TRƯỜNG ĐẠI HỌC BÁCH KHOA TP HCM

KHOA ĐIỆN – ĐIỆN TỬ

Bộ môn Viễn Thông

BÁO CÁO BÀI TẬP LỚN

MÔN: TRÍ TUỆ NHÂN TẠO

NHẬN DIỆN BỆNH NHÂN COVID QUA TIẾNG HO

Giảng viên: **TS. PHẠM VIỆT CƯỜNG**

SVTH: **MAI CHÍ BẢO**

Lớp: **L01**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2021

MỤC LỤC

DANH MỤC CÁC HÌNH VẼ

DANH MỤC CÁC TỪ VIẾT TẮT

CNN	Covolution Neural Network
CRNN	Convolutional Recurrent Neural Network
FCN	Fully - Convolutional Neural Network
LSTM	Long Short Term Memory networks
OCR	Optical Character Recognition
RNN	Recurrent Neural Network
VGG16	Visual Geometry Group from Oxford

CHƯƠNG 1. TỔNG QUAN VỀ ĐỀ TÀI

1.1 Giới thiệu về đề tài

Tính tới thời điểm này, đã có hơn 255 triệu ca nhiễm Covid – 19, trong đó có hơn 5 triệu ca tử vong đã cho thấy sự nghiêm trọng của loại dịch bệnh này lên con người. Từ đó xã hội đặt ra một dấu hỏi lớn về những biện pháp phòng chống, phát hiện sớm và ngăn ngừa sự lan rộng của biến chung vi rút Covid – 19.

Hiện nay đã có những phương pháp phòng ngừa hiệu quả như sự ra đời của các loại vắc xin Astrazeneca, Pfizer… hay các bộ xét nghiệm với độ chính xác cao chỉ trong thời gian ngắn. Tuy nhiên những biện pháp trên vì tính đắt đỏ, khó khăn, chậm chạp trong việc tổ chức và triển khai đặc biệt là ở những địa phương xa xôi và khó khăn.

Dẫn đến cần có một phương pháp gia tăng khả năng phòng vệ của xã hội, nhưng đồng thời cũng phải rẻ, dễ áp dụng cho mọi người, mọi nhà, ở mọi nơi. Vì thế trong môn học này, em chọn đề tài "Nhận diện bệnh nhân Covid – 19 qua tiếng ho" để nghiên cứu

1.2 Nội dung nghiên cứu

Đề tài "Nhận diện bệnh nhân Covid – 19 qua tiếng ho" gồm các nội dung cụ thể như sau:

- Tìm hiểu tổng quan về cơ chế ho, các đặc trưng âm thanh cần sử dụng
- Xây dựng mô hình nhận diện bệnh nhân
- Tìm hiểu các kỹ thuật training nhằm tăng khả năng nhận diện
- Thực hiện tiền xử lý dữ liệu, so sánh kết quả của các phương pháp xử lý dữ liệu

CHƯƠNG 2. KẾT QUẢ NGHIÊN CỨU

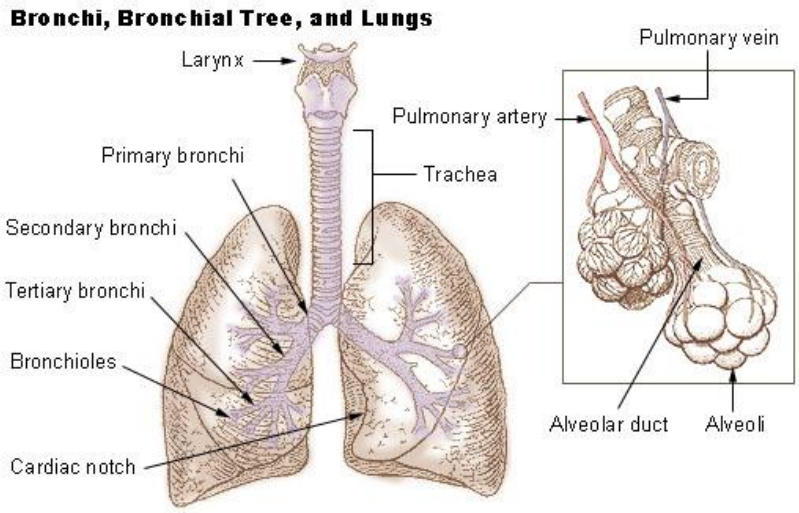
2.1 Tổng quan về ho và cơ chế tiếng ho

Ho là một cơ chế phản vệ của cơ thể nhằm ngăn không cho các cơ thể hít phải những vật lạ như bụi, vi trùng, kí sinh trùng… bằng một cơ chế nén và đẩy khí ra ngoài với tốc độ 160 km/h (nhanh hơn tốc độ của một quả bòng chày)

Để hiểu cơ chế của ho, trước tiên ta cần hiểu về cấu tạo của phổi. Chỉ đơn giản gồm hai phần chính là đường ống (khí quản, phế quản…) và túi khí.

Cơ chế của ho:

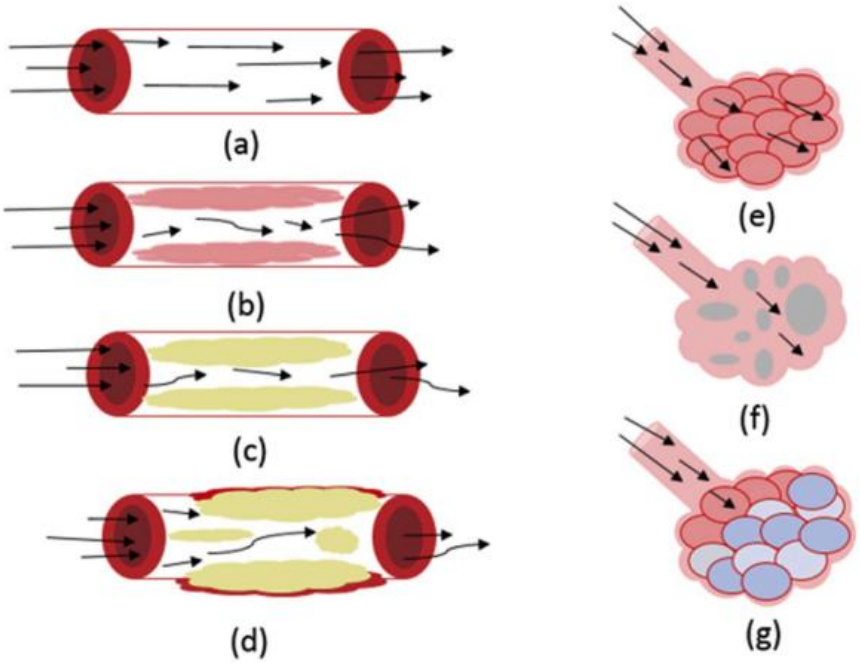
- Các tế bào cảm quan khi cảm thấy nhưng protein lạ (có trên bề mặt của vi rút, kí sinh trùng) sẽ gửi thông tin đến não để kích hoạt việc ho
- Các cơ hoành mở ra nhằm đưa một lượng khí nhất định vào lồng ngực
- Nắp thanh quản đóng lại. Đồng thời cơ xương sườn, cơ bụng ép chặt làm tăng áp suất trong lồng ngực
- Nắp thanh quản mở, khi này áp suất cao sẽ đẩy khi ra ngoài với tốc độ rất nhanh



Hình 2. 1 Cấu tạo của phổi

Từ cơ chế trên, chính việc cấu tạo của khí quản, hay túi khí gặp vấn đề làm thay đổi áp suất và dẫn đến gấy ra các tiếng ho khác nhau với các bệnh khác nhau

- Khí quản ổn định: Tiếng ho thường phân làm hai tiếng rõ ràng, thời gian ho cũng ngắn so với người có bệnh về phổi
- Hẹp khí quản: Áp suất cao dẫn đến tiếng ho rất ho, đồng thời với một lần hít vào sẽ có rất nhiều cơn ho theo sau. Khó khăn trong việc hít thở
- Giãn khí quản: Áp suất thấp dẫn đến tiếng ho đầu thường to, các tiếng ho sau nhỏ. Hít thở khò khè. Thường đi kèm với nhiều chất nhầy trong cổ họng
- Xơ cứng phổi: Phổi không thể co giãn lịnh hoạt dẫn thế không thể hít thở sâu, áp suất thấp làm cho năng lượng của tiếng ho nhỏ. Thời lượng của một đợt ho sẽ rất dài. Bệnh nhân thường có triệu chứng thiếu Oxy lên não và các cơ quan khác vì quá trình đưa Oxy vào máu bị ảnh hưởng
- Phổi có nước: "Nước" ở đây là chất thải của vi trùng trong túi phổi. Làm cho quá trình trao đổi Oxy không thể diễn ra bình thường và sẽ có các triệu chứng như ở bệnh xơ cứng phổi. Tiếng ho thường sẽ mạnh ở cơn ho đầu nhưng rồi năng lượng sẽ giảm dần ở các cơn ho sau



