

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Xây dựng mô-đun hỗ trợ hệ thống giám sát thi cử EdTech

LÊ ĐỨC TÙNG

tung.ld183852@sis.hust.edu.vn

Ngành Công nghệ thông tin

Giảng viên hướng dẫn: TS. Đặng Tuấn Linh

Chữ ký GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ thông tin và Truyền thông

HÀ NỘI, 08/2023

LỜI CẢM ƠN

Trước khi bắt đầu, tôi muốn bày tỏ lòng cảm ơn chân thành đến tất cả những cá nhân đã giúp đỡ tôi trong quá trình thực hiện đồ án này.

Đầu tiên, tôi muốn gửi lời cảm ơn đến chính tôi, vì đã mạnh mẽ theo đuổi một thứ bản thân mình không thích thú, đam mê.

Tiếp đó, tôi muốn gửi lời cảm ơn đến giảng viên hướng dẫn tôi, TS.Đặng Tuấn Linh, vì đã luôn nhẹ nhàng đưa ra những đường lối chiến lược, những lời khuyên và hướng dẫn tôi trong quá trình nghiên cứu và thực hiện đồ án.

Tôi muốn gửi lời cảm ơn đến gia đình và bạn bè của tôi, vì đã động viên và hỗ trợ tôi trong suốt quá trình thực hiện đồ án. Những lời khuyên và động viên đó đã giúp tôi vượt qua những thử thách trong quá trình muôn từ bỏ và hoàn thành đồ án.

Một lần nữa, tôi xin chân thành cảm ơn tất cả những người đã giúp đỡ tôi trong quá trình thực hiện đồ án này.

TÓM TẮT NỘI DUNG ĐỒ ÁN

Trong thời đại công nghệ ngày càng phát triển, việc áp dụng các công nghệ mới vào giáo dục ngày càng phát triển. Đi đôi với sự dạy và học là hệ thống bài thi trực tuyến cùng với hệ thống chống gian lận trong quá trình thi.

EdTech là một trong những trung tâm nghiên cứu của Bách Khoa Hà Nội tiên phong trong việc áp dụng công nghệ để chống gian lận trong thi cử. Trung tâm đã phát triển một hệ thống quản lý khóa học đồng thời áp dụng các công nghệ mới để phát hiện và ngăn chặn các hành vi gian lận.

Một trong những tính năng nổi bật của hệ thống này là khả năng phát hiện âm thanh tiếng nói khi sinh viên trao đổi bài. Tính năng này sử dụng công nghệ nhận dạng giọng nói để phát hiện các từ khóa liên quan đến gian lận. Khi phát hiện các tiếng nói, hệ thống sẽ gửi cảnh báo cho người vi phạm và gửi giáo viên để xử lý.

Hệ thống này cũng có tính năng điểm danh và ghi lại những gian lận mà sinh viên đã vi phạm. Tính năng này sử dụng camera để ghi lại hình ảnh của sinh viên trong quá trình thi cử. Nếu sinh viên có bất kỳ hành vi gian lận nào, hệ thống sẽ tự động ghi lại hình ảnh để giáo viên có một cái nhìn trực quan.

Hệ thống quản lý khóa học của EdTech đã và đang được và đang được phát triển hoàn thiện để đưa vào triển khai trong quá trình học tập ở Trường đại học Bách Khoa Hà Nội và hy vọng mang lại hiệu quả cao trong việc chống gian lận trong thi cử. Hệ thống này giúp đảm bảo tính công bằng của các kỳ thi và giúp nâng cao chất lượng giáo dục.

Một số lợi ích của việc sử dụng hệ thống quản lý khóa học của EdTech để chống gian lận trong thi cử bao gồm: Đảm bảo tính công bằng của các kỳ thi: hệ thống sử dụng công nghệ để phát hiện và ngăn chặn gian lận, giúp đảm bảo tính công bằng của các kỳ thi. Nâng cao chất lượng giáo dục: Hệ thống giúp ngăn chặn gian lận, giúp học sinh tập trung vào việc học tập và đạt được kết quả tốt hơn. Tiết kiệm thời gian và chi phí cho giáo viên: Hệ thống giúp giáo viên tiết kiệm thời gian và chi phí trong việc coi thi.

