COVID-19 Classification Based on Cough Sound

COMP5331 Course Project Group 2

The Hong Kong University of Science and Technology — September 21, 2021

1 Basic Information

1.1 Project information

Our project title is the same as this proposal title. Table 1 describes other project and group information.

Table 1: Project and group information

Project topic	Project type	Group number
Classification	Implementation	2

1.2 Group memeber information

Group 2 consists of six group members: CHAO Chung-chi, LI Jiabao, MO Zongchao, TANG Jihong, WANG Yubo, YANG Lingyun. The detailed information about these members could be shown as follows.

Group Member 1

(a) Student ID: 20562119

(b) Student Name: Chung-chi CHAO

(c) FYP supervisor: Prof. Shing-Chi CHEUNG

(d) FYP topic and explanation:
My FYP topic is deep testing of DNN systems. We are aiming to provide a set of detection criteria to uncover faults in DNN systems especially in the data aspect. The work will mostly be focused on designing and implementing rules to detect and categorize potential data faults given an input dataset, which is not related to this group project.

(e) Declaration statement:
I declare that this project is done solely within the course but not other scopes.

Group Member 2

(a) Student ID: 20718615

(b) Student Name: Jiabao LI

(c) Research supervisor: Prof. Jiguang WANG

(d) Research topic and explanation:
My research topic is to uncover the evolution process within biology, mostly in cancer genomics and radiomics. I am working on the simulation and the calculation of the brain tumor by machine learning, which is unrelated to the group project.

(e) Declaration statement:
I declare that this project is done solely within the course but not other scopes.

Group Member 3

(a) Student ID: 20755950

(b) Student Name: Zongchao MO

(c) FYP supervisor: Prof. Jiguang WANG

(d) FYP topic and explanation:
Computational biology. My research focuses mainly on computational biology. We analyze cancer genome sequencing data to identify both germline and somatic genome alteration contributing to tumorigenesis and development. And we use computational methods to investigate the dynamic expression change during cancer treatment.

(e) Declaration statement:
I declare that this project is done solely within the course but not other scopes.

Group Member 4

(a) Student ID: 20815724

(b) Student Name: Jihong TANG

(c) Research supervisor: Prof. Jiguang WANG

(d) Research topic and explanation:
As one first year PhD student, my research topic has not been focused. But my supervisor and my lab have focused on brain tumor related computational methods development and data analysis. Therefore, my future research work will be related to the brain tumor with a high probability, which is not related with the course project.

(e) Declaration statement:
I declare that this project is done solely within the course but not other scopes.

Group Member 5

(a) Student ID: 20840755

(b) Student Name: Yubo WANG

(c) Research supervisor: Prof. Lei CHEN

(d) Research topic and explanation:
Knowledge Extraction. The knowledge extraction procedure is a way that we could obtain information and data from sources like web pages or html files. Because the data contained in the data sources are usually not formatted, therefore we develop various methods or algorithms to extract them. Hence my research topic is nothing related to the classification work.

(e) Declaration statement:
I declare that this project is done solely within the course but not other scopes.

Group Member 6

(a) Student ID: 20715584

(b) Student Name: Lingyun YANG

(c) Research supervisor: Prof. Wei WANG

(d) Research topic and explanation:
My research interests include machine learning for systems, cloud computing and resource management for large-scale clusters. Currently, I am working on the research mainly about resource management for large-scale gpu clusters. In this project, our group chooses a topic on COVID-19 cough classification, which is different from my own research.

(e) Declaration statement:
I declare that this project is done solely within the course but not other scopes.

2 Project description

2.1 Background

Since December 2019, the coronavirus disease 2019 (COVID-19) has been the health hotspot worldwide. Caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), COVID-19 has been highly transmissible and spread fast around the whole world. Since 2021, variants of the virus have emerged and become dominant in many countries, with the Delta, Alpha and Beta variants being the most virulent. By the writing time of the proposal, COVID-19 has confirmed cases exceeding 228 million world-wide [1]. Different variants of the virus gave them the ability to survive under the pressure of vaccines. Despite the global vaccination, the COVID-19 has caused more than 4.68 million deaths, making it one of the deadliest pandemics in history.