Hình 2. 2 a) Khí quản bình thường; b), c) Hẹp khí quản; d) Giãn khí quản; e) Túi khí bình thường; f) Xơ cứng phổi; g) Phổi chứa nước

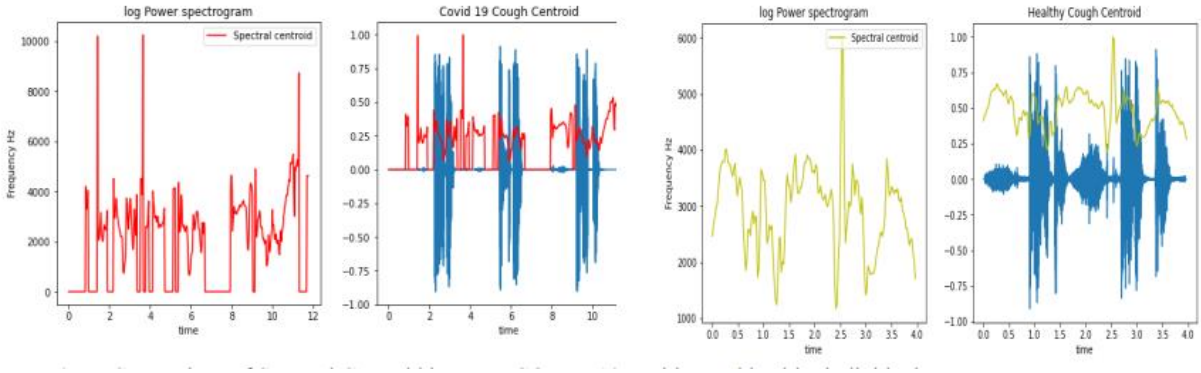
Bệnh nhân bị Covid – 19 thường sẽ có triệu chứng ho khan (không đờm nhầy) và khó thở. Từ những hiểu biết như trên về tiếng ho ở các loại bệnh, em thấy rằng năng lượng của tiếng ho, thời gian của từng cơn ho và cả đợt ho, âm thanh tiếng thở, số lần lặp lại của tiếng ho… đều mang lại thông tin cần thiết và từ đó em có thể đưa ra model và cách xử lý dữ liệu cho bài toán.

2.2 Đặc trưng âm thanh.

2.2.1 Spectral Centroid

Spectral Centroid: là dãy phổ trung bình (khác với trung vị) của dãy phổ bất kỳ. Những âm thanh trong và sáng hơn thường có dãy phổ centroid cao hơn.

Ví dụ như đồ thị kết quả phổ centroid của người bình thường sẽ cao hơn vì người nhiễm covid sẽ có biểu hiện ho khàn

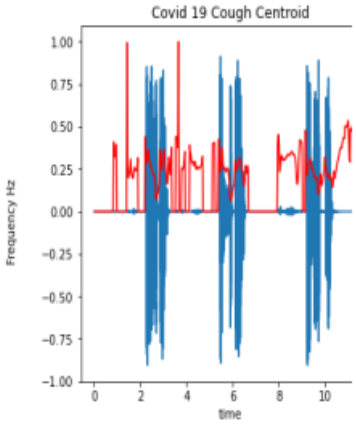


Hình 2. 3 So sánh Spectral Centroid trong tiếng ho của người khỏe mạng và bệnh nhân Covid

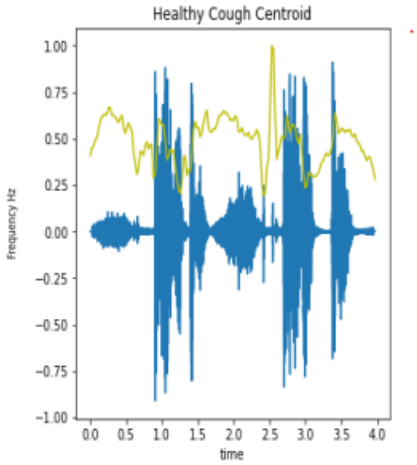
2.2.2 Spectral Bandwidth

Spectral bandwidth: là khoảng cách giữa tần số thấp và tần số cao trong khoảng thời gian liên tục. Để có thể nhận biết âm thanh một cách trực quan và đánh giá file âm thanh qua dạng ảnh ta áp dụng spectral bandwidth để training trong model

Ví dụ bên dưới đồ thị màu xanh là khoảng cách biểu thị giữa tần số cao và tần số thấp thấy bệnh nhân covid ho thường ngắn và dứt khoát hơn người thường



Hình 2. 4 Spectral Bandwidth trong tiếng ho của bệnh nhân Covid

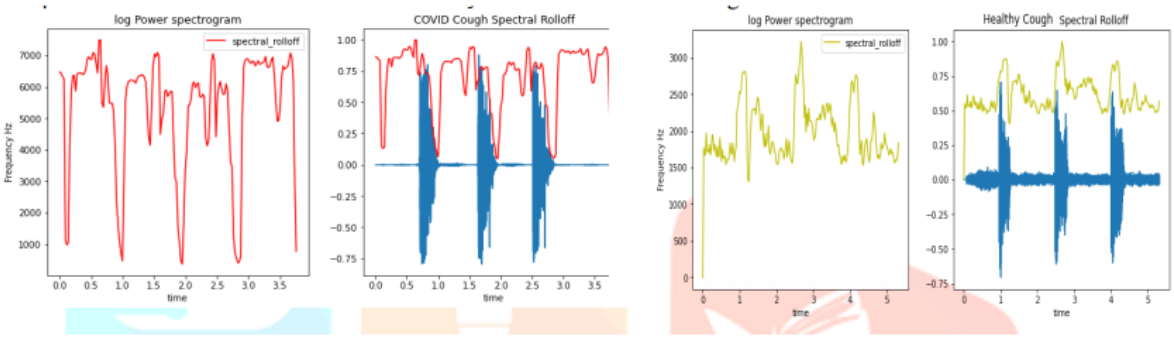


Hình 2. 5 Spectral Bandwidth trong tiếng ho của người khỏe mạnh

2.2.3 Spectral Roll - off

Spectral roll-off: những tần số dưới chứa 99% năng lượng của phổ

Ta có ví dụ khi phân bệnh nhân covid, tại dãy tần số thấp và cao của người mắc bệnh covid thì năng lượng của họ cao hơn nhiều so với người bình thường khi ho



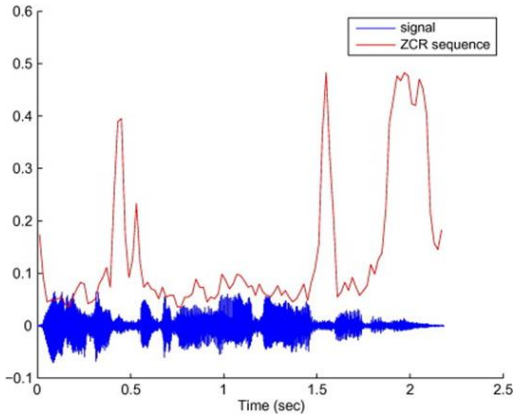
Hình 2. 6 So sánh Spectral Roll - off trong tiếng ho của người khỏe mạng và bệnh nhân Covid

2.2.4 Zero - Crossing Rate

Zero-Crossing Rate (ZCR) là tỉ lệ (tốc độ) thay đổi tín hiệu của tín hiệu trong khung. Nói cách khác, nó là số lần tín hiệu thay đổi giá trị, từ tích cực sang tiêu cực và ngược lại, chia cho độ dài của khung hình. Phương pháp đơn giản nhất để phân biệt giữa

giọng nói có lồng tiếng và không có lồng tiếng là phân tích ZCR. Một số lượng lớn các điểm giao nhau bằng 0 ngụ ý rằng không có dao động tần số thấp ở đây.

