

Nhận diện COVID-19 qua tiếng ho

Nguyễn Thành Trung^{a, b}, Hoàng Đình Toàn^{a, b}, and Đào Minh Triết^{a, b}

^aĐại học Bách Khoa - Đại học Quốc gia Thành phố Hồ Chí Minh

^b{trung.nguyendx, toan.hoangcsgl, triet.dao.04042000}@hcmut.edu.vn

August 21, 2021

TÓM TẮT NỘI DUNG

Với sự lây lan nhanh của COVID-19 hiện nay, việc xét nghiệm xem một người có dương tính hay không là rất cần thiết để truy vết và cách ly để thực hiện mục tiêu kép của Đảng và Nhà nước đề ra. Tuy nhiên, việc này vẫn chưa thực sự thuận tiện về nhiều mặt cho người xét nghiệm cũng như người được xét nghiệm vì nhiều lý do. Báo cáo cuối kì - học kì 202 này trình bày một mô hình học máy Light GBM để có thể nhận diện được người dương tính với COVID-19 chỉ qua một đoạn ghi âm tiếng ho của họ. Mô hình triển khai được áp dụng để tham gia cuộc thi nhận diện COVID-19 qua tiếng ho của AICOVIDVN và đạt được thứ hạng 16. Toàn bộ mã nguồn cũng như các tệp tin liên quan có thể được tìm thấy tại: <https://github.com/dee-ex/EE3063-SEM202-FINAL-PROJECT>.

Từ khoá: COVID-19, AICOVIDVN, học máy, quang phổ Mel, MFCCs, sắc độ, SVM SMOTE, Light GBM

1. GIỚI THIỆU

Hiện nay, dịch bệnh COVID-19 ở Việt Nam đang diễn biến rất phức tạp. Số ca nhiễm mỗi ngày luôn được tính ở đơn vị ngàn và vẫn chưa có dấu hiệu giảm. Một trong điều then chốt trong công cuộc phòng và chống dịch hiện nay đó là việc xét nghiệm cho người dân. Cụ thể, để được xét nghiệm nhanh kháng nguyên (hay còn gọi là kiểm tra nhanh COVID-19), người dân đến các bệnh viện và trung tâm y tế ở quận, huyện, thành phố. Nếu muốn xét nghiệm khẳng định SARS-CoV-2 (hay còn gọi là phương pháp xét nghiệm RT-PCR) thì đến các phòng xét nghiệm của các đơn vị được Bộ Y tế cho phép xét nghiệm khẳng định SARS-CoV-2. Về mức giá xét nghiệm, Sở Y tế cho biết, xét nghiệm bằng phương pháp RT-PCR bằng mức giá thanh toán bảo hiểm y tế theo quy định tại Công văn 4356/2020 của Bộ Y tế về việc hướng dẫn mức giá thanh toán chi phí thực hiện xét nghiệm COVID-19. Xét nghiệm nhanh kháng nguyên: bằng giá dịch vụ được quy định tại Phụ lục đính kèm Thông tư số 13/2019 và Thông tư số 14/2014 của Bộ Y tế là 238.000 đồng/mẫu [1].

Chúng tôi thấy rằng, với tình hình hiện tại, việc xét nghiệm không chỉ khó khăn về tài chính mà còn là sự thiếu hụt về dụng cụ. Liệu có cách nào để thực hiện việc này hiệu quả hơn? Các nhà nghiên cứu trên thế giới đã chỉ ra rằng có thể phát hiện COVID-19 qua tiếng ho, qua điện thoại di động bằng trí tuệ nhân tạo (AI) với độ chính xác trên 90% [2]. Trên những cơ sở đó, chúng tôi triển khai một mô hình học máy giúp phân loại liệu một người có dương tính với COVID-19 hay không.

2. NGHIÊN CỨU LIÊN QUAN

Trong bài báo được xuất bản trên IEEE Journal of Engineering in Medicine and Biology [3], những nhà nghiên cứu đã huấn luyện một mô hình với khoảng một vạn mẫu tiếng ho. Mô hình này khi đem đi thử nghiệm với những tiếng ho của người đã được xác nhận là dương tính với COVID-19 đã cho kết quả nhạy đến 98.5%. Nhóm nghiên cứu trên cũng đã và đang làm việc để kết hợp mô hình này vào một ứng dụng thân thiện với người dùng. Nếu được FDA chấp thuận và áp dụng trên quy mô lớn thì có thể là một công cụ sàng lọc trước miễn phí, tiện lợi để xác định những người có khả năng dương tính với COVID-19.

Theo như chúng tôi tìm hiểu, chưa thực sự có một cơ sở khoa học rõ ràng nào cho điều này, nhưng những kết quả đã đạt được từ [3] cho thấy phương pháp nhận diện COVID-19 qua tiếng ho là chuyện rất khả thi. Báo cáo này sẽ đóng góp thêm những kết quả về việc áp dụng học máy trong việc nhận diện người nhiễm COVID-19 qua tiếng ho ở Việt Nam.

