2 Fundamentals of Cough Detection and COVID-19  
Cough Recognition  
Tiếng ho là một triệu chứng phổ biến của nhiều loại bệnh liên quan về đường hô  
hấp. Vì vậy tiếng ho mang những thông tin quan trọng về những thay đổi của  
đường hô hấp và tình trạng của bệnh nhân nên ngày nay có nhiều hệ thống được  
phát triển áp dụng công nghệ AI, máy học hiện đại để xử lý dữ liệu về tiếng ho  
với mục đích trích xuất các thông tin hữu ích. Với thực trạng thế giới đối mặt  
đại dịch COVID-19, các công trình nghiên cứu xoay quanh xử lý dữ liệu tiếng  
ho được đẩy mạnh hơn với các nghiên cứu liên quan đến tiếng ho và COVID-19,  
tiêu biểu có thể kể đến là dự đoán kết quả nhiễm bệnh qua tiếng ho với công cụ  
là trí tuệ nhân tạo, máy học.

Cough is a common symptom of various respiratory diseases. Since cough sound contains vital aspects related to respiratory change in particular and current patient’s circumstance in general, a variety of modern AI system and state-of-the-art machine learning have been developed with the aim to extract useful features assisting later research. During the situation of coronavirus breakout, research in sound track especially in cough sound is considered one the focal keys to reduce pandemic stress. AI gadgets and machine learning systems have been generating many viable solutions in diagnosing illnesses through patient’s cough sound.

2.1 Cough Detection  
Bên cạnh bài toán phân loại kết quả nhiễm bệnh của bệnh nhân COVID-19,  
chúng tôi muốn đến đề cập đến một vấn đề đi kèm cũng không kém phần quan

trọng là phát hiện tiếng ho. Tại đề mục này chúng tôi trích dẫn một số nghiên  
cứu có liên quan về phương pháp sử dụng công cụ máy học, học sâu để phát  
hiện tiếng ho trong bài toán phân loại kết quả nhiễm bệnh COVID-19, Qua đó  
liệt kê các phân tích và kết quả đạt được của các nhóm tác giả.

Along with illness classification issue conducted in COVID-19 patients, we also mention several other undoubtedly essential aspects in cough detection. In this section, we cite some related research employing machine learning, deep learning to detect cough with the aim to serve COVID-19 classification process. Regarding these elements, we make a list of analysis and results achieved by each group of coauthors.

Hiện nay đã có nhiều công trình nghiên cứu và phân tích về việc áp dụng công  
nghệ trí tuệ nhân tạo. Tiên phong như bài báo A Machine Hearing System for  
Robust Cough Detection Based on a High-Level Representation of Band-Specific  
Audio Features của Jesús và các cộng sự là một trong những công trình quan  
trọng về việc áp dụng các kỹ thuật máy học, học sâu lên dữ liệu âm thanh ho.  
Bài báo đề xuất một hệ thống máy nghe cho phân đoạn ho dựa trên âm thanh,  
phương pháp cough detection của hệ thống được thực hiện trong hai bước. Đầu  
tiên, tập hợp thuộc tính của quang phổ ngắn hạn (short-term spectral) được  
tính toán riêng biệt trong năm dải tần số được xác định trước, phương pháp  
lựa chọn và kết hợp thuộc tính được áp dụng để làm cho tập hợp thuộc tính  
mang thông tin đủ hiệu quả trong các tình huống nhiễu khác nhau. Thứ hai,  
biểu diễn dữ liệu cấp độ cao đạt được bằng cách tính toán giá trị trung bình  
và độ lệch chuẩn của các bộ mô tả ngắn hạn trong các khung dài hạn 300ms.  
Cuối cùng, cough detection được thực hiện bằng cách sử dụng Support Vector  
Machine được đào tạo với dữ liệu từ các tình huống nhiễu khác nhau. Hệ thống  
đạt được độ nhạy 92.71%, độ đặc hiệu 88.58% và 90.69% là một kết quả nghiên  
cứu vượt trội trong các phương pháp hiện đại (SOTA), mở đường cho việc tạo  
ra một thiết bị theo dõi ho trong các tình huống thực tế.