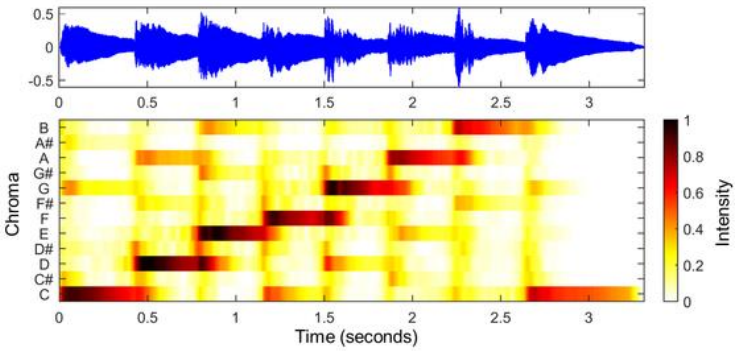
- ZCR được dùng cùng với mức năng lượng để phân biệt đâu có tiếng ho và đâu là khoảng lặng như 1 cơ chế attention giúp model làm việc tốt hơn.
- ZCR có thể được hiểu là thước đo độ ồn của tín hiệu, phản ánh đặc tính phổ của tín hiệu.
- Các giá trị của ZCR cao hơn đối với các phần nhiễu của tín hiệu, trong khi trong khung giọng nói, các giá trị ZCR tương ứng thường thấp hơn



Hình 2. 7 Ví dụ về ZCR

2.2.5 Chroma

Chroma (Sắc độ): Một cao độ âm nhạc có thể được chia làm 2 thành phần: độ cao của âm (pitch) và sắc độ (chroma). Một chroma là một thuộc tính của nốt. Trong dải tần âm thanh các tần số có thể được sắp xếp giống như một chiếc thang, với mỗi bậc thang là một giá trị cao độ và cứ sau 12 bậc thì sắc độ sẽ được lặp lại.



Hình 2. 8 Ví dụ thu được chroma của tín hiệu piano

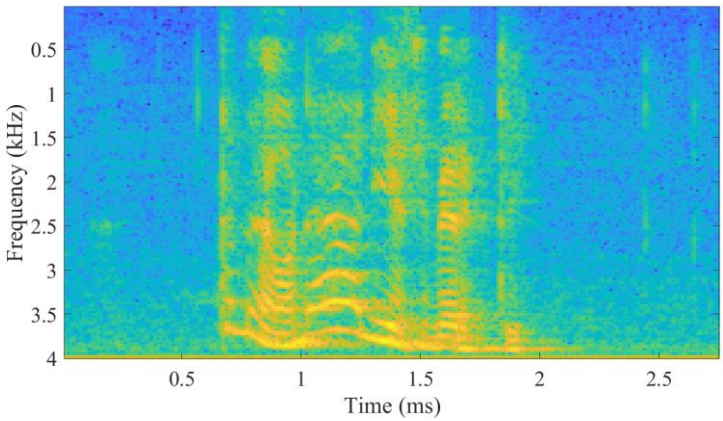
Xác định được các sắc độ của một âm thanh cho thấy mức độ cụ thể của các sắc độ trong âm sắc, từ đó dễ dàng hơn trong việc phân tích và xử lý tín hiệu âm thanh.

2.2.6 Mel Frequency cepstral coefficients và Mel Frequency Spectrogram

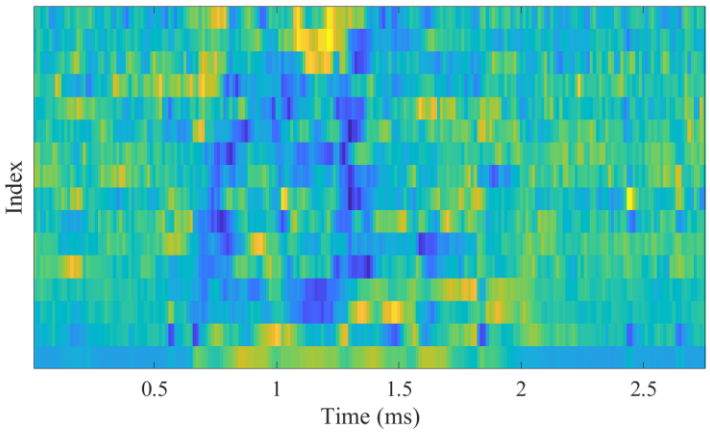
Cách trích xuất các feature giọng nói được sử dụng phổ biến nhất là Mel Frequency Cepstral Coefficents (MFCC), nó phổ biến nhất và mạnh mẽ do ước tính chính xác của nó về các tham số giọng nói và mô hình tính toán hiệu quả của lời nói.

Mỗi frame ta đã extract được 12 Cepstral features làm 12 feature đầu tiên của MFCC, feature thứ 13 là năng lượng của frame đó.

Mel frequency spectrogram thường có 39 đặc trưng, các feature có tính độc lập cao, ít nhiễu, đủ nhỏ để đảm bảo tính toán, đủ thông tin để đảm bảo chất lượng cho các thuật toán nhận dạng.



Hình 2. 9 Mel frequency spectrogram



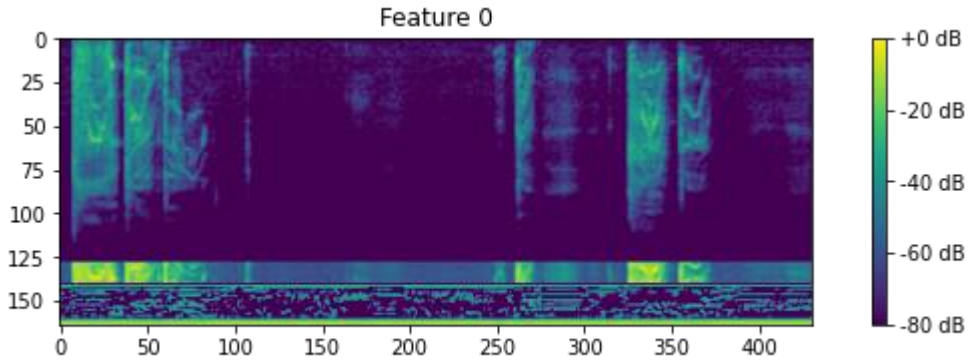
Hình 2. 10 MFCCs

2.2.7 Cách sắp xếp các đặc trưng

Trong đề tài này, em quyết định các đặc trưng của âm thanh được xếp theo cột dọc theo thứ tự ở dưới và có cùng độ dài thời gian để giữ lại được thông tin về thời gian của bức ảnh để model CRNN được làm việc tốt hơn:

- Mel Frequency Spectrogram
- Chroma
- Mel-frequency cepstral coefficients
- Zero-crossing rate
- Spectral centroid
- Spectral bandwidth
- Spectral roll off

Mỗi đặc trưng của data được xếp theo cột dọc với tổng cộng là 164 hàng, số cột tùy thuộc vào độ dài thời gian của từng file âm thanh.



Hình 2. 11 Dữ liệu sau khi được trích xuất các đặc trưng

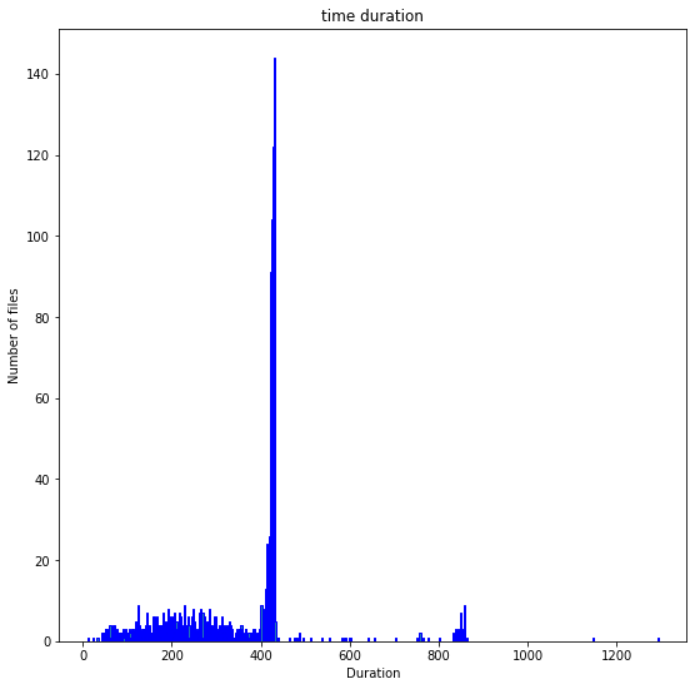
2.3 Bộ dữ liệu

Data của em có từ 3 nguồn: cuộc thi AICovidVN-115M Challenge (1194 file), COVID-19 Cough Recordings (170 file) và của em thêm vào (mỗi người 1 file âm tính, tổng 5 file). File được em dùng là file âm thanh có đuôi .wav.