3. TẬP DỮ LIỆU

Dự án cộng đồng AICOVIDVN tổ chức cuộc thi “AICV-115M Challenge” đã thu thập và công bố một tập dữ liệu đã được dán nhãn gồm 4.504 mẫu dữ liệu kèm thêm một tập dữ liệu gồm 1233 mẫu dữ liệu chưa được dán nhãn.* Tập dữ liệu chưa được dán nhãn là tập dữ liệu kiểm tra để nộp lên ban tổ chức để tính điểm và vì thế trong báo cáo này, tập dữ liệu này sẽ không được sử dụng để đánh giá do sự thiếu sót thông tin, mà chỉ làm việc với tập dữ liệu được dán nhãn. Để miêu tả về tập dữ liệu, mỗi mẫu dữ liệu trong đó là một tệp tin âm thanh có 1 cho tới 4 tiếng ho. Dù ban tổ chức đã rất cố gắng, nhưng vẫn tồn tại những một số mẫu không có tiếng hoặc tiếng không rõ. Sau khi loại bỏ những mẫu không đạt yêu cầu nhờ thông tin được cuộc thi cung cấp, số lượng mẫu còn lại là 4.068. Trong số những mẫu được giữ lại, có 3399 mẫu được dán nhãn là âm tính và 669 được dán nhãn là dương tính, tỷ lệ giữa hai nhãn khoảng 8/2 (chính xác hơn là 8.4/1.6). Sau đó, tập dữ liệu này được chia ra thành 2 tập là huấn luyện và thử nghiệm theo một tỷ lệ xấp xỉ một tỷ lệ phổ thông là 8/2 gồm 3268 dữ liệu huấn luyện và 800 dữ liệu kiểm tra. Tập huấn luyện gồm có 2735 mẫu âm tính và 533 mẫu dương tính. Cũng giống như nhiều tập dữ liệu về y tế, đây là một tập dữ liệu không cân bằng giữa hai nhãn, dẫn tới việc xây dựng và đánh giá mô hình sẽ khó khăn hơn đôi chút so với những tập dữ liệu cân bằng [4]. Một số phương pháp giải quyết hiện tượng dữ liệu mất cân bằng sẽ được đề cập trong Phần 5.1.2.

4. TRÍCH XUẤT ĐẶC TRƯNG

So với dữ liệu hình ảnh - một loại dữ liệu quen thuộc với đại đa số những người làm về học máy/học sâu, âm thanh là một kiểu dữ liệu khá đặc biệt. Nếu mỗi điểm ảnh của một bức hình được coi là một đặc trưng, ta có thể gom tất cả các điểm ảnh có được thành một véc-tơ đầu vào và trong rất nhiều trường hợp chuyện này rất đối bình thường và cũng không kém phần hiệu quả. Còn âm thanh thì không đơn giản như vậy, dữ liệu ta có là một mảng dài với mỗi phần tử là biên độ của sóng được lấy mẫu tại một thời điểm nhất định. Tần số lấy mẫu, tức nghịch đảo của khoảng thời gian để ta ghi lại biên độ giữa lần thứ t và $t + 1$ thường sẽ là 44.1 KHz (chất lượng tốt) [5]. Việc đưa tất cả những giá trị có được của một tệp âm thanh vào một mô hình học máy/học sâu trực tiếp để nhận diện cấu trúc thường không cho ra những kết quả khả quan. Thay vào đó, một số những phương pháp trích xuất đặc trưng quan trọng của âm thanh đã được áp dụng. Chúng tôi đã thử nghiệm qua một số những đặc trưng thường được sử dụng, tuy vậy không phải đặc trưng nào cũng đóng góp được nhiều và tích cực cho bài toán này. Những mục sau của Phần 4 sẽ tập trung giới thiệu sơ bộ 3 đặc trưng có ảnh hưởng lớn tới mô hình của chúng tôi: **Quang phổ Mel**, **MFCCs** và **Sắc độ**. Về phần trích lọc các đặc trưng này trong quá trình lập trình, thư viện chúng tôi sử dụng để xử lý tác vụ này là librosa[†] - một thư viện mạnh mẽ để xử lý tín hiệu âm thanh.

4.1 Đặc trưng Quang phổ Mel

Quang phổ, hay còn được gọi là ảnh phổ, là một cách trực quan để biểu thị độ lớn của tín hiệu theo thời gian ở các tần số khác nhau. Nếu phổ chỉ có thể cho ta giá trị mức năng lượng của các tần số khác nhau thì quang phổ còn giúp ta xem được các mức năng lượng đó thay đổi như thế nào theo thời gian. Hiểu một cách trực quan, quang phổ là phổ theo từng thời điểm.

Mel là một thang đo tần số được đưa ra để chuyển thang đo tần số thông thường về thang đo tần số phù hợp với con người do Volkmann và Newman [6] đề xuất. Điều này dựa trên thực tế về khả năng phân biệt các tín hiệu tần số của con người, khi chúng ta giỏi hơn trong việc phát hiện sự khác biệt ở tần số thấp hơn là tần số cao. Ví dụ, giữa 500Hz và 1000Hz thì con người bình thường dễ dàng phân biệt được, nhưng giữa 10.000Hz và 10.500Hz thì lại không như thế dù khoảng cách đều là 500Hz. Thang đo Mel được triển khai giúp cho máy học có thể mô phỏng lại cách nhận thức của con người với tín hiệu.