Currently, there have been many authors applying AI technology in their research and analysis. As a pioneer, “A Machine Hearing System for  
Robust Cough Detection Based on a High-Level Representation of Band-Specific  
Audio Features” carried out by Jesús et al. is one of typical related research using cough sound as input in their machine learning and deep learning models. This paper constructed a machine hearing system to segment cough sound. Specifically, cough detection solution is conducted in two steps. In the first step, it calculates a short-term spectral feature set separately for five ranges of predefined frequency. Feature selection and combination help this feature set contain sufficient effective information in different context of noisy space. In the second step, high-level data representation results from the mean and standard deviation computation of short-term descriptors in 300 ms long-term frames. Eventually, various noisy scenarios are applied in training Support Vector Machine model for cough detection phase. The system gets outstanding result in several considerations which are 92.71%, 88.58% and 90.69% for sensitivity, specificity and Area Under Receiver Operating Charcteristic (ROC) curve (AUC) respectively. This research opens the route for detecting and monitoring cough in reality.

Cough Detection Using Selected Informative Features from Audio Signals  
của Menghan Hu và các cộng sự đề xuất một hệ thống phân loại có nhiệm vụ  
phát hiện âm thanh ho trong các bản ghi âm. Các mô hình trong thí nghiệm  
được đào tạo bởi bộ dữ liệu kết hợp bộ dữ liệu ESC-50 với các đoạn ghi âm ho  
tự ghi, bộ dữ liệu kiểm thử chứa các bản ghi âm ho của bệnh nhân. Để thực hiện  
thí nghiệm, Menghan Hu và các cộng sự áp dụng các phương pháp lựa chọn đặc  
trưng và trích xuất đặc trưng khác nhau gồm có Random Frog, Uninformative  
Variable Elimination (UVE) và Variable influence on projection (VIP) nhằm  
mục đích tối ưu hóa không gian thuộc tính của dữ liệu.

Cough Detection Using Selected Informative Features from Audio Signals conducted by Menghan Hu et al. sets foot on the system bringing the mission of cough events detection in a general cough audio. The training process employs combined ESC-dataset which recordings were self-recorded, and the test set contains inpatient’s cough recordings. In the experiment, Menghan Hu and her coauthors used many algorithms including Random Frog, Uninformative  
Variable Elimination (UVE) and Variable influence on projection (VIP) to choose different feature numbers for constructing 15 cough detectors with the purpose of mitigating feature dimension.

2.2 COVID-19 Cough Recognition  
Nhiều nhóm tác giả cũng cấp các nghiên cứu về cả bộ dữ liệu và các thực nghiệm  
khác nhau áp dụng các phương pháp ML và DL trong nỗ lực cung cấp một giải  
pháp sàng lọc COVID-19 hiệu quả và tiết kiệm. Tại điểm mục này, chúng tôi  
trích dẫn một số nghiên cứu về bộ dữ liệu và các phương pháp phân loại COVID  
dựa vào tiếng ho. Qua đó liệt kê các kết quả và phân tích về phương pháp tiếp  
cận của các nhóm tác giả cho bài toán.

Many authors also provide the COVID-19 screening system with numerous research, datasets and different promising experimental results in machine learning and deep learning with the desire to diagnose faster, money-saving and more effectively. In this section, we cite a number of dataset research and current COVID cough classification methods in order to make a general list of analysis and its results in the aforementioned problem.

Dự án Coswara [1] được tạo ra bởi Neeraj Sharma và các công sự với mục tiêu  
cũng cấp cho cộng đồng nghiên cứu khoa học một bộ dữ liệu về tiếng ho, hơi thở  
và giọng nói phục vụ các nghiên cứu liên quan tới COVID-19. Dự án thu thập dữ  
liệu được đóng góp bởi người dùng trên toàn thế giới qua một giao diện website  
được cung cấp bởi nhóm tác giả (https://coswara.iisc.ac.in/). Các bản thu âm  
sẽ được thu thập thông qua các thiết bị điện tử như điện thoại thông minh hoặc

máy tính có kết nối Internet. Tại thời điểm thực hiện nghiên cứu của chúng tôi,  
chúng tôi sử dụng 4,465 audios được cung cấp từ dự án Coswara ( hai loại tiếng  
ho: shallow cough and heavy cough được lưu trữ dưới dạng filename.wav). Theo  
nghiên cứu [2], bộ dữ liệu Coswara có tiếng ho của người dương tính với COVID  
ngắn hơn tiếng ho của người bình thường từ 15% tới 20%. Nhóm tác giả cũng  
cung cấp một file metadata chứa các thông tin cá nhân cũng như tình trạng  
bệnh lý do người đóng góp tiếng ho cung cấp.