Có một số file âm thanh có độ dài hơn một phút nhưng đa số file âm thanh có thời gian khoảng 20 giây trở xuống bao gồm tiếng ho và tiếng thở ở phía sau tiếng ho, không có tiếng nói.

Như vậy tổng dataset gồm 1369 dữ liệu. Trong đó có 481 dữ liệu positive (dương tính) và 888 dữ liệu negative (âm tính). Dữ liệu positive chỉ chiếm 35.14%. Lượng dữ liệu này sẽ được chia làm 3 phần: Test (15%), Validation (20%), Train (65%). Tuy nhiên, sau đó em còn thực hiện Data Augmentation trên tập file Train để so sánh.

Hiện tại COVID-19 đang rất nguy hiểm nên việc tiếp xúc với người bệnh rất khó khăn nên việc thu thập dữ liệu positive trở nên khó, dẫn tới việc dữ liệu positive ít hơn nhiều so với negative. Chính vì vậy nên dữ liệu hiện đang bị mất cân bằng.



Hình 2. 12 Số lượng file theo thời gian (thời gian đã được tính sau khi hopping và lấy sampling rate)

2.4 Tiền xử lý dữ liệu

2.4.1 Các phương pháp chung

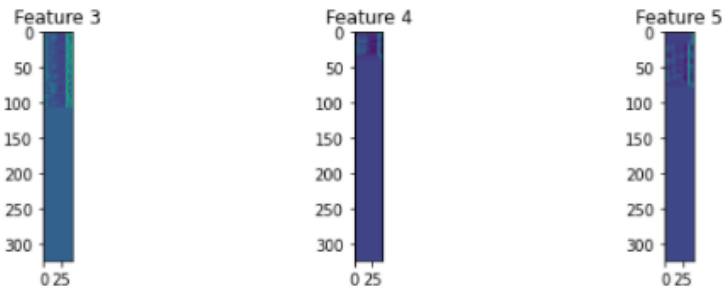
Resize (điều chỉnh kích thước) ảnh là một bước tiền xử lý quan trọng. Em chọn điều chỉnh kích thước nhỏ hơn 4 lần. Em nhận thấy số param trong model cần tính toán đã giảm đi gần 3 lần từ 48 triệu xuống còn 17 triệu param

Normalization (chuẩn hóa): để train một mô hình tốt rất khó khăn vì nó có cấu trúc gồm rất nhiều layer, trong quá trình training, phân bố dữ liệu qua các layer bị thay đổi rất nhiều.

- Ở đây em dùng standard vì min max đảm bảo tất cả các feature đều cùng 1 tỉ lệ chính xác (từ 0 - 1) nhưng nó rất nhạy cảm với các noise. Điều đó sẽ ảnh hưởng rất lớn đến cả mô hình.
- Model dễ bị biased trong trường hợp giá trị của một số feature cao đột biến so với các feature còn lại (giá trị của Mel chỉ khoảng 1 – 2, nhưng giá trị của ZCR thì đến hàng ngàn), giúp mô hình unbiased đối với các feature có giá trị cao.

Padding: Do các ảnh đưa vào có thời gian thực khác nhau, dẫn đến kích thước các feature map cũng khác nhau. Padding là việc thêm số 0 vào các biên của đầu vào để các feature đưa vào mô hình có kích thước bằng nhau. Em padding tất cả các file theo file âm thanh có độ dài lớn nhất.

Transpose: khi đưa model vào Tensorflow thì khi khai báo yêu cầu chiều rộng trước, chiều cao sau, ngược lại so với kích thước của feature trong model, do đó ta cần hoán vị các trục của mảng feature lại trước.



Hình 2. 13 Ảnh đã được tiền xử lý để đưa vào model

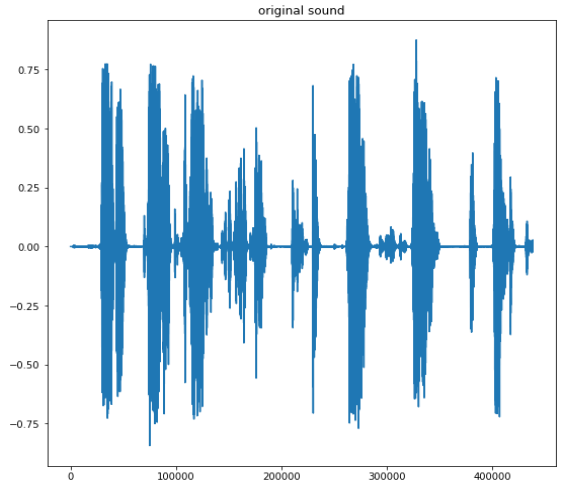
2.4.2 Data Augmentation

Trong 1369 dữ liệu, dữ liệu positive chỉ chiếm 35.14% (481 file) còn dữ liệu negative chiếm 64.86% (888 file).Khi train model sẽ có xu hướng nghiêng hẳn về bên phía negative, làm cho kết quả của model không còn đáng tin cậy. Có nhiều cách để cân bằng lại dữ liệu như thay đổi metric, đổi mô hình và phương pháp thực hiện, nhưng hiệu quả nhất là thêm vào dữ liệu positive hoặc giảm dữ liệu negative.

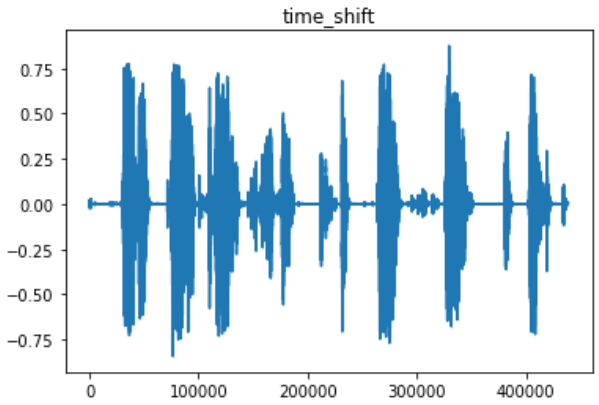
Hiện tại dịch bệnh COVID vẫn đang rất nguy hiểm nên việc tiếp cận F0 để có thêm dữ liệu positive là vô cùng khó khăn. Nếu ta chấp nhận giảm đi dữ liệu negative thì khi đó chỉ còn 962 file (50% positive và 50% negate) với số lượng ít như vậy việc training sẽ không đem lại kết quả như ta mong muốn. Chúng ta chỉ có thể tăng cường dữ liệu positive bằng các phương pháp phương pháp SMOTE, Mixup, Data synthesis, Random crop... Trong project này thì em đã tăng cường dữ liệu (Data Augmentation) trên tập Train bằng các phương pháp sau:

- Time Shift (dịch thời gian): âm thanh sẽ dịch sang trái hoặc sang phải 1 khoảng nhỏ.Trong project em đã dịch trong khoảng (-1000,1000)
- Adding background noise (tạo thêm tiếng ồn xung quanh): Tạo thêm tiếp ồn xung quanh để tránh trường hợp khi training mô hình lý tưởng hóa (chỉ có tiếng ho và tiếng thở). Còn khi test còn có cả âm thanh xung quanh.
- Stretching the sound (kéo dài âm thanh):Phương pháp này sẽ kéo dài âm thanh ra 1 chút.
- Changing Gain: tăng hoặc giảm âm lượng của âm thanh

Lý giải cho việc tại sao có nhiều phương pháp tăng cường dữ liệu như SMOTE, MIXUP, Random crop,... nhưng em chỉ dùng những phương pháp như dịch thời gian, kéo dài thời gian, tạo tiếng ồn. Nguyên nhân là do đối với phương pháp SMOTE nó sẽ dùng những dữ liệu lân cận có sẵn để tạo ra 1 dữ liệu mới. Còn với Mixup nó sẽ trộn dữ liệu âm thanh khác nhau lại với nhau để tạo nên dữ liệu mới. Với cách làm như vậy về mặt bản chất là sai vì nó tạo ra âm thanh ảo, không có trong cuộc sống thực, việc sử dụng những dữ liệu như vậy sẽ ảnh hưởng đến hoạt động thực tiễn của model nên em ko dùng những phương pháp.