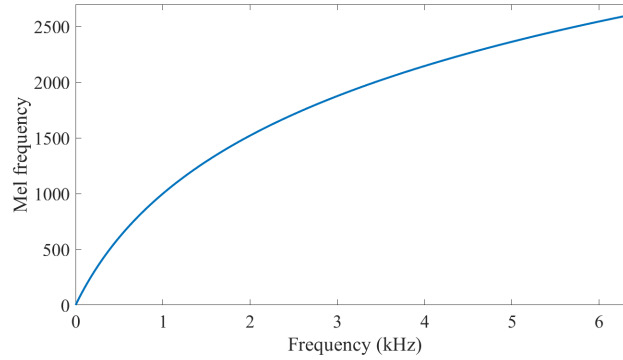
Công thức chuyển đổi từ thang đo Hz sang Mel [7]:

$$f_{\text{Mel}} = 2595 \log_{10} \left(1 + \frac{f_{\text{Hz}}}{700} \right) \quad (1)$$

từ Phương trình (1) hoặc qua Hình 1, ta có nhận xét đó là các tín hiệu ở tần số thấp có sự chênh lệch lớn hơn

*AICovidVN 115M Challenge: Covid Cough Detection Challenge (Final Round), đường dẫn: <https://aihub.vn/competitions/22#participate>.

[†]librosa, đường dẫn: <https://librosa.org/doc/latest/index.html>.



Hình 1. Thang đo Mel so với thang tần số Hz thông thường [7].

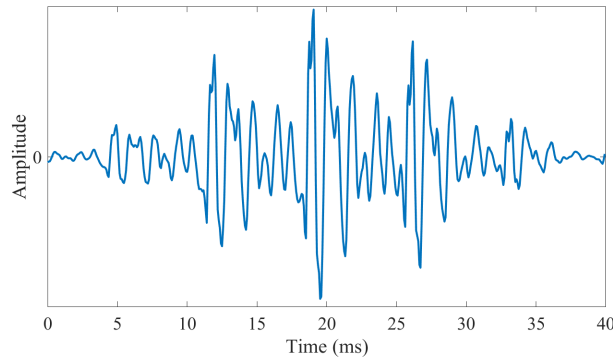
khi chuyển sang thang đo Mel. Ngược lại ở tần số cao, sự chênh lệch này được thu hẹp lại khi chuyển sang thang đo Mel. Điều này gần như tương tự với khả năng nhận biết sự khác biệt giữa các tín hiệu tần số khác nhau của con người.

Việc áp dụng thang đo Mel vào trích xuất quan phổ sẽ cho ta một quang phổ Mel, giúp tăng tối đa mức thông tin mà một quan phổ có thể cung cấp.

4.2 Đặc trưng Mel-frequency cepstral coefficients (MFCCs)

Đặc trưng này được trích xuất và ứng dụng trong nhiều nghiên cứu về nhận diện giọng nói [8]. Với những nghiên cứu về tín hiệu âm thanh, đặc trưng MFCCs luôn là một đặc trưng nên thử nghiệm. Đúng như tên gọi của nó, MFCCs sẽ cho ra các **hệ số** (coefficients) của **cepstral** (một cách chơi chữ của spectral) từ thang **tần số Mel** (Mel-frequency Phần 4.1).

Để có một cepstral từ một tín hiệu âm thanh, bước đầu sẽ phải chuyển đổi từ miền thời gian (Hình 2) sang miền tần số bằng biến đổi Fourier[‡]. Sau bước biến đổi này, ta có **phổ** (spectrum) (Hình 3).

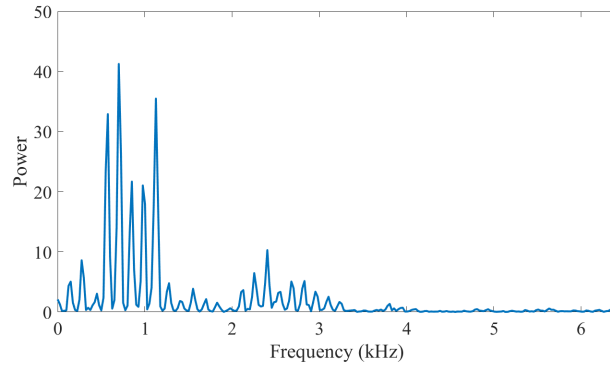


Hình 2. Một tín hiệu âm thanh trên miền thời gian [7].