Coswara project established by Neeraj Sharma et al. aims at providing scientific community with a dataset of cough, breath and voice serving for COVID-19-pertaining research. The author created a website (<https://coswara.iisc.ac.in/>) that global users can contribute their recordings from anywhere at ease. Those recordings are collected through internet-connected digital devices such as smart phones or computers. During the course of researching period, we employ 4,465 audios from Coswara project (2 types of coughs: shallow cough and heavy cough stored following filename.wav). About the study, there is a special insight between COVID cough and non-COVID cough, which is in the duration of them, statistics shows that COVID cough audios are shorter than non-COVID cough audios 15% to 20%. The authors also send a metadata file including personal information and illness situation together with contributor’s audios.

Với nỗi lực cung cấp một bộ dữ liệu lớn và chất lượng cho việc nghiên cứu  
các giải pháp phân loại COVID-19 qua tiếng ho, nhóm tác giả Lara Orlandic và  
các cộng sự đã xây dựng bộ dữ liệu COUGHVID [3] cung cấp hơn 25,000 tiếng  
ho thu thập từ cộng đồng. Bên cạnh đó, có bốn chuyên gia có kinh nghiệm trong  
tiến hành đánh giá, phân tích và gán nhãn cho hơn 2,8000 bản thu âm tiếng ho  
khác nhau. Qua đó nhóm tác giả tin rằng COUGHVID là bộ dữ liệu lớn nhất  
và uy tín cho các phân tích và thửc nghiệm các giải pháp phân loại COVID  
qua tiếng ho. Ngoài bộ dữ liệu, nhóm tác giả còn cung các kết quả từ các giải  
pháp nhận diện tiếng ho được xây dựng trên bộ dữ liệu để tối ưu hoá công việc  
tiền xử lý trên bộ dữ liệu. Qua đó, các nhóm tác giả khi sử dụng bộ dữ liệu  
có thể tham khảo và đánh giá. Nhóm tác giả cũng cung cấp một tập metadata  
gồm các thông tin của người tham gia đóng góp tiếng ho kèm theo đó là các kết  
quả nhận diện mức độ tiếng ho trong các audio tại field cough\_detected, đây là  
những thông tin vô cùng hữu ích cho các nhóm nghiên cứu khác khi thực hiện  
nghiên cứu của mình trên bộ dữ liệu này.

Intended to generate a large and high-quality dataset for further research in COVID-19 solutions through cough, Lara Orlandic et al. created COUGHVID with more than 25,000 coughs collected from the community. Besides, there are four experienced experts taking the responsibility for evaluating, analyzing and labeling more than 28,000 different cough sounds. Based on this dataset, the authors claimed that COUGHVID is the largest and most reliable for cough COVID analysis and experiments. In addition to the dataset, the authors also present experimental results from many cough detection remedies generated by this dataset in order to optimize dataset preprocessing phases itself. As a result, in the following research, other researchers can refer and compare evaluation. As above research, this research also brings along a metadata file containing individuals’ information who join the project, and detection results of audios in cough\_detected filed. These are invaluable information for further research carried out by the following groups.

Ali Imran và các cộng sự trong nghiên cứu [4] đã đề xuất một quy trình phân  
loại COVID-19 qua tiếng ho gồm hai thành phần Cough collection and detection  
và COVID-19 dyagnosis - đây cũng là operations tương tự như chúng tôi xây  
dựng trong nghiên cứu này. Ali Imran và các cộng sự đã chỉ ra rằng tiếng ho của  
bệnh nhân nhiễm COVID-19 có những đặc trưng riêng biệt khi so sánh với tiếng  
ho của các bệnh nhân nhiễm các bệnh về hô hấp khác như Laryngitis, Tracheitis,  
Lung abscess, Lung tumor, Pleural diseases, Interstitial lung disease, etc. Đây là  
một phát hiện quan trọng chứng minh những nỗ lực trong việc phân tích và áp  
dụng AI trong việc phân loại bệnh nhân COVID-19 thông qua tiếng ho là hoàn  
toàn khả thi. Nghiên cứ [4] cũng cung cấp kết quả phân loại COVID-19 với độ  
chính xác trên 80% tới nhỏ hơn 93%. Tuy nhiên điểm hạn chế từ nghiên cứu này  
là bộ dữ liệu sử dụng còn quá ít (70 COVID-19, and 247 normal cough samples  
from different people) đối với việc phát triển một giải pháp AI giải quyết bất  
cứ bài toán nào. Bên cạnh đó bộ dữ liệu cũng không được public để các nhóm  
nghiên cứu khác có thể thực hiện các thử nghiệm và đánh giá khác trên bộ dữ  
liệu.