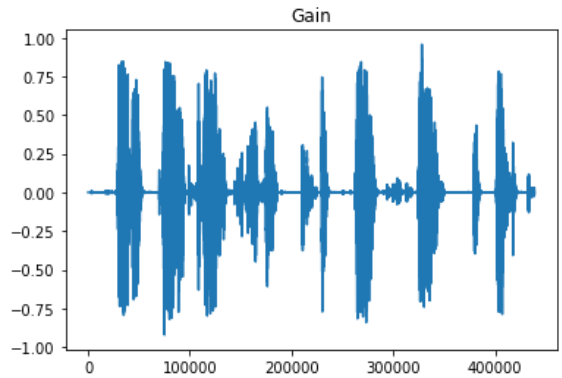


Hình 2. 14 Âm thanh gốc



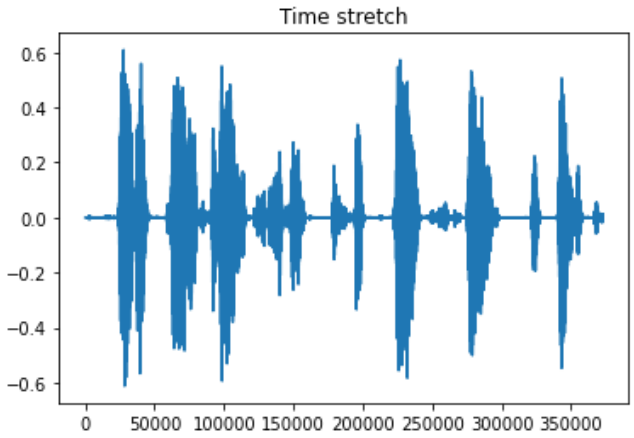
Hình 2. 15 Time Shift

Em dịch sang phải hoặc trái để tạo ra thêm 1 dữ liệu mới, nhìn vào hình thấy ảnh đã được dịch sang trái 1 chút so với ảnh gốc



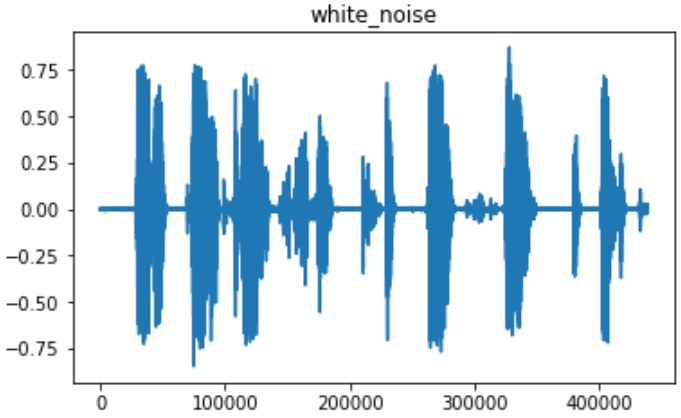
Hình 2. 16 Changing Gain

Phương pháp tăng cường, so với ảnh gốc biên độ của cường độ được tăng hơn



Hình 2. 17 Time Stretch

Phương pháp kéo dài thời gian, ở đây là hình minh họa cho việc thu gọn thời gian lại.



Hình 2. 18 Adding background noise

Tạo ra thêm những tiếng ồn xung quanh. So với ảnh gốc ban đầu, ảnh này có những chỗ đều nhau về cường độ âm thanh là do có tạp âm bên ngoài

Bằng cách phương pháp trên em đã tạo ra thêm 409 dữ liệu giả để cân bằng dữ liệu. Khi này posivite đã có 890 và negative có 888. Như vậy dữ liệu đã được cân bằng.

2.5 Mô hình Convolutional Recurrent Neural Network (CRNN) kết hợp lớp Attention

Từ những hiểu biết về tiếng ho đã nói ở mục 2.1 em thấy rằng để phân biệt tiếng ho cần dựa vào năng lượng của tiếng ho, thời gian của từng cơn ho và cả đợt ho, âm

thanh tiếng thở, số lần lặp lại của tiếng ho… Chúng đều là thông tin hình ảnh dạng chuỗi, mà model thích hợp nhất để sử dụng là CRNN

Mạng thần kinh tái phát liên tục (CRNN) là sự kết hợp của mạng CNN, RNN cho các tác vụ nhận dạng chuỗi dựa trên hình ảnh, như nhận dạng văn bản cảnh, OCR và âm thanh. Ngoài ra em còn kết hợp thêm lớp attention nhằm nâng cao khả năng nhận diện những thông tin cần thiết cho việc phân loại.

Kiến trúc mô hình được sử dụng ở đề tài này gồm 4 giai đoạn chính gồm mạng thần kinh tích chập (Convolution Neural Network), lớp attention, mạng hai chiều LSTM (Bidirectional Long-Short Term Memory Network), lớp phân loại (Fully Connected Layer)

2.5.1 Mạng thần kinh tích chập (Convolutional Neural Network)

Ở bài này em chọn lớp CNN của mô hình VGG16 (Xem kẽ 7 lớp Convolutional với max pooling và batch normalization) nhằm trích xuất các đặc trưng từ hình ảnh đầu vào.

Từ ảnh gốc khi qua các lớp tích chập sẽ tạo ra được các Feature Map và từ đó một chuỗi các Feature Vector sẽ được tạo ra gọi là Feature Sequence để đưa vào mạng RNN. Mỗi Feature Vector có thể tương ứng với một vùng Receptive Field ở ảnh gốc.

Mọi hình ảnh phải được chuẩn hóa về cùng chiều cao với độ dài có thể khác nhau.

2.5.2 Cơ chế Attention

Tuy Bi-LSTM có thể cải thiện được khả năng nhớ những thông tin phía trước nhiều hơn RNN, nhưng hiệu suất của nó vẫn ngày càng giảm khi thông tin phải nhớ tăng lên rất nhiều.

Cơ chế của attention là đánh trọng số cho từng thông tin đã training, khi có dữ liệu cần đánh giá ta cũng sẽ đánh trọng số cho chúng nên dựa vào 2 trọng số đã có để có thể dễ dàng tập trung những vị trí cần đánh giá

2.5.3 Mạng hai chiều LSTM (Bidirectional Long-Short Term Memory Network)

Bi-LSTM còn được gọi là Bi-directional LSTM sử dụng kết giữa 2 LSTM xử lý data theo chiều thuận hoặc chiều ngược. LSTM này một dạng đặc biệt của RNN

Những ưu điểm của Bi-LSTM mang lại giúp quyết bài toán:

- Thừa hưởng được đặc tính của RNN là dựa vào những đặc trưng của data ở phía trước để cập nhật kết quả thay vì đánh giá những đặc trưng riêng lẻ.
- Cải thiện được nhược điểm của RNN trong bài toán có chuỗi thông tin dài bằng cách loại bỏ những thông tin không cần thiết, đồng thời giữ lại hết những thông tin quan trong nếu cần (là Long - term memory)
- Do LSTM chỉ đánh giá data theo một chiều, nên để đánh giá data có hiệu quả tốt hơn nên Bi - LSTM sử dụng 2 LSTM kết hợp đánh giá data theo chiều thuận và theo chiều ngược lại của data.
- Ví dụ như bệnh nhân mắc covid có triệu chứng ho khan đi cùng phía sau là tiếng thở ngắn, khò khè. Nếu như ta dự đoán chỉ qua tiếng ho khàn ở phía trước mà đưa ra luôn kết quả thì có thể ta sẽ dự đoán sai. Do đó theo chiều ngược lại ta tiếp tục đánh giá tiếng thở ở phía sau khi có đầy đủ thông tin thì ta có thể đưa ra kết luận chính xác hơn

2.5.4 Lớp Fully Connected Layer

Lớp này giúp chuyển đổi mỗi khung hình (Per-frame Prediction) được tạo bởi mạng hai chiều LSTM ở trước thành chuỗi kết quả cuối cùng.

Ở đây em là phẳng ma trận đầu ra của Bi – LSTM thành một véc tơ có độ dài 20224 phần tử. Từ đó làm đầu vào của một lớp gồm 256 node trước khi đưa ra kết quả phân loại ở lớp cuối gồm 2 node. Đồng thời ở em còn thêm một lớp Batch normalization và đã đuược một số ưu điểm, sẽ được nêu ra ở các phần sau.