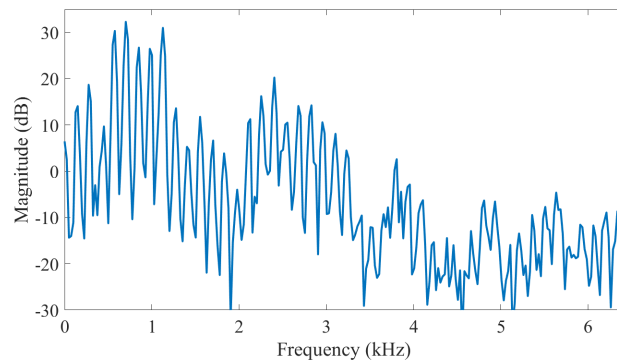
Từ phổ này, ta thực hiện lấy log các giá trị (Hình 4) rồi dùng biến đổi Fourier ngược đưa về miền Quefrency (*Quefre* ngược của *Frequen* trong Frequency - tần số) (Hình 5). Đó cũng là lí do cái ta có được gọi là cepstrum (*ceps* ngược của *spec* trong spectrum - phổ).

Còn để có được MFCCs, ta chỉ cần thêm bước lọc bằng bộ lọc tần số Mel trước khi tiến hành bước biến đổi Fourier ngược. Thông thường ta sẽ lấy 13 hệ số, hoặc bội của 13 để cho ra được kết quả tốt nhất. Riêng báo cáo này, chúng tôi sử dụng số hệ số là 13.

[‡]Fourier transform, đường dẫn: https://en.wikipedia.org/wiki/Fourier_transform



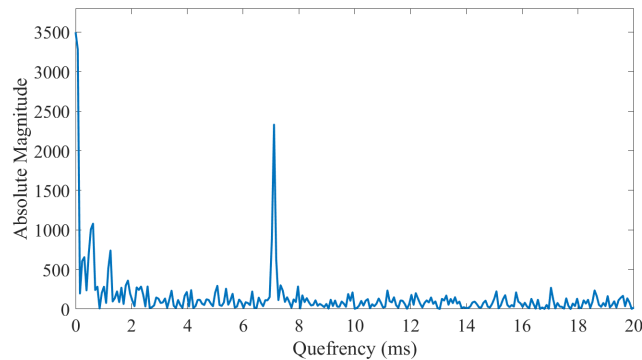
Hình 3. Phổ tín hiệu sau khi qua biến đổi Fourier [7].



Hình 4. Kết quả sau khi lấy log phổ tín hiệu [7].

4.3 Đặc trưng Sắc độ

Quan sát cơ bản thì tai người cảm nhận hai độ cao âm nhạc có màu sắc giống nhau khi chúng cách nhau một **quãng tám** (an octave). Dựa trên quan sát này, một cao độ âm nhạc có thể được chia làm 2 thành phần: **độ cao của âm** (pitch) và **sắc độ** (chroma). Giả sử tỉ lệ về khoảng cách giữa 2 nốt trong một quãng tám là bằng nhau, ta có 12 giá trị sắc độ được biểu diễn bởi tập hợp: $\{C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B\}$. Như vậy, trong dải tần âm thanh mà con người có thể nghe thấy (từ 20Hz - 20KHz), 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



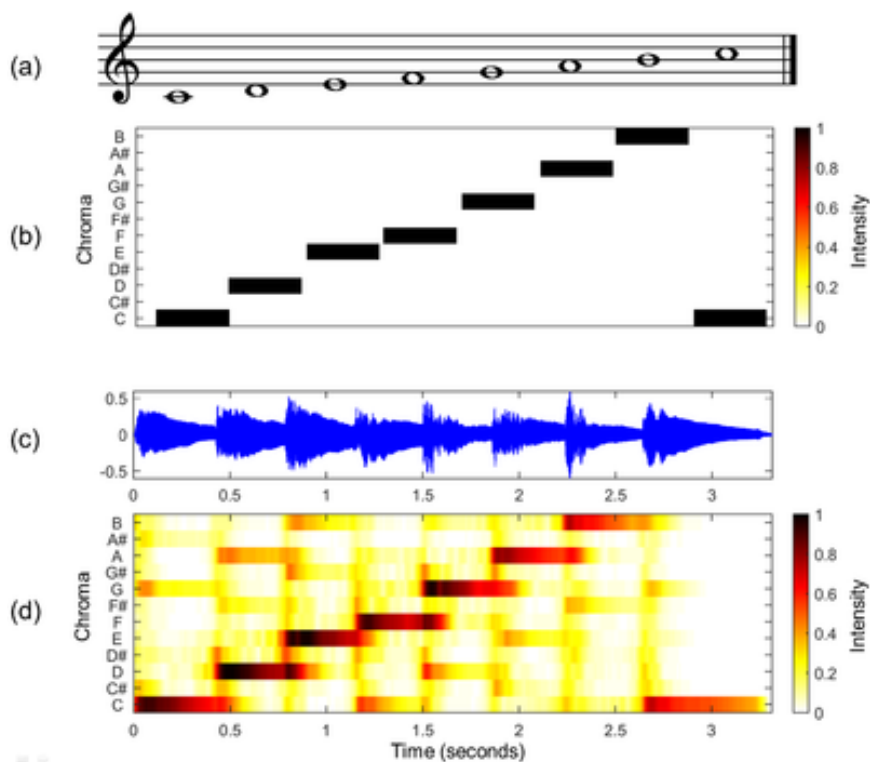
Hình 5. Cepstrum của tín hiệu [7].