Đồ án này triển khai bài toán phát hiện tiếng nói khi sinh viên trao đổi bài cũng như hệ thống điểm danh và ghi lại những gian lận mà sinh viên đã vi phạm và triển khai các mô đun đó trên hệ thống Edtech.

Sinh viên thực hiện
(Ký và ghi rõ họ tên)

ABSTRACT

In the rapidly evolving era of technology, the integration of new technologies into education is advancing. Alongside the teaching and learning process, online examination systems and anti-cheating measures have become essential. EdTech, a research center at the Hanoi University of Science and Technology, is at the forefront of applying technology to combat cheating during examinations. The center has developed a course management system that utilizes cutting-edge technologies to detect and prevent cheating behaviors.

One prominent feature of this system is its ability to detect voice interactions among students during exams. This feature employs voice recognition technology to identify keywords related to cheating. Upon detecting suspicious conversations, the system issues warnings to the violators and notifies the instructors for appropriate action.

Furthermore, the system includes attendance tracking and records any cheating instances by capturing images of students during the exam. If any cheating behavior is detected, the system automatically records visual evidence for the instructors to review.

The course management system developed by EdTech is continuously being refined and implemented in the academic process at Hanoi University of Science and Technology. The goal is to ensure fairness in examinations and enhance the overall quality of education.

The benefits of using EdTech's course management system to combat cheating in exams include ensuring fairness in assessments, enhancing educational quality by deterring cheating, and saving time and costs for instructors in exam invigilation.

This project tackles the challenge of detecting voice interactions during exams and implementing attendance tracking and cheating detection modules within the EdTech system. The abstract provides an overview of the system's capabilities in safeguarding academic integrity, fostering focused learning, and achieving improved academic outcomes.

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	2
1.2.1 Mục tiêu.....	2
1.2.2 Phạm vi đề tài	2
1.3 Định hướng giải pháp.....	2
1.3.1 Giải pháp mục tiêu bài toán phát hiện tiếng nói.....	2
1.3.2 Giải pháp mục tiêu bài toán điểm danh.....	3
1.4 Bố cục đồ án	3
CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU.....	4
2.1 Tổng quan về hệ thống	4
2.2 Khảo sát hiện trạng bài toán phát hiện giọng nói.....	6
2.2.1 Lý thuyết	6
2.2.2 Tổng quan bài toán	8
2.3 Tổng quan bài toán điểm danh.....	10
2.3.1 Biểu đồ use case bài toán điểm danh	11
2.3.2 Quy trình nghiệp vụ	12
2.3.3 Đặc tả usecase	14
2.4 Yêu cầu phi chức năng	15
2.4.1 Yêu cầu trải nghiệm người dùng	15
2.4.2 Yêu cầu về bảo mật và tải đường truyền.....	16
2.4.3 Yêu cầu khả năng mở rộng và nâng cấp, bảo trì.....	16

CHƯƠNG 3. CÔNG NGHỆ SỬ DỤNG.....	17
3.1 JavaScript	17
3.1.1 Lí do lựa chọn.....	17
3.1.2 Vai trò trong đồ án	17
3.2 SpringBoot	18
3.2.1 Lí do lựa chọn.....	18
3.2.2 Vai trò trong đồ án	18
3.3 MySQL	19
3.3.1 Lí do lựa chọn.....	19
3.3.2 Vai trò trong đồ án	19
3.4 Python	19
3.4.1 Lí do lựa chọn.....	20
3.4.2 Vai trò trong đồ án	20
3.5 Docker.....	20
3.5.1 Lí do lựa chọn.....	21
3.5.2 Vai trò trong đồ án	21
CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ	22
4.1 Thiết kế kiến trúc luồng điểm danh	22
4.1.1 Lựa chọn kiến trúc phần mềm	22
4.1.2 Thiết kế tổng quan.....	24
4.1.3 Thiết kế chi tiết giao diện.....	24
4.1.4 Thiết kế chi tiết gói	27
4.1.5 Thiết kế cơ sở dữ liệu	27
4.2 Bài toán phát hiện tiếng nói	29
4.2.1 Khối dữ liệu	30
4.2.2 Khối huấn luyện mô hình.....	31