2.6 Quán trình huấn luyện (Training)

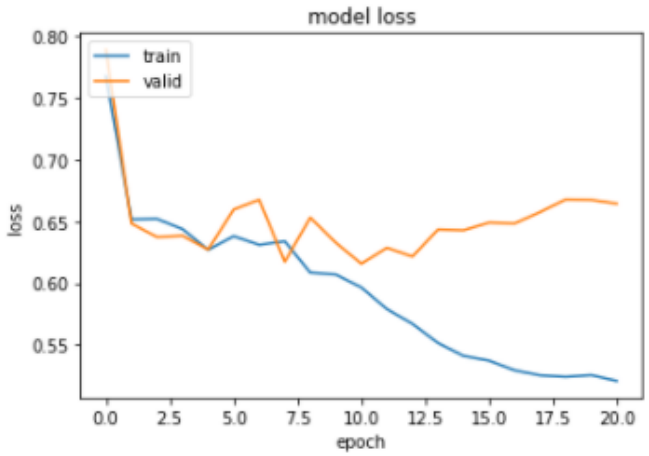
Chọn số epoch là 50 (có thể thay đổi tự do nhưng để giảm quá trình training, chọn số epoch phù hợp với early stopping)

Chọn số Batch size là 32, Là size trong mini-batch gradient descent, nghĩa là dùng bao nhiêu dữ liệu cho mỗi lần tính và cập nhật hệ số. Ở đây em chọn 32 vì 64, 128 rất lớn, nó làm model ổn định những sẽ dễ mất cực trị toàn cục. Không chọn 16 là vì thời gian train khá lâu và cũng không thực sự cần thiết.

Changing Learning rate: Learning rate quá nhỏ có thể dẫn đến quá trình training lâu dài có thể gặp khó khăn, ngược lại giá trị quá lớn có thể dẫn đến việc học quá nhanh hoặc quá trình training không ổn định. Để tăng sự chính xác em bắt đầu với learning rate tương đối cao và giảm dần trong quá trình training.

- Lợi ích: Tăng tốc độ đạt được cực tiểu, tránh được việc bỏ qua cực tiểu lý tưởng khi learning rate lớn. Tối ưu hóa được quá trình training và giải quyết được vấn đề "hỗn loạn" trong vùng cực tiểu.
- Hạn chế: Nếu lạm dụng Changing Learning rate sẽ dễ khiến model bị overfitting. Giảm learning rate sẽ làm cho quá trình học chậm hơn.
- Trong project thì em có learning rate khởi tạo 10^-3 trong 10 epoch đầu. sau đó trong 5 epoch tiếp theo em đã giảm learning rate xuống còn 10^-4 để dễ hội tụ tại cực trị, cuối cùng sẽ là 10^-5 trong quá trình còn lại. Tuy nhiên lần thay đổi thứ 3 này thường gây ra overfitting, cập nhật chậm nên em thường không lấy model ở số epoch này để dự đoán

Khi sử dụng model CRNN nhưng không có Batch Normalization ở lớp phân loại và train với original data: Ta thấy hàm loss thu được không tốt và xảy ra hiện tượng overfitting. Model bắt đầu overfitting từ epoch 10 (loss valid tăng nhẹ còn loss train giảm mạnh).

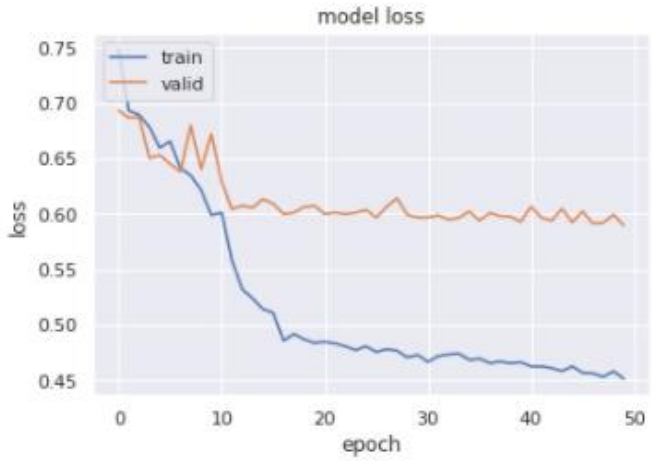


Hình 2. 19 Training model CRNN với original data khi chưa có Batch Normalization

Early stopping sẽ cho phép dừng training khi hiệu suất model không cải thiện, tiết kiệm thời gian training, cải thiện tính tổng quát của mạng. Tránh hiện tượng overfitting và giúp model đạt hiệu suất cao nhất. Trong project, em chọn early stopping là 10.

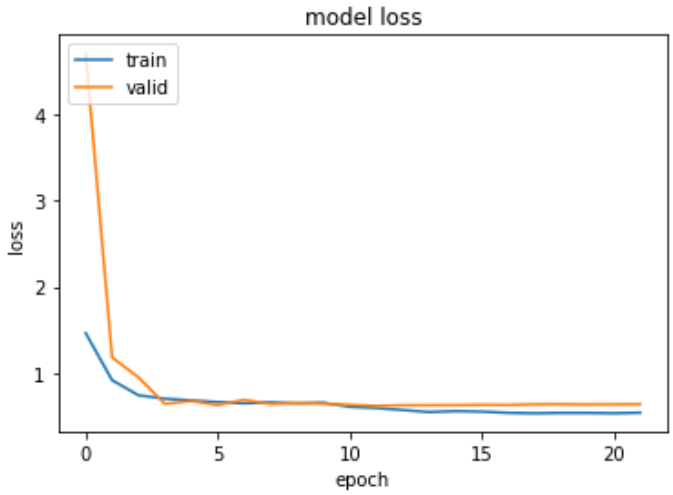
Lượng data khá ít và mất cân bằng giữa dương tính và âm tính (481 so với 888). Vì vậy để đạt được độ cân bằng và tăng hiệu quả của project thì Data Augmentation là cần thiết.

Khi train với model CRNN mà không sử dụng Batch Normalization ở output nhưng sử dụng data augment: Khi sử dụng Data Augmentation thì lượng dữ liệu được nhiều hơn, bài toán dễ huấn luyện hơn (dễ hội tụ), hiện tượng overfitting được hạn chế đi đáng kể (loss_train và loss_valid giảm nhẹ).



Hình 2. 20 Train với CRNN có sử dụng data augment nhưng không có Batch Normalization

Kết quả của quá trình training: Project sử dụng Batch Normalization ở output của model CRNN, thay đổi learning rate, chọn Batch size... một cách hiệu quả, hợp lý đã giúp cải thiện được độ ổn định của bài toán so với trước đây. Loss train và loss Valid sau khi hội tụ thì thay đổi không đáng kể.



Hình 2. 21 Train với CRNN có sử dụng data augment và có Batch Normalization

Khi sử dụng CRNN và Batch Normalization thì kết quả đã tốt hơn khi không sử dụng Batch Normalization và Data Augmentation, loss được cải thiện rõ rệt và ổn định hơn. Ngoài ra kỹ thuật changing learning rate giúp cho bài toán nhanh hội tụ và giảm số lượng epoch cần thiết cho việc hội tụ.

2.7 Các chỉ số đánh giá

Accuracy chỉ cho ta biết được bao nhiêu phần trăm dữ liệu được phân loại đúng mà không đưa ra được cụ thể mỗi lớp được phân loại thế nào, lớp nào được phân loại đúng nhiều hơn, dữ liệu lớp nào thường bị nhầm vào lớp khác. Ở trong đề tài này, trong tập dữ liệu chiếm 68% là negative, và model dự đoán negative cho tất cả các lần đoán thì accuracy cũng đạt 68%. Vì thế để đánh giá một mô hình tốt hay không, cần đánh giá tổng quát tất cả các chỉ số đánh giá sau.

Với một cách xác định một lớp là positive, precision được định nghĩa là tỉ lệ số điểm true positive trong số những điểm được phân loại là positive. Precision càng cao thì mô hình càng chính xác trong việc tìm các điểm.

Recall được định nghĩa là tỉ lệ số điểm true positive trong số những điểm thực sự là positive. Recall càng cao thì đồng nghĩa với việc bỏ sót các điểm positive thực sự càng thấp. Một mô hình phân lớp tốt là mô hình có cả Precision và Recall đều cao, và càng gần 1.

F1 – score là harmonic mean (trung bình điều hòa) của precision và recall. Do đó nó đại diện hơn trong việc đánh giá độ chính xác trên đồng thời precision và recall. F1 – score càng cao thì mô hình càng tốt.