Ali Imran et al. in the study [4] have suggested a 2-part COVID cough classification process which are Cough collection and detection and COVID-19 dyagnosis. This is also the similar operations to implement in our research. Ali Imran et al. point out that COVID-19 infected patient’s coughs have distinct features when they make a comparison with other respiratory diseases’ sounds namely Laryngitis, Tracheitis,  
Lung abscess, Lung tumor, Pleural diseases, Interstitial lung disease, etc. . This is an important finding proving that extensive effort in analyzing and applying AI into classifying COVID-19 coughs results in workable remedies. The study [4] also possibly returns COVID-19 classification results with the accuracy from above 80% to less than 93%. Nevertheless, the weakness of this research is the small size of employed data (70 COVID-19, and 247 normal cough samples from different people) in relation to the development of a comprehensive AI solution for all the issues. Another problem is that it is not allowed to access the dataset for further experiments and evaluation.

Một nghiên cứu khác của nhóm tác giả Madhurananda Pahar và các cộng sự  
[2] đã cung cấp các kết quả thực nghiệm với phương pháp tiếp cận ML và DL  
cho giải pháp phân loại COVID-19 qua tiếng ho. Nhóm tác giả đã thực hiện các  
thử nghiệm của mình trên bộ dữ liệu Coswara (tại thời điểm nhóm tác giả thực  
hiện nghiên cứu Coswara cũng cấp 92 mẫu tiếng ho của bệnh nhân dương tính  
với COVID-19 và 1079 mẫu âm tính) và Sarcos cái mà được thu thập chủ yếu  
từ người dân ở South Africa bao gồm 18 mẫu tiếng ho có nhãn dương tính với  
COVID-19 và 26 mẫu âm tính. Tổng hợp các mẫu tiếng ho từ cả hai tập dữ liệu,  
Madhurananda Pahar và các cộng sự đã có được mẫu tiếng ho tới từ người dân  
tới từ bốn lục địa khác nhau (Asia, Europe, Oceania và America). Các tác giả  
đã đưa ra các kết quả thực nghiệm của mình với các mô hình ML và DL cùng  
với phương pháp khắc phục việc thiếu dữ liệu khi sử dụng synthetic minority  
oversampling technique (SMOTE) [5] và phân tích đặc trưng sử dụng sequential  
forward selection (SFS). Kết quả tốt nhất nhóm tác giả đạt được là với mô hình  
Resnet50 với AUC of 0.976 khi huấn luyện và đánh giá kết quả trên tập dữ liệu  
Coswara, AUC of 0.938 khi đánh giá trên bộ dữ liệu Sarcos. Nghiên cứu của  
Madhurananda Pahar và các cộng sự cung cấp các thực nghiệm đa dạng với các  
giải pháp ML và DL cùng với các phương pháp trích xuất đặc trưng và xử lý  
mất cân bằng dữ liệu. Tuy nhiên, giống như [4] số lượng dữ liệu để xây dựng  
các giải pháp là quá nhỏ do đó các kỹ thuật xử lý dữ liệu cũng không thể khắc  
phục hoàn toàn nhược điểm này. Bên cạnh đó bộ dữ liệu Sarcos cũng không  
được publish, vì vậy các nghiên cứu sau không thẻ thiết kế lại các thực nghiệm  
để kiểm chứng các kết quả được báo cáo.