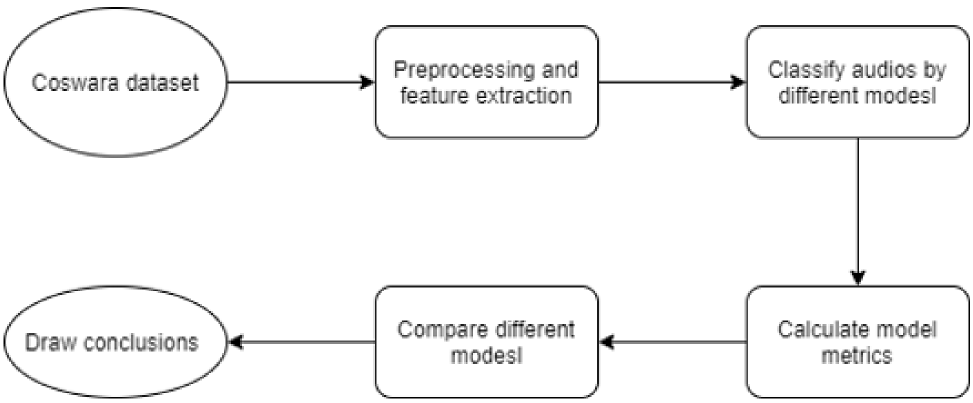
The diagnosis of COVID-19 is made primarily by direct detection of SARS-CoV-2 RNA by nucleic acid amplification tests (NAATs). In addition to the direct detection method, the research and industry community has developed multiple methods focusing on the fast, easily accessible, and possible contactless diagnosis of COVID-19. Among these, the usage of cough sound collected from smartphone apps, trained with machine learning or deep learning models, for detecting and classifying COVID-19 has become popular recently [2]. The cough sound-based diagnosis method shows its advantage in different areas. Firstly, it may decrease the demands for facilities and resources compared to NAAT methods, such as medical supplies and experienced workers. Secondly, it may reduce the transmissible risk for its contactless data collection procedure.

However, more data preprocessing and modeling development work needed to be done to increase the accuracy of such methods. Therefore, we propose focusing on the diagnosis method based on cough sound in this course project. We aim to implement the most popular classification models in related papers, and give our evaluations based on their performances.

2.2 Data preprocessing

To fulfill our aim, the coswara dataset [3] is used in our project. The coswara data has the public part and the private part. For the released coswara data, the metadata and the audio of the participants can be found in Github (https://github.com/iiscleap/Coswara-Data) or Kaggle (https://www.kaggle.com/janashreeananthan/coswara). Updated on Sept 14th, 2021, the coswara dataset includes the audio data with 40 different dates, and about 4000 samples are collected. For the audio data, the breathing sounds, cough sounds, and phonation sounds are recorded. For the metadata, the age, gender, location, health status, and the related medical information of each participant are preserved in the coswara dataset.

However, there is still a long way to use the cough audio straightforward to classify the different labels. In our project, we are going to utilize the audio information as the features and the meta-information as the labels to fulfill our classification of the COVID-19 audio. In the audio preprocess step, the digital signal processing methods in the references [4] and the related package (https://github.com/jameslyons/python_speech_features) will be applied to our project so as to get the formatted input features. After that step, the distribution of the public coswara dataset (including the features and the labels) will be visualized by the statistical and dimensional reduction methods.



2.3 Modeling work

After the data preprocessing work, the classification models can be applied to the formatted dataset to explore the potential subgroups of the COVID-19 cough audio.

2.3.1 VAE

For the variational autoencoders (VAE) model, the encoder and the decoder are included in the model, which are popular model in the audio and music areas [5]. The VAE model may also be able to classify our coswara dataset with multiple clinical labels, and the VAE model will be implemented by packages such as pytorch and keras.

2.3.2 VGGish

VGGish [6] is an audio classification model that adopts the structure of VGG [7], which is a convolutional neural network originally used for image classification. VGG was the first runner-up of ILSVR2014 in the ImageNet classification task. There are slight modifications to the architecture of VGGish compared to that of VGG, including the input size being changed to better meet the needs of log mel spectrogram inputs, dropping the last group of convolutional and maxpool layers, and using a 128-fully connected layer instead of a 1000-fully connected layer. As VGGish was trained on a large collection of audio data, AudioSet [8], VGGish is commonly used as a feature extractor to convert raw audio inputs to 128-D embeddings. The pre-trained weights are also publicly available.

There has been previous work adopting VGGish extracted features, handcrafted features, and combined features to classify respiratory sounds [6]. There has also been previous work on lung sound recognition using VGGish and bidirectional GRU [9]. In this project, we aim to leverage VGGish to distinguish between covid cough sounds and non-covid cough sounds.