ROC là đường cong biểu diễn khả năng phân loại của mô hình tại các ngưỡng threshold. Đường cong này dựa trên 2 chỉ số TPR (true positive rate) và FPR (false positive rate). Một mô hình hiệu quả khi có FPR thấp và TPR cao.

AUC là chỉ số đại diện cho khả năng phân lớp, nó cho biết mô hình có khả năng phân biệt giữa các lớp như thế nào. AUC là phần diện tích nằm dưới đường cong ROC. AUC càng cao thì mô hình càng tốt.

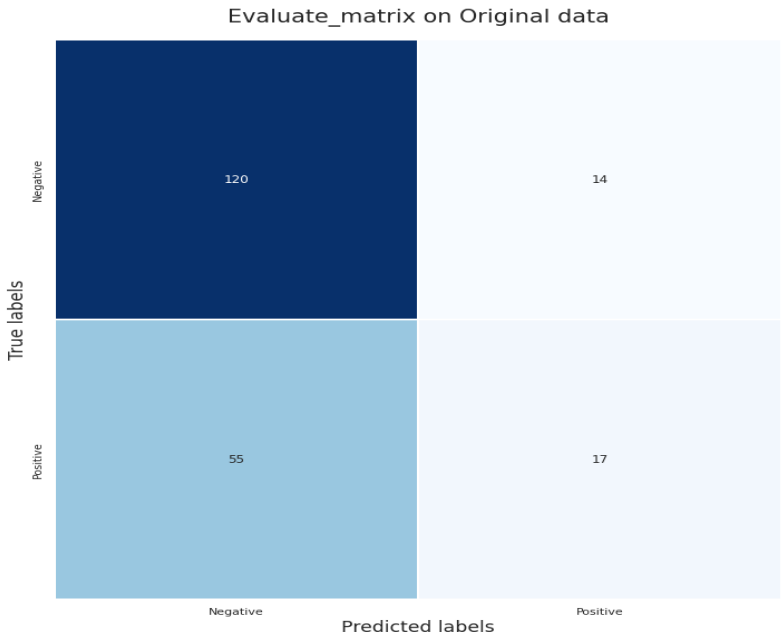
2.8 Đánh giá kết quả

Dưới đây là so sánh kết quả đánh giá giữa 2 data: Original data (dữ liệu gốc gồm 1369 file âm thanh chưa qua Data Augmentation) và Augmented data (Được thêm 409 file âm thanh đã qua Data Augmentation)

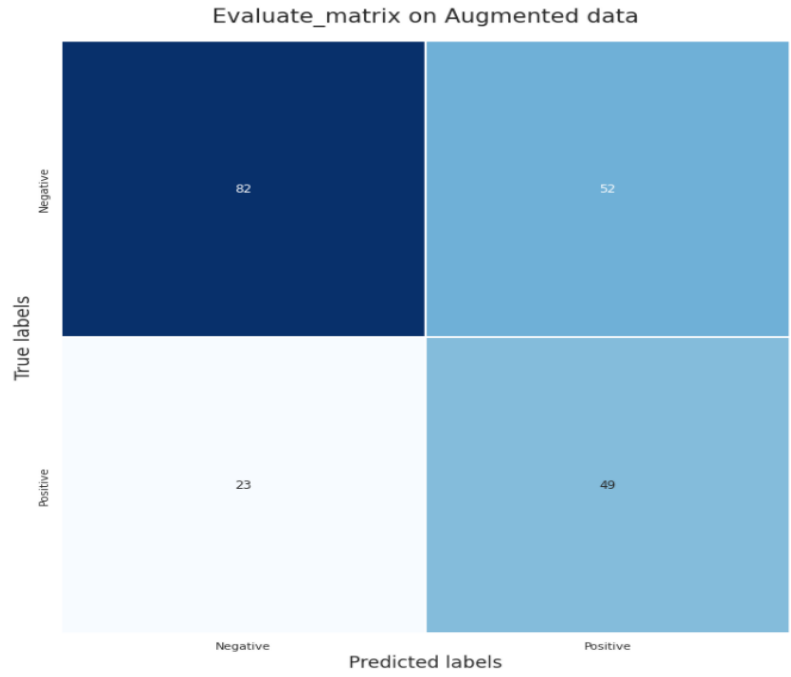
So sánh dựa trên confusion matrix

- Xét ở true negative (người âm tính thực sự), original dự đoán chính xác 120 người và sai 14 người, ở aug thì đoán trúng chỉ 82 và sai 52 người. Ở khoảng này thì original data thực tế làm việc tốt hơn
- Nhưng nhìn vào giá trị true positive (người bị dương tính thực sự), ta thấy ở original data dự đoán đúng chỉ 17 người mà sai đến 55 người, còn ở aug thì đoán đúng 49 người và sai chỉ 23 người. Chứng tỏ Augmented data hoạt động tốt hơn, ta có thể chấp nhận tăng việc đoán sai âm tính thực sự nhưng nhất định phải giảm tối thiểu việc bỏ sót người bị dương tính thực, vì sẽ nguy hiểm hơn nhiều nếu thả nhầm quá nhiều người bị dương tính ra khỏi bệnh viện

- Khi đánh giá mô hình trên AUC, có thể thấy con số khác biệt ít, chỉ tăng từ 66,8% lên 68,7%. Tuy nhiên về mặt ý nghĩa nó thay đổi rất lớn. Nó giúp ta phát hiện chính xác hơn và ít bỏ sót những người bị dương tính thực sự hơn rất nhiều.



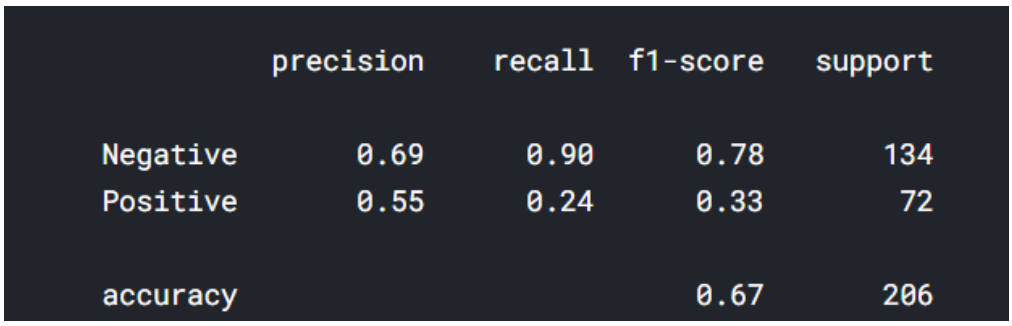
Hình 2. 22 Confusion matrix trên Original data



Hình 2. 23 Confusion matrix trên Augmented data

So sánh dựa trên F1, recall, precision, AUC, ROC curve:

- Xét trong original data thì precision và recall của Positive là 55% và 24%, độ chênh lệch nhau lớn, dẫn đến F1 – score nhỏ, chứng tỏ các điểm positive mà mô hình tìm được chính xác khá cao nhưng tỉ lệ bỏ sót các điểm positive cũng rất cao. F1 – score bằng 33% cho ta thấy khả năng phân loại chỉ ở mức kém.
- Xét trong Positive của augmented data thì precision = 49% và recall = 68%, lệch nhau không nhiều, F1 – score = 57% cũng nằm ở mức trung bình, chứng tỏ các điểm positive tìm được chính xác khá cao và tỉ lệ bỏ sót các điểm positive thấp, tốt hơn ở original data rất nhiều. Mô hình có recall tăng đồng nghĩa với việc bỏ sót dương tính ít hơn.
- Precision, recall và negative ở original data là (69%; 90%; 78%) khác với ở augmented data (78%; 61%; 69%), cho thấy độ chính xác khi tìm negative ở original thấp hơn nhưng ngược lại, tỉ lệ bỏ sót negative của augmented lại cao hơn rất nhiều.
- Precision, recall và F1 – score của Negative ở cả 2 data đều cao hơn Positive rất nhiều chứng tỏ mô hình phát hiện tốt ở những người âm tính hơn là những người dương tính với Covid.
- Khi đánh giá mô hình trên AUC, có thể thấy con số khác biệt ít, chỉ tăng từ 66,8% lên 68,7%. Tuy nhiên về mặt ý nghĩa nó thay đổi rất lớn. Nó giúp ta phát hiện chính xác hơn và ít bỏ sót những người bị dương tính thực sự hơn rất nhiều.