4.2.3 Khối triển khai trên trình duyệt.....	35
4.3 Thư viện và công cụ sử dụng	37
4.4 Minh họa các kết quả đạt được	37
4.4.1 Bài toán điểm danh	37
4.4.2 Kết quả đạt được bài toán phát hiện tiếng nói.....	39
4.5 Kiểm thử.....	40
4.6 Triển khai	42
4.6.1 Cách triển khai các tính năng	42
4.6.2 Triển khai trên thực tế.....	42
CHƯƠNG 5. CÁC GIẢI PHÁP VÀ ĐÓNG GÓP NỔI BẬT.....	44
5.1 Các công hiến.....	44
5.1.1 Xây dựng tính năng phát hiện tiếng nói cụ thể áp dụng vào trong trường học	44
5.1.2 Bộ dữ liệu đánh nhãn các file âm thanh	45
5.1.3 Xây dựng tính năng điểm danh cũng như quan sát các lỗi vi phạm	46
5.2 Các hạn chế.....	46
CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	47
6.1 Kết luận	47
6.2 Hướng phát triển.....	47

DANH MỤC HÌNH VẼ

Hình 2.1	Hệ thống chống gian lận	5
Hình 2.2	UseCase tổng quan bài toán	6
Hình 2.3	Tần số và biên độ	6
Hình 2.4	Âm sắc	7
Hình 2.5	Mel Spectrogram	8
Hình 2.6	Mô hình phân loại âm thanh	9
Hình 2.7	Ví dụ các mô hình	10
Hình 2.8	UseCase tổng quan	11
Hình 2.9	Quy trình nghiệp vụ chụp chân dung	12
Hình 2.10	Quy trình nghiệp vụ xác nhận thông tin	13
Hình 4.1	Mô hình khách -chủ	22
Hình 4.2	Mô hình MVC trong máy chủ	23
Hình 4.3	Tổng quan luồng tính năng	24
Hình 4.4	Giao diện chụp ảnh	25
Hình 4.5	Giao diện xem lỗi vi phạm	26
Hình 4.6	Thiết kế chi tiết gói	27
Hình 4.7	Thiết kế cơ sở dữ liệu	28
Hình 4.8	Tổng quan các khối	30
Hình 4.9	Chuyển file audio thành các mel spectrogram	31
Hình 4.10	Cấu trúc mạng VGG-16	32
Hình 4.11	Độ chính xác và măt măt của cấu trúc mạng VGG-16	32
Hình 4.12	Cấu trúc mạng CNN	33
Hình 4.13	Độ chính xác và măt măt của cấu trúc mạng CNN	34
Hình 4.14	Cấu trúc mạng CNN của Google	35
Hình 4.15	File model.json	36
Hình 4.16	Giao diện chụp ảnh của sinh viên	37
Hình 4.17	Giao diện xem ảnh của sinh viên	38
Hình 4.18	Giao diện xem thông tin tham gia thi và các lỗi vi phạm của sinh viên	39
Hình 4.19	Nhãn và hình ảnh melspectrogram tương ứng	40
Hình 4.20	Trải nghiệm hệ thống giám sát thi	43
Hình 4.21	Kết quả thử nghiệm hệ thống	43
Hình 5.1	Tài nguyên dùng để phát hiện tiếng nói	44
Hình 5.2	Dữ liệu âm thanh	45

Viết tắt	Tên tiếng Anh	Tên tiếng Việt
API	Application Programming Interface	Giao diện lập trình ứng dụng
CPU	Central Processing Unit	Bộ xử lý trung tâm máy tính
GPