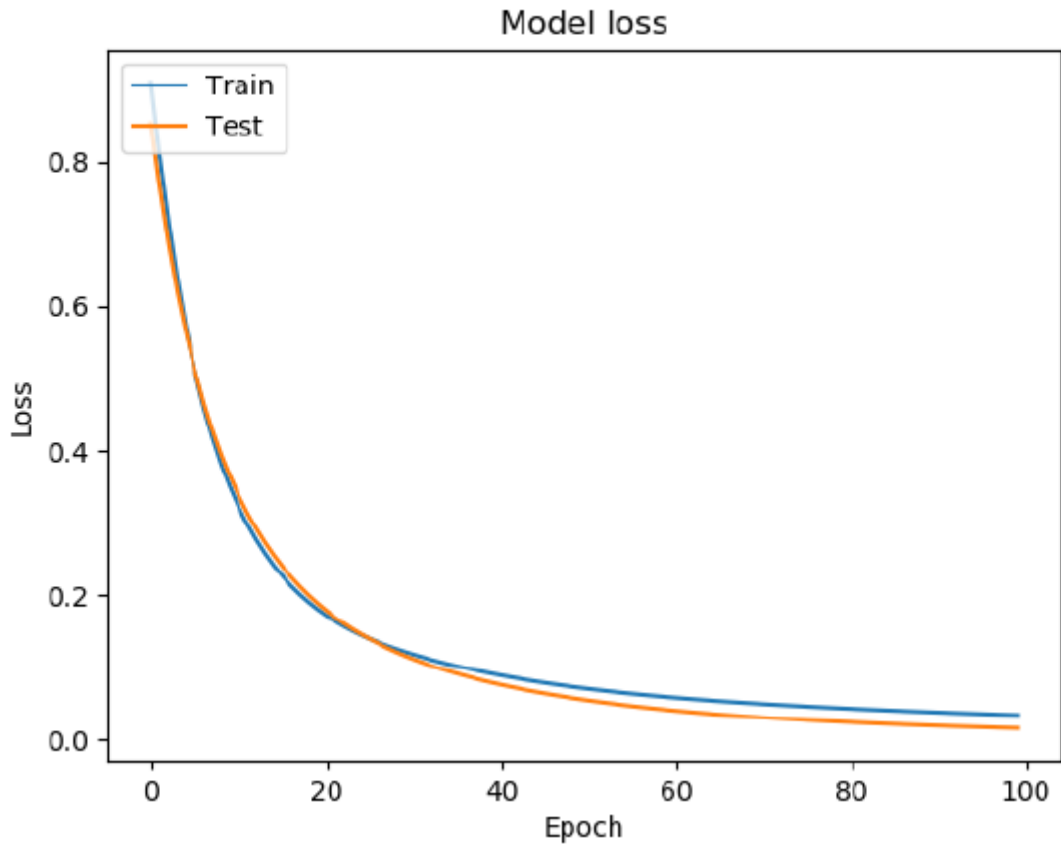Confusion matrix showing label classification accuracy

Figure 4

CNN training loss plot

Figure 5

CNN training accuracy plot

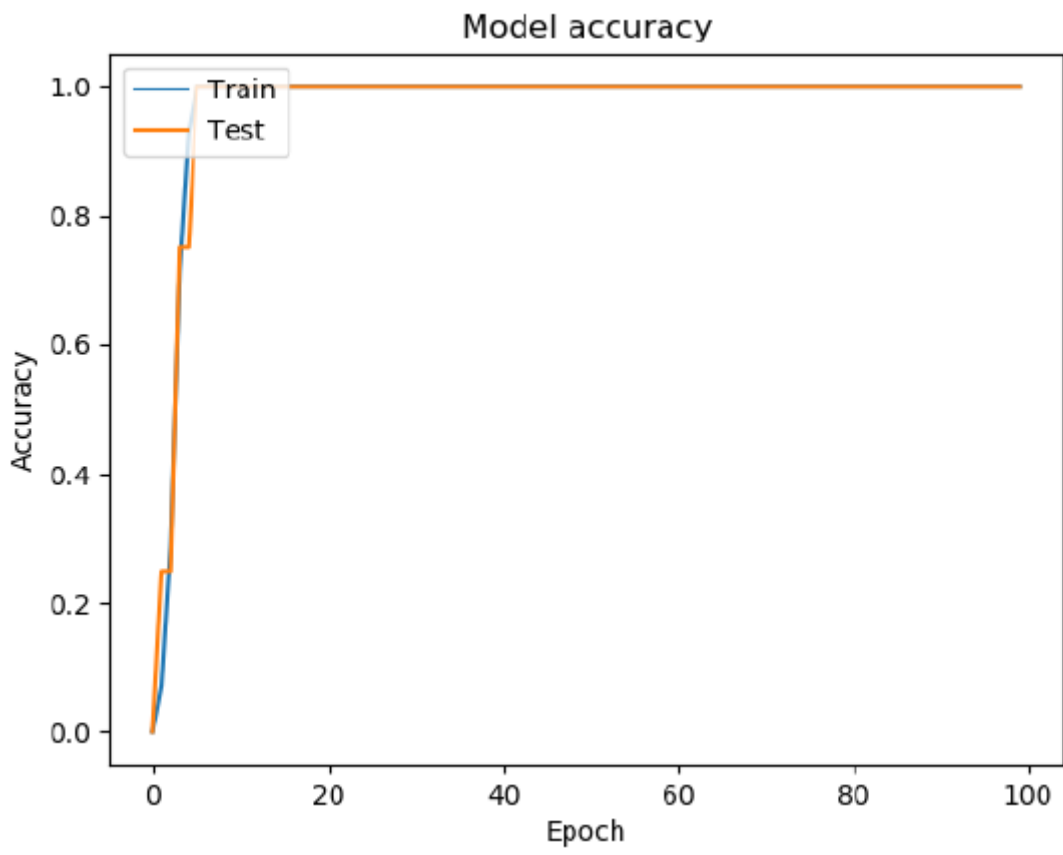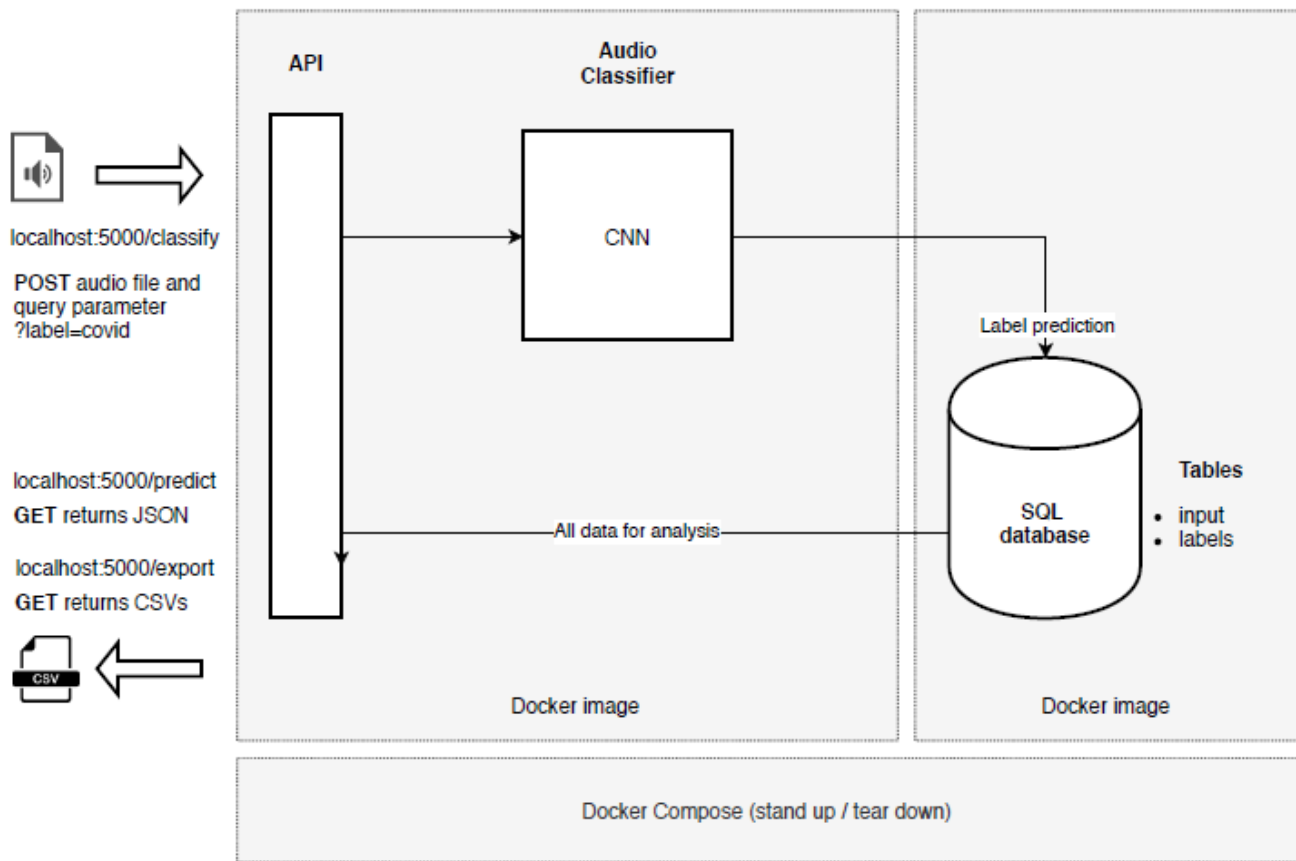