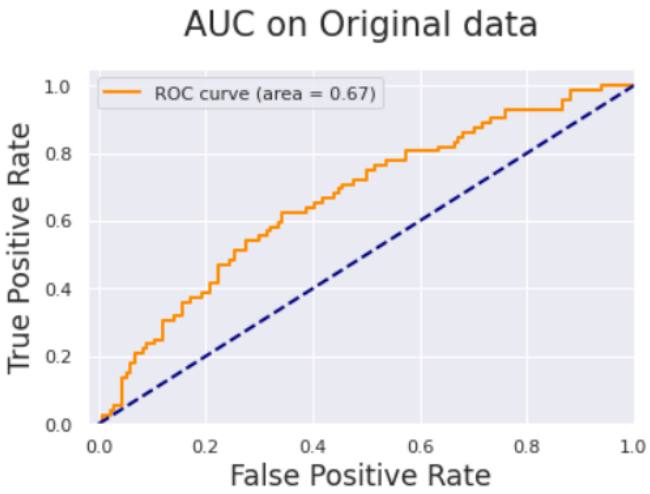


	precision	recall	f1-score	support
Negative	0.69	0.90	0.78	134
Positive	0.55	0.24	0.33	72
accuracy			0.67	206

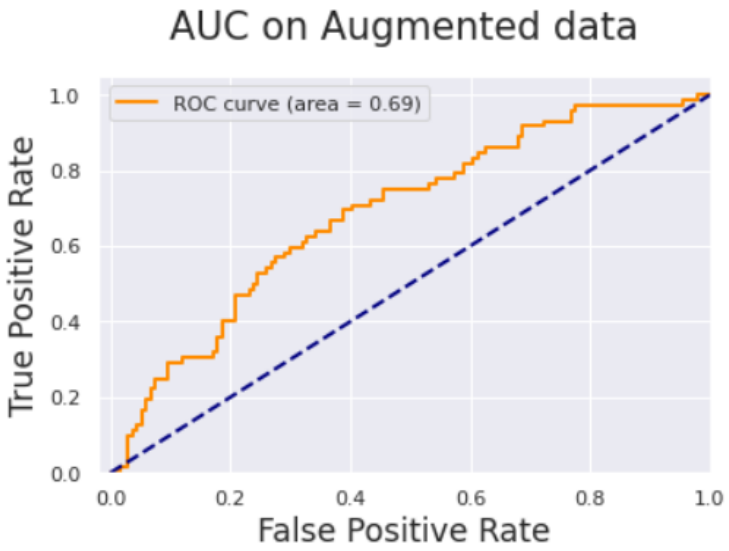
Hình 2. 24 Kết quả với F1, precision, recall, accuracy trên original data

	precision	recall	f1-score	support
Negative	0.78	0.61	0.69	134
Positive	0.49	0.68	0.57	72
accuracy			0.64	206

Hình 2. 25 Kết quả với F1, precision, recall, accuracy trên augmented data



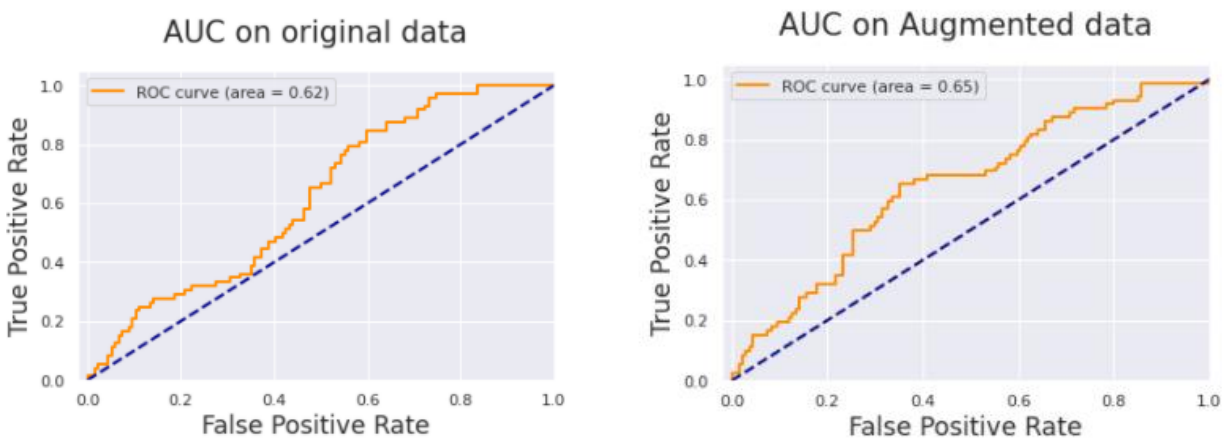
Hình 2. 26 AUC trên original data



Hình 2. 27 AUC trên augmented data

So sánh về kết quả khi với model không có lớp Batch Normalization:

- Khi dùng original data (không dùng Batch Normalization) thì accuracy thấp hơn so với khi đã dùng augment data (không dùng Batch Normalization), AUC (Khu vực dưới đường cong ROC) khi có augment cao hơn (0.65 so với 0.62). Tác dụng của việc data augment rất lớn, giúp cải thiện được hiệu suất của project.



Hình 2. 28 Kết quả AUC (không dùng Batch Normalization) cho model CRNN khi chưa augment và khi đã augment

TÀI LIỆU THAM KHẢO

[1] P. Bagad *et al.*, "Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds," no. Ml, 2020, [Online]. Available: http://arxiv.org/abs/2009.08790.

[2] S. B. Brosnahan, A. H. Jonkman, M. C. Kugler, J. S. Munger, and D. A. Kaufman, "Covid-19 and respiratory system disorders current knowledge, future clinical and translational research questions," *Arterioscler. Thromb. Vasc. Biol.*, no. September 2020, pp. 2586–2597, 2020, doi: 10.1161/ATVBAHA.120.314515.

[3] T. M. I. T. Faculty *et al.*, "COVID-19 Artificial Intelligence Diagnosis," 2020.

[4] A. Imran *et al.*, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *Informatics Med. Unlocked*, vol. 20, pp. 1–27, 2020, doi: 10.1016/j.imu.2020.100378.

[5] N. Melek Manshouri, "Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study," *Cogn. Neurodyn.*, vol. 0123456789, 2021, doi: 10.1007/s11571-021-09695-w.

[6] M. Melek, "Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound," *Neural Comput. Appl.*, vol. 33, no. 24, pp. 17621–17632, 2021, doi: 10.1007/s00521-021-06346-3.

[7] E. A. Mohammed, M. Keyhani, A. Sanati-Nezhad, S. H. Hejazi, and B. H. Far, "An ensemble learning approach to digital corona virus preliminary screening from cough sounds," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-95042-2.

[8] P. Mouawad, T. Dubnov, and S. Dubnov, "Robust Detection of COVID-19 in Cough Sounds," *SN Comput. Sci.*, vol. 2, no. 1, pp. 1–13, 2021, doi: 10.1007/s42979-020-00422-6.

[9] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 Detection in Cough, Breath and Speech using Deep Transfer Learning and Bottleneck Features," 2021, [Online]. Available: http://arxiv.org/abs/2104.02477.

[10] C. Pelaia, C. Tinello, A. Vatrella, G. De Sarro, and G. Pelaia, "Lung under attack by COVID-19-induced cytokine storm: pathogenic mechanisms and therapeutic implications," *Ther. Adv. Respir. Dis.*, vol. 14, pp. 1–9, 2020, doi: 10.1177/1753466620933508.

[11] N. S. Rajput, "Winter Semester 2020-21 CSE3031 - Artificial Intelligence - Project Review – 3 Team Members : TITLE : Keywords : Type of work :," 2020.

[12] G. Rudraraju *et al.*, "Cough sound analysis and objective correlation with spirometry and clinical diagnosis," *Informatics Med. Unlocked*, vol. 19, p. 100319, 2020, doi: 10.1016/j.imu.2020.100319.

[13] P. Sadhukhan, M. T. Ugurlu, and M. O. Hoque, "Effect of covid-19 on lungs: Focusing on prospective malignant phenotypes," *Cancers (Basel).*, vol. 12, no. 12, pp. 1–17, 2020, doi: 10.3390/cancers12123822.

[14] S. Tian, W. Hu, L. Niu, and H. Liu, "Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- company ' s public news and information website . Elsevier hereby grants permission to make all its COVID-19-r," no. January, 2020.