2.3.3 GRU

Compared to Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) can better model sequences of data since they are able to handle a variable-length sequence input. In recent research work [10, 11], RNNs have already been adopted for COVID-19 cough classification. Gated Recurrent Unit (GRU) was proposed by the reference [12], which is a gating mechanism in RNNs [13] and is similar to long short-term memory (LSTM) [14] but has fewer parameters. It can make each recurrent unit adaptively capture dependencies of different time scales. GRU has gating units which modulate the information flow inside the unit, however, it does not have any separate memory cells. Based on the experiments [13], GRU mostly achieves comparable performance to LSTM, sometimes even better performance on some datasets in terms of convergence in CPU time and parameter updates and generalization. It has been shown that GRU can achieve better performance on certain smaller and less frequent datasets [14]. In this project, we will include GRU model in the comparison to evaluate its performance on COVID-19 cough classification.

2.3.4 Transformer

Similar to RNNs, transformer [15] is designed to handle a variable-length sequential input data. This machine learning model has been widely used in many fields, such as computer vision and natural language processing. The key of its success is the adoption of the attention mechanism, which differentially weighs the significance of each part of the input data and empower the model to focus on important areas. Without using any RNN structure, transformer can achieve dominant performance when handling the sequential data. We will also apply the transformer to COVID-19 cough classification to evaluate its performance.

2.3.5 Transformer-CP

Transformer-CP is a model enabling contrastive pre-training. A contrastive learning method can benefit from large batch size and can avoid overfitting problems in the downstream network. By introducing a random masking mechanism, the feature encoder would be more robust. The masking generator generates a masking matrix with a specific masking rate. Based on the masking matrix and the masking rate, some of the inputs are randomly masked and removed from the attention calculation in the Transformer. The loss function used in this phase for contrastive learning is a multi-class cross-entropy function working together with the similarity metric.

2.4 Comparison between different models

In the end, we will compare different models (e.g., VAE, GRU, Transformer) by the same metrics to evaluate this classification problem, such as the common metrics [16]. The preprocessing methods, models and classification metrics will be implemented by us and we will also compare the models by the metrics, which is a work contributed by us equally. Eventually, by comparing the performance of the models which we have implemented, the potential suggestions on diagnosis methods or detection COVID-19 by audio information may also be given in the final part of our project.

References

[1] Wikipedia contributors. Covid-19 pandemic by country and territory — Wikipedia, the free encyclopedia, 2021. [Online; accessed 19-September-2021].

[2] H. Xue and F. Salim. Exploring self-supervised representation ensembles for covid-19 cough classification, 2021.

[3] N. Sharma, P. Krishnan, R. Kumar, et al. Coswara–a database of breathing, cough, and voice sounds for covid-19 diagnosis. *arXiv preprint arXiv:2005.10548*, 2020.

[4] E. A. Mohammed, M. Keyhani, A. Sanati-Nezhad, et al. An ensemble learning approach to digital corona virus preliminary screening from cough sounds. *Scientific Reports*, 11(1):1–11, 2021.

[5] L. Girin, S. Leglaive, X. Bie, et al. Dynamical variational autoencoders: A comprehensive review. *arXiv preprint arXiv:2008.12595*, 2020.

[6] S. Hershey, S. Chaudhuri, D. P. W. Ellis, et al. Cnn architectures for large-scale audio classification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[8] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, et al. Audio set: An ontology and human-labeled dataset for audio events. In *Proc. IEEE ICASSP 2017*, New Orleans, LA, 2017.

[9] L. Shi, K. Du, C. Zhang, et al. Lung sound recognition algorithm based on vggish-bigru. *IEEE Access*, 7:139438–139449, 2019.

[10] A. Hassan, I. Shahin, and M. B. Alsabek. Covid-19 detection system using recurrent neural networks. In *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, pages 1–5, 2020.

[11] M. Pahar, M. Klopper, R. Warren, et al. Covid-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine*, 135:104572, 06 2021.

[12] K. Cho, B. van Merriënboer, D. Bahdanau, et al. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar, October 2014. Association for Computational Linguistics.

[13] J. Chung, C. Gulcehre, K. Cho, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.

[14] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

[15] A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, et al., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[16] M. Hossin and M. N. Sulaiman. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2):1, 2015.