lặp lại (nốt thứ 1 là Đô (C) thì nốt thứ 13 cũng là Đô nhưng ở tần số cao hơn). Tập hợp tất cả các cao độ của cùng 1 sắc độ được gọi là lớp cao độ (pitch class) (Hình 6). Ví dụ, tập hợp tất cả các cao độ của sắc độ C : $\{C_n : n \in \mathbb{Z}\} = \{\dots, C_2, C_1, C_0, C_1, C_2, C_3, \dots\}$. Như vậy, ta có thể thay thế thang tần số trong phân tích



Hình 6. Lớp cao độ của nốt Đô [9].

phổ của một tín hiệu âm thanh thành thang các sắc độ. Biểu đồ cường độ các sắc độ theo thời gian được gọi là biểu đồ sắc độ của âm thanh: Ở Hình a) và b) trong Hình 7 là âm giai của nốt Đô và biểu đồ sắc độ tương ứng



Hình 7. Ví dụ về sắc độ của âm thanh phát ra từ đàn piano [10].

(lý tưởng). Hình c) trong hình 7 là bản ghi âm của âm giai nốt Đô trên đàn piano và hình d) trong hình 7 là đồ thị sắc độ tương ứng.

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.

5. PHƯƠNG PHÁP

5.1 Tiền xử lý dữ liệu

5.1.1 Kết hợp đặc trưng và Chuẩn hoá dữ liệu

Mỗi mẫu dữ liệu đầu vào sẽ được trích lọc các đặc trưng là quang phổ Mel - \mathbf{X}_{Mel} , MFCCs - $\mathbf{X}_{\text{MFCCs}}$ và Sắc độ - $\mathbf{X}_{\text{Chroma}}$. Mỗi đặc trưng này đều là một mảng 2D với số chiều khác nhau, chúng tôi đã tiến hành lấy trung bình mỗi cột trong mảng 2D đưa về thành một véc-tơ (dạng dữ liệu 1D). Ví dụ như mới quang phổ Mel, thì ta sẽ có $\mathbf{x}_{\text{Mel}} = \mathbb{E} \left\{ (\mathbf{X}_{\text{Mel}})_{*,*} \right\}$. Những đặc trưng khác được tiến hành tương tự như đặc trưng quang phổ Mel. Một trong những lý do để lấy trung bình các cột của các đặc trưng là để kết hợp những đặc trưng này lại với nhau. Tóm lại, mỗi mẫu dữ liệu đầu vào sẽ cho một vector đặc trưng $\mathbf{x} = [\mathbf{x}_{\text{Mel}}; \mathbf{x}_{\text{MFCCs}}; \mathbf{x}_{\text{Chroma}}]$.

Giá trị của các đặc trưng cũng cách biệt nhau khá lớn nên chúng tôi đã triển khai hai loại chuẩn hoá dữ liệu là **chuyển khoảng giá trị** (khoảng $[0, 1]$) (Phương trình (2)).

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (2)$$

5.1.2 Cân bằng dữ liệu

Ngoài việc huấn luyện với tập dữ liệu không cân bằng, chúng tôi cũng thử nghiệm trên tập dữ liệu được cân bằng thông qua phương pháp cân bằng dữ liệu tổng hợp lấy mẫu quá mức [11]. Cụ thể, với mỗi mẫu thuộc nhóm thiểu số, lựa chọn ra k mẫu thiểu số láng giềng gần nhất với nó và sau đó thực hiện tổ hợp tuyến tính để tạo ra mẫu giả lập. Giả sử một mẫu thuộc nhóm thiểu số \mathbf{x} và một trong những mẫu thiểu số láng giềng là \mathbf{z} thì ta có thể sinh ra một mẫu mới cũng thuộc nhóm thiểu số là $\mathbf{x}' = \mathbf{x} + \delta(\mathbf{z} - \mathbf{x})$ với δ là một số ngẫu nhiên trong khoảng $(0, 1)$. Việc xác định k mẫu thiểu số láng giềng và số mẫu sinh ra trong quá trình cân bằng của chúng tôi sẽ sử dụng máy hỗ trợ véc-tơ (SVM - **S**upport **V**ector **M**achine). Sau khi cân bằng, số lượng mẫu dương tính bằng với số lượng mẫu âm tính và bằng 2735 mẫu. Vậy tổng số lượng dữ liệu huấn luyện sẽ là 5470.

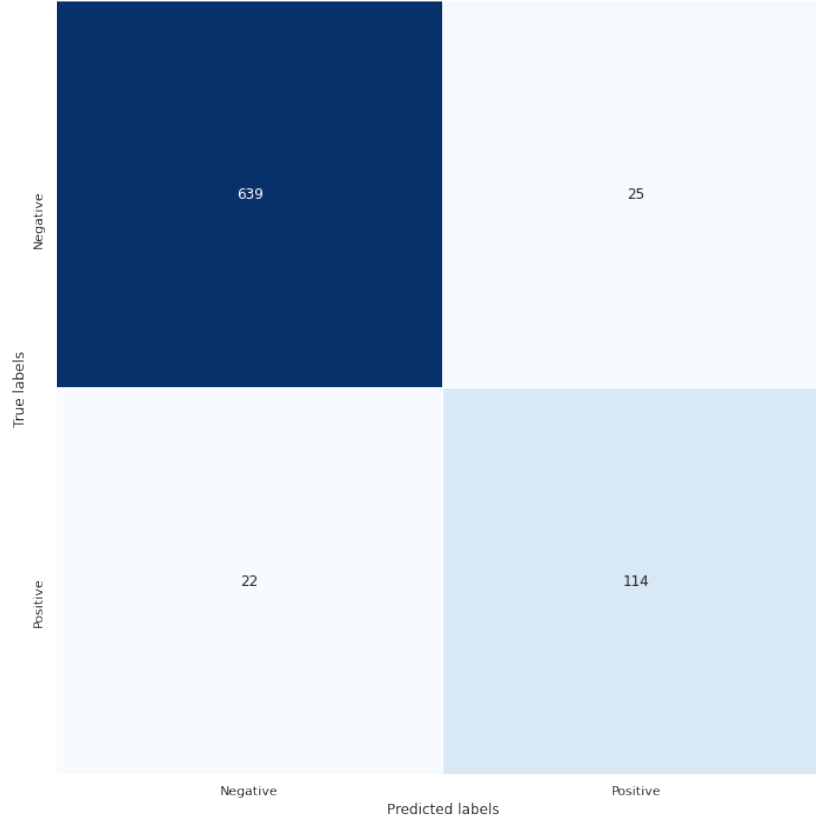
5.2 Mô hình phân loại

Mô hình được sử dụng là Light GBM - **L**ight **G**radient **B**oosting **M**achine, là một mô hình Gradient Boosting phân tán dành cho học máy do Microsoft phát triển. Mô hình dựa trên các thuật toán cây quyết định kết hợp một kỹ thuật độc đáo được gọi là lấy mẫu một phía trên Gradient. Light GBM đang được sử dụng rộng rãi trong nhiều giải pháp chiến thắng của các cuộc thi về học máy. Các thực nghiệm so sánh trên các tập dữ liệu công khai cho thấy Light GBM có thể làm tốt hơn các khung tăng cường hiện có về cả hiệu suất và độ chính xác, với mức tiêu thụ bộ nhớ thấp hơn đáng kể [12].

Thông số cài đặt chính cho mô hình Light GBM như sau: "objective": "binary", "boosting_type": "gbdt", "metric": "auc", "learning_rate": 0.03, "subsample": 0.68, "tree_learner": "serial", "colsample_bytree": 0.28, "early_stopping_rounds": 100, "subsample_freq": 1, "reg_lambda": 2, "reg_alpha": 1, "num_leaves": 500 + delta. Biến delta là số gia sẽ được đề cập trong Phần 5.3.

5.3 Cách thức huấn luyện

Chúng tôi sử dụng xác thực chéo (k-fold cross validation) và chọn $k = 10$ dựa vào kinh nghiệm những nhà nghiên cứu đi trước. Kết quả đầu ra của phiên xác thực chéo sẽ được tính theo kiểu bầu chọn mềm (soft voting) [13], tức là giá trị đầu ra sau mỗi phiên xác thực chéo là $y = 0.1 \times \sum_{i=1}^{10} y^{(i)}$, trong đó $y^{(i)}$ là kết quả của lần xác thực thứ i . Không dừng lại ở đó, chúng tôi còn lồng các phiên xác thực chéo lại với nhau (nested cross validation) để cho ra kết quả cuối cùng. Trong thực nghiệm, chúng tôi triển khai 10 phiên xác thực chéo đánh số thứ tự 0 tới 9, ở mỗi phiên này, giá trị cài đặt delta sẽ bằng số thứ tự phiên bội với 100. Một cách rõ ràng, giá trị cài đặt num_leaves sẽ lần lượt nhận các giá trị 500, 600, ..., 1300, 1400 qua các phiên. Chúng tôi cũng tiếp tục sử dụng bầu chọn mềm giữa các phiên, do đó kết quả cuối cùng sẽ là trung bình của 10 phiên.



Hình 8. Unnormalized confusion matrix.

6. KẾT QUẢ VÀ THẢO LUẬN

6.1 Đánh giá trên tập kiểm tra

Từ unnormalized confusion matrix (Hình 8), ta không khó để tính ra các giá trị quan trọng để đánh giá là độ chính xác - accuracy (Phương trình (3)), độ chuẩn xác - precision (Phương trình (4)), độ phủ hay độ nhạy - recall (Phương trình (5)) và điểm F1 - F1 score (Phương trình (6)) là:

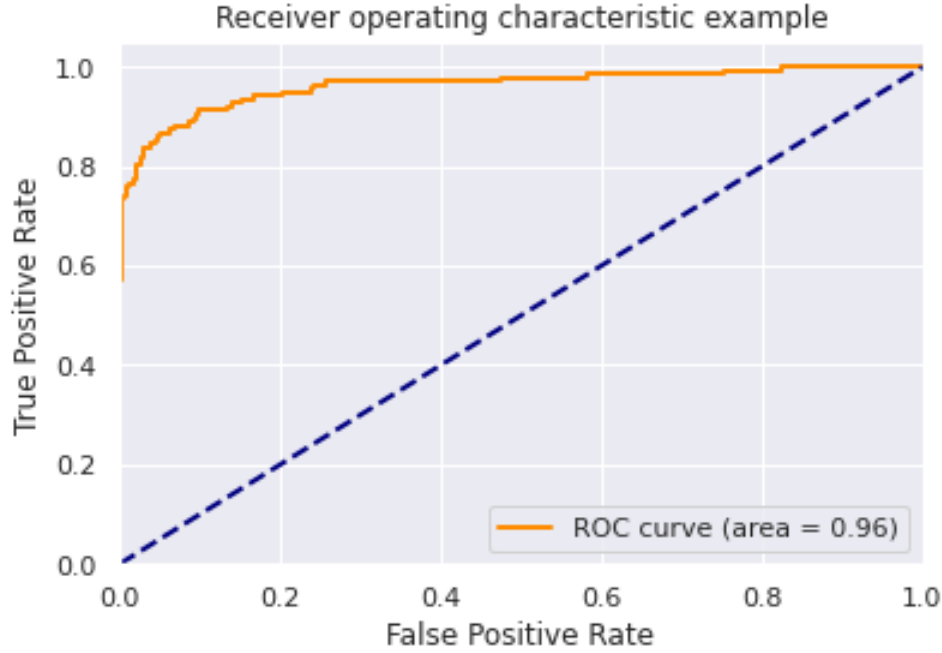
$$\text{Accuracy} = \frac{114 + 639}{800} = 0.94 \quad (3)$$

$$\text{Precision} = \frac{114}{114 + 25} = 0.82 \quad (4)$$

$$\text{Recall} = \frac{114}{114 + 22} = 0.84 \quad (5)$$

$$\text{F1 score} = 2 \times \frac{0.84 \times 0.82}{0.84 + 0.82} = 0.83 \quad (6)$$

Với độ chính xác 94%, có vẻ như bộ phân loại của chúng ta làm việc khá ổn. Tuy nhiên, với một bộ dữ liệu mất cân bằng, chỉ xem xét mỗi độ chính xác rất dễ dẫn đến những nhận định sai lầm. Không quá bất ngờ khi độ chuẩn xác và độ phủ thấp hơn độ chính xác khá nhiều (khoảng 10%). Điều này cho thấy việc tỷ lệ tìm được và bỏ sót của mô hình vẫn chưa thực sự ở mức độ xuất sắc. Chính xác hơn, mô hình tìm sai 25 mẫu âm tính và bỏ sót 22 mẫu dương tính. Việc bỏ sót $\frac{22}{22+114} \approx 16\%$ mẫu dương tính trong những bài toán về y tế cho thấy mô hình vẫn còn nhiều hạn chế.



Hình 9. Receiver Operating Characteristic curve và Area Under the Curve.

Nếu đây là một bài toán phân loại thông thường thì với ROC (Hình 9) cho thấy đây là một bộ phân loại có chất lượng tốt với AUC là 0.96, gần bằng 1.0. Nhưng như đã đề cập, với một vấn đề liên quan về y tế thì kỳ vọng hiển nhiên phải cao hơn những bài toán thông thường.

Ngoài những giá trị độ đo thông thường, chúng tôi ngoài ra cũng tính toán thêm một hệ số hay được sử dụng với những dữ liệu mất cân bằng, gọi là hệ số Cohen's Kappa (Cohen's Kappa coefficient) [14]. Dựa vào ma trận ở Hình 8, ta tính được hệ số Cohen's Kappa k như sau:

$$p_0 = \frac{639 + 114}{800} = 0.94 \quad (7)$$

$$p_e = \frac{639 + 22}{800} \times \frac{639 + 25}{800} + \frac{25 + 114}{800} \times \frac{22 + 114}{800} = 0.72 \quad (8)$$

$$\mathcal{K} = \frac{p_0 - p_e}{1 - p_e} = 0.79 \quad (9)$$

Nếu để ý, p_0 (Phương trình (7)) chính là độ chính xác, còn p_e (Phương trình (8)) là thước đo sự thống nhất giữa dự đoán của mô hình và nhãn. Trên thực tế, hệ số \mathcal{K} (Phương trình (9)) cố gắng điều chỉnh sai lệch đánh giá bằng cách tính đến việc phân loại chính xác bằng một dự đoán ngẫu nhiên, điều dễ xảy ra với các tập dữ liệu mất cân bằng. Với giá trị $\mathcal{K} = 0.79$, mô hình có thể được coi là có khả năng phân loại **khá đáng kể** [15] khi nằm trong đoạn [0.61, 0.80].