A different study of Madhurananda Pahar et al. publishes several experimental results in ML-based and DL-based COVID-19 classification through coughs. The authors make experiments on Coswara dataset (there are 92 positive COVID-19 samples and 1079 negative samples in Coswara at the researching time) and Sarcos, which is mainly collected from South Africa residents, includes 18 positive COVID cough samples and 26 negative samples. After collecting cough samples from both datasets, Madhurananda Pahar et al. receive samples from four different continents (Asia, Europe, Oceania and America). They combine ML and DL models with synthetic minority oversampling technique (SMOTE) [5] – a method to solve imbalance dataset, and sequential forward selection (SFS) – a feature analysis method, in order to get better performance. The highest score of AUC using Resnet50 are 0.976 and 0.938 when these are trained on Coswara and Sarcos respectively. Although Madhurananda Pahar et al. employ various different ML and DL solutions along with feature extraction and imbalance handles, alike [4], a handful number of samples are not able to return outstanding results. As a result, the shortage of data samples is still existed as well as Sarcos dataset is not published - more obstacles to any other researchers in not only further development but also reimplementation to confirm its results.

Trong nghiên cứu [6] của Jostein Leirgulen và các cộng sự, nhóm tác giả  
đã đề xuất việc sử dụng các metadata như age, gender, pre-existing respiratory  
conditions, specified symptoms để xây dựng các thực nghiệm với các mô hình  
ML trên tập dữ liệu COUGHVID. Mặc dù kết quả còn khá khiêm tốn - 0.66  
accuracy với mô hình XGBoost - nhưng các tác giả đã đề xuất tới việc tăng  
cường độ chính xác của các giải pháp chuẩn đoán COVID-19 qua tiếng ho khi  
sử dụng các thông tin từ metadata. Hơn nữa, nghiên cứu cũng cung cấp insight  
khi phân chia dữ liệu segmentation (tránh việc giữ liệu segmentation tới từ cùng  
một subject nằm trong cả dữ liệu training, validation và testing splits) để có thể  
đưa ra những đánh giá khách quan nhất khi huấn luyện các mô hình ML. Các  
tác giả cũng chỉ ra việc sử dụng các kỹ thuật segmentation khác nhau cũng sẽ  
ảnh hưởng tới độ chính xác của các mô hình phân loại.  
Nghiên cứu [7] đưa ra thêm một hướng tiếp cận khác cho bài toán nhận  
diện COVID-19 qua tiếng ho bằng cách sử dụng phương pháp Stacked Ensemble  
Model. Qua đó tối ưu các ưu điểm của các mô hình khác nhau để tổng hợp kết  
quả chính xác nhất. Bộ dữ liệu COUGHVID đã được sử dụng để huấn luyện các  
mô hình. Kết quả tốt nhất mà nhóm tác giả đạt được 0.7986 accuracy và 0.797  
cho ROC-AUC.

In the research of Jostein Leirgulen et al., authors recommend using other metadata fields like age, gender, pre-existing respiratory conditions and specified symptoms to build ML experiments on COUGHVID dataset. Though results are not impressive which is 0.66 accuracy for XGBoost, this is still an improvement with the presence of metadata fields in the context of COVID-19 diagnosis through coughs. Moreover, the research also emphasizes the segmentation information to avoid the risk of leaking data into validation and test set, subsequently, it helps ML model return more objective results. Besides, different segmentation techniques also have different impact on the accuracy.

Research [7] presents another approach to the old COVID-19 cough diagnosis problem which comes along with Stacked Ensemble Model – utilizing different models’ strengths in one synthesized model with better result. In this paper, COUGHVID dataset is employed and then get the achievement of 0.7986 and 0.797 for accuracy and ROC-AUC respectively.

3 Proposed Lightweight Ensemble Models  
3.1 Proposed System Architecture  
Từ kiến trúc hệ thống được sử dụng trong [4] chúng tôi đề xuất thêm việc sử  
dụng các metadata dùng trong việc đưa ra chuẩn đoán. Hình 1 miêu tả chi tiết  
hệ thống đề xuất của chúng tôi.  
Đối với hệ thống của chúng tôi, khi người dùng nhập thông tin và thu âm  
tiếng ho, âm thanh ho sẽ được đưa qua module Cough Detector để đảm bảo âm  
thanh được thu vào là âm thanh ho. Sau đó, bản ghi âm thanh ho sẽ được xử lý  
và trích xuất thành các đặc trưng, kết hợp với các đặc trưng từ thông tin của  
người dùng để đưa vào mô hình phân loại. Các kết quả trả về từ các mô hình  
phân loại sẽ được tổng hợp và đưa ra kết quả nhận định cuối cùng.