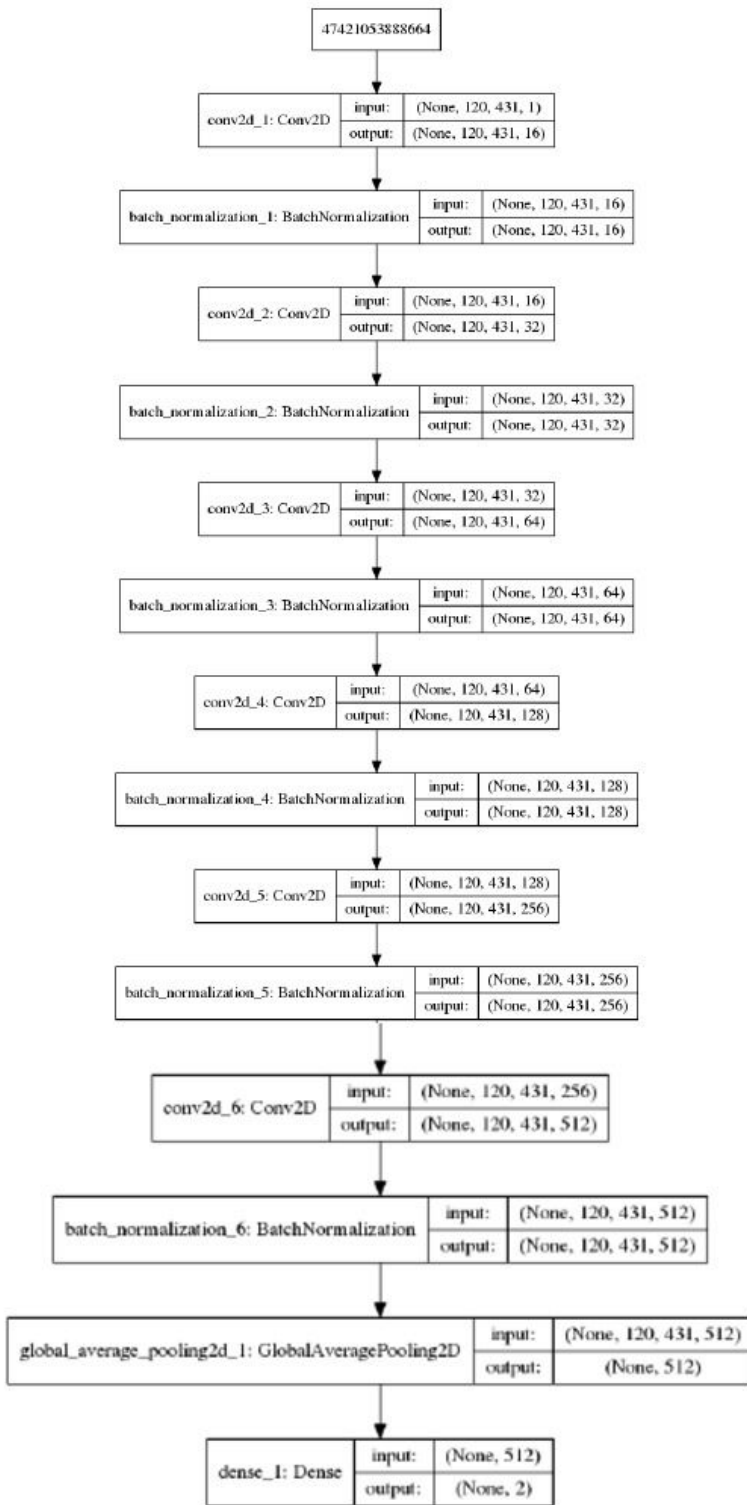**Figure 6**

Model testing framework

Figure 7

Deep Convolutional Neural Network model architecture