6.2 Đánh giá trên tập kiểm tra công khai của ban tổ chức AICOVIDVN

Chúng tôi sử dụng mô hình và cách thức huấn luyện đã triển khai, áp dụng lên toàn bộ tập dữ liệu gồm 4.068 mẫu. Sau đó cho kết quả trên tập kiểm tra của ban tổ chức (tập gồm 1233 mẫu). Kết quả AUC nhận được là 0.88, xếp hạng 16 trên bảng tổng sắp.[§] Vì chi tiết kết quả ban tổ chức không công khai, vậy nên chúng tôi không thể đánh giá chi tiết trên tập này.

[§]Bảng tổng sắp, đường dẫn: <https://aihub.vn/competitions/22#results>.

7. KẾT LUẬN

Với kết quả đạt được, chúng tôi thấy mô hình vẫn còn nhiều hạn chế nếu đưa vào áp dụng thực tế. Dẫu vậy, báo cáo vẫn cung cấp những hiểu biết về các loại đặc cho bài toán nhận diện COVID-19 qua tiếng ho.

Trong tương lai, hạn chế của mô hình được trình bày trong báo cáo có thể được cải thiện bằng cách sử dụng các mô hình tiền huấn luyện, hoặc sử dụng dữ liệu dưới dạng 2D và đưa số chiều hợp lý để có thể kết hợp với nhau. Ngoài ra, thu thập thêm dữ liệu là điều rất cần thiết để cải thiện được mô hình.

LỜI CẢM ƠN

Xin chân thành cảm ơn giáo viên bộ môn của môn *Trí tuệ nhân tạo trong điều khiển* trong học kì 202, thầy **TS. Phạm Việt Cường** đã truyền đạt nhiều kiến thức quan trọng của ngành học máy. Nhờ vào đó, chúng tôi có được nền tảng để tìm hiểu và phát triển cho bài toán nhận diện COVID-19 qua tiếng ho cũng như đạt được thứ hạng tốt trên bảng tổng sắp của cuộc thi đang tham gia.

TÀI LIỆU THAM KHẢO

- [1] K.Vân, “Địa chỉ và mức giá xét nghiệm COVID-19 cho người dân ra khỏi TP.HCM.” Bộ Y tế, 07 July 2021 <https://ncov.moh.gov.vn/en/-/6847426-5408> (2021). (Accessed: 13 August 2021).
- [2] Chu, J., “Artificial intelligence model detects asymptomatic Covid-19 infections through cellphone-recorded coughs.” MIT News, 29 October 2020 <https://news.mit.edu/2020/covid-19-cough-cellphone-detection-1029> (2020). (Accessed: 13 August 2021).
- [3] Laguarda, J., Hueto, F., and Subirana, B., “Covid-19 artificial intelligence diagnosis using only cough recordings,” *IEEE Open Journal of Engineering in Medicine and Biology* **1**, 275–281 (2020).
- [4] Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., and Herrera, F., [*Learning from imbalanced data sets*], vol. 10, Springer (2018).
- [5] Devi, R. and Pugazhenth, D., “Ideal sampling rate to reduce distortion in audio steganography,” *Procedia Computer Science* **85**, 418–424 (2016). International Conference on Computational Modelling and Security (CMS 2016).
- [6] Stevens, S. S., Volkman, J., and Newman, E. B., “A scale for the measurement of the psychological magnitude pitch,” *The journal of the acoustical society of america* **8**(3), 185–190 (1937).
- [7] Bäckström, T., “Cepstrum and mfcc.” Aalto University Wiki, 21 May 2019 <https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC> (2019). (Accessed: 17 August 2021).
- [8] Tiwari, V. T., “Mfcc and its applications in speaker recognition,” *Int. J. Emerg. Technol.* **1** (01 2010).
- [9] Wikipedia, “Scientific pitch notation — Wikipedia, the free encyclopedia.” <http://en.wikipedia.org/w/index.php?title=Scientific%20pitch%20notation&oldid=1039089704> (2021). [Accessed: 17 August 2021].
- [10] Wikipedia, “Chroma feature — Wikipedia, the free encyclopedia.” <http://en.wikipedia.org/w/index.php?title=Chroma%20feature&oldid=1011214694> (2021). [Accessed: 17 August 2021].
- [11] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., “Smote: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research* **16**, 321–357 (Jun 2002).
- [12] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y., “Lightgbm: A highly efficient gradient boosting decision tree,” in [*NIPS*], (2017).
- [13] Islam, R. and Shahjalal, M., “Soft voting-based ensemble approach to predict early stage drc violations,” 1081–1084 (08 2019).
- [14] Wang, J., Yang, Y., and Xia, B., “A simplified cohen’s kappa for use in binary classification data annotation tasks,” *IEEE Access* **7**, 164386–164397 (2019).
- [15] Rau, F., Hartmann, A., Tellhelm, B., Zwick, M., Klumpp, S., Neumann, A., Oltersdorf, B., von Pückler, K., Failing, K., and Kramer, M., “Observer variability and sensitivity of radiographic diagnosis of canine medial coronoid disease,” *Tierärztliche Praxis. Ausgabe K, Kleintiere/Heimtiere* **39**, 313–22 (01 2011).