According to the architectural system used in [4], we suggest employing additional metadata fields in diagnosing. Figure 1 presents our system in detail.

About our system, when users log their personal information and record their coughs, signals will be ensured that they are cough ones by the module Cough Detector. Subsequently, cough recordings will be processed and extracted into features, then combined with corresponding user information before being fed into a classification model. Returned results from the classification models will be synthesized and the final result is produced.

3.2 Proposed data preprocessing  
Cough segmentation: Sau khi quan sát các dữ liệu có trong hai bộ dữ liệu  
Coswara và COUGHVID, chúng tôi nhận thấy rằng audio tiếng ho sẽ chứa nhiều  
tiếng ho cách nhau bởi các khoảng nghỉ hoặc các âm thanh nhiễu như trong Fig.  
2. Vì vậy, việc cắt các audio thành các chunk nhỏ hơn chỉ chứa một tiếng ho  
mỗi chunk giúp cải thiện chất lượng dữ liệu huấn luyện cũng như loại bỏ các  
phần tạp âm trong dữ liệu. Chúng tôi sử dụng API pydub [8] để tiến hành việc  
segment các audio tiếng ho thành các chunk tiếng ho. Chúng tôi thực nghiệm  
với nhiều bộ tham số khác nhau trên dữ liệu từ Coswara và COUGHVID, từ đó,  
quan sát các file chunk được tạo ra để tìm ra bộ tham số tốt nhất.

Cough segmentation: After observing data in both Coswara and COUGHVID datasets, we realize that cough audios contain many low-energy or noisy intervals as a gap between two times of cough as depicted in Fig. 2. Therefore, separating cough audios into each chunk of cough helps improve training dataset’s quality and simultaneously eliminate noisy gaps. We exploit \textit{pydub API} [8] to implement audio segmentation into chunks of coughs. We also experiment with a variety of different parameters on Coswara and COUGHVID, then observe generated chunks to choose the most optimized group of parameters.

Với bộ tham số khi sử dụng API pydub gồm min\_silence\_len = 100, si-  
lence\_thresh = -30. Phần âm thanh trắng sẽ được thêm vào các cough chunk để  
đảm bảo rằng độ dài của các chunk là 1 - 1.3 giây. Qua quá trình thực nghiệm  
và quan sát trên kết quả thực nghiệm, chúng đề xuất quy trình trích xuất đặc  
trưng từ cough chunk như Fig. 3.  
By using \textit{pydub API}, we have the set of parameters: $min\\_silence\_len = 100, silence\\_thresh = -30$. Empty gaps will be added to the cough chunk in order to make sure that chunk duration belongs to the range of 1 – 1.3 seconds. Throughout the experimental process and observation, we recommend a feature extraction system illustrated in Fig. 3.

Sử dụng kỹ thuật segmentation giúp chúng tôi cả thiện số lượng dữ liệu hiện  
có cũng như giải quyết một phần nhỏ các dữ liệu tiếng ho có âm thanh nhiễu.  
Kết quả đạt được sau khi segmentation: Coswara với 4,465 cough audio ban đầu  
chúng tôi tạo ra 22,878 cough chunks, COUGHVID từ 20,072 cough audio đã  
tạo sinh 84,981 cough chunks.

Moreover, chúng tôi thiết kết thêm các thực nghiệm trên các kỹ thuật xử lý  
đặc trưng khác nhau như sử dụng standard scaler và SMOTE để đánh giá việc  
sử dụng đặc trưng nguyên mẫu được trích xuất từ dữ liệu so sánh với các đặc  
trưng đã qua xử lý có ảnh hưởng như thế nào tới bài toán nhận diện COVID-19  
qua tiếng ho.

Applying the segmentation technique helps enhance the number of available data and filter a fraction of noisy coughs. Here are the sizes of the two aforementioned datasets before and after segmentation: the number of Coswara recordings increases from 4,465 cough audios to 22,878 cough chunks, and the figure for COUGHVID rises from 20,072 cough audios to 84,981 cough chunks.

Moreover, we design experiments on different feature processing techniques such as Standard scaler and SMOTE in order to make a comparison between using original features extracted from the datasets and employing processed features for the COVID-19 coughs recognition issue.

Standard scaler: là kỹ thuật giúp điều chỉnh (chuẩn hoá) các giá trị của  
một vector đặc trưng dạng số học về một thang đo tiêu chuẩn có μ = 0 và  
σ = 1. Điều này giúp tránh sự rời rạc trong dữ liệu dạng số, tăng tính tương  
quan giữa các điểm dữ liệu với nhau. Việc chuẩn hoá dữ liệu trước khi tiến hành  
thực nghiệm trên các mô hình ML hoặc DL cũng là một bước quan trọng để đạt  
được kết quả tốt điều này cũng đã được nghiên cứu trong [9].

Standard scaler: a technique to normalize values of a numeric vector to satisfy a standard scale with μ = 0 and σ = 1. This technique prevents numeric data from discreteness and enhances correlation among data points. Data standardization prior to pushing them into ML or DL experiments also plays an essential role and contributes to substantial improvements (typically proved in [9]).

SMOTE: được phát triển bởi NV Chawla và các cộng sự trong [5], SMOTE  
là một phương pháp giúp cải thiện, khắc phục mất cân bằng nhãn trong dữ liệu.  
Phương pháp này tạo sinh mẫu dữ liệu nhằm gia tăng kích thước mẫu của nhóm  
thiểu số trong trường hợp xảy ra mất cân bằng mẫu. Để gia tăng kích thước  
mẫu, với mỗi một mẫu thuộc nhóm thiểu số ta sẽ lựa chọn ra k mẫu láng giềng  
gần nhất với nó và sau đó thực hiện tổ hợp tuyến tính để tạo ra mẫu giả lập.  
Metadata preprocessing: Chúng tôi tiến hành phân tích các metadata  
được cung cấp từ hai bộ dữ liệu Coswara và COUGHVID để đưa ra một thống  
kê cơ bản trong table ??. Với bộ dữ liệu Coswara, các field dữ liệu dạng số học  
không bị thiếu, các trường dữ liệu dạng chuỗi sẽ được chúng tôi fill null bằng  
giá trị ’unknown’ với cách điền khuyết này chúng tôi chấp nhận giá trị null là  
một giá trị hợp lệ vì các thông tin từ metadata là rất nhạy cảm bởi nó là các  
thông tin về bệnh lý của người bệnh nên chúng tôi quyết định không sử dụng  
các phương pháp dự đoán dữ liệu khuyết.

SMOTE: a worth-considering technique designed and developed by NV Chawla [5] assists researchers in dealing with label-imbalanced datasets. SMOTE will generate data points (oversampling) with the aim of increasing the size of the minority and resulting in label balance. In order to increase the number of samples, each sample of the minority class will need its $k$ closest neighbors and then carry out a linear combination, finally returning virtual samples.

Metadata preprocessing: We decide to analyze provided metadata from Coswara and COUGHVID, and pop up a statistics table \ref{}. We accept null values as valid information due to the sensitivity of patients’ situations, we will not employ any missing data imputation techniques to fulfill these null values. Therefore, with regard to fields with missing data, these fields will be filled by the value ‘unknown’.

3.3 Proposed model for Cough Detection  
Đối với giải pháp cho cough detection, sau khi thực hiện các thí nghiệm phương  
pháp máy học và học sâu khác nhau và đạt được các kết quả, chúng tôi đề xuất  
sử dụng mô hình có kết quả tốt nhất là mô hình kết hợp của hai mô hình mạng  
thần kinh tích chập (CNN) 1 chiều và 2 chiều, có tác vụ lần lượt là xử lý các  
thuộc tính của âm thanh ho Mel-frequency cepstrum (mfcc) và Mel-spectrogram.

Ensemble Model Mô hình đề xuất cho cough detection được chúng tôi xây  
dựng dựa trên việc kết hợp 2 mô hình CNN với nhau. Đầu tiên là 2 lớp đầu vào  
lần lượt cho các thuộc tính mfcc và mel-spectrogram. Tiếp theo, các thuộc tính  
của MFCC được đưa qua 4 lớp xử lý của mô hình thứ nhất. Trong mô hình thứ  
hai là các thuộc tính đại diện mel-spectrogram được xử lý song song với mô hình  
thứ nhất qua 16 lớp. Sau đó, kết quả ở lớp cuối của mỗi mô hình sẽ được kết  
hợp lại qua 1 lớp kết hợp và tiếp tục đưa qua 2 lớp ẩn để đến lớp đầu ra cuối  
cùng.

Regarding the Cough Detection component, after training ML and DL models as well as evaluating on their results, we recommend our highest-score model which is the combination of convolution neural network (CNN) 1-d and 2-d. In particular, Mel Frequency Cepstral Coefficient (MFCC) and Mel-spectrogram, which are cough sound features, are handled by 1-d CNN and 2-d CNN respectively.

Ensemble Model: We desire to boost the performance of Cough Detector models, therefore, we make an attempt to join two CNN models. At first, there are two input layers to receive MFCC and Mel-spectrogram features. Next, MFCC features are carried through four layers of the first model. In the second model - the sixteen-layer model, Mel-spectrogram features are processed in parallel with the first model. Then, both results of the two models are combined thanks to a blending layer before being pushed through the last two hidden layers.

3.4 Proposed model for COVID-19 Cough Recognition  
Như đề cập tại Mục 1, chúng tôi đề xuất việc sử dụng mô hình MobileNet để  
giúp việc triển khai áp dụng thực tế giải pháp mà chúng tôi phát triển được  
thuận lợi hơn nhờ vào các lợi thế của MobileNet [10]. Chúng tôi xây dựng một  
kiến trúc mô hình đặc biệt kết hợp giữa MobileNet và 1d- CNN để có thể sử  
dụng cùng lúc đặc trưng về tiếng ho và metadata. Ngoài MobileNet, chúng tôi  
cũng sử dụng thêm các mô hình ML như LightGBM, RF, SVM. Chúng tôi thực  
nghiệm trên bộ dữ liệu audio tiếng ho gốc và audio tiếng ho được cắt ra thành  
các chunk để đáng giá và so sánh phương pháp xử lý dữ liệu, trích suất đặc  
trưng cũng như tìm ra mô hình tốt nhất cho bài toán.

As discussed in Sec. 1, we highly recommend using MobileNet [10] in realistic deployment based on its advantageous characteristics. We construct a special version of MobileNet and 1-d CNN - LSTM with the purpose of utilizing both cough features and metadata fields. Alongside MobileNet, we also employ alternative ML models such as LightGBM, RF, and SVM. All of the listed models will be implemented on the original coughs dataset and cough chunks dataset so as to score and evaluate among them. Eventually, results will be compared to comprehend feature extraction techniques, feature processing techniques as well as models for the COVID-19 Cough Recognition methods.

Proposed stacking ensemble system Trong y khoa việc chuẩn đoán mẫu  
có dương tính hay âm tính khi thực hiện sàng lọc ban đầu cần rất thận trọng và  
chính xác. Ví dụ, mẫu âm tính với COVID-19 nhưng bị dự đoán là dương tính  
thì vẫn có thể phần nào chấp nhận được vì người bị chuẩn đoán nhầm cũng sẽ  
phải thực hiện những xét nghiệm chuyên sau đó nên việc chuẩn đoán nhầm này  
sẽ sớm được phát hiện. Tuy nhiên, với những bệnh nhân COVID-19 nhưng bị  
chuẩn đoán âm tính thì sẽ cực kỳ nguy hiểm vì sẽ dẫn đến việc lây lan dịch bệnh  
từ bệnh nhân này cho những người tiếp xúc xung quanh. Fig. 4 miêu tả chi tiết  
hệ thống tổng hợp dự đoán từ các phương pháp ML và DL (Level-1 models), kết  
quả dự đoán (xác suất của các nhãn) sẽ được sử dụng để train Logistic regression  
model (Level-2 model) để đưa ra kết quả dự đoán cuối cùng cho người dùng.

Proposed stacking ensemble system: In the medical system, positive or negative diagnosis in screening for disease requires thoroughness and accurateness. For instance, if a negative COVID-19 sample is predicted as positive, it is acceptable that this individual will still be further examined by other medical tests, as a result, the wrong diagnosis will soon be corrected. In contrast, a false positive sample will result in a serious spread of disease in the community because the patient would interact with others while having the virus. Fig. 4 demonstrates an elaborate system synthesizing estimations of different ML and DL remedies (Level-1 models). Subsequently, estimated probabilistic results will undergo Logistic regression model (Level-2 model) training in order to pinpoint whether an individual contracted COVID-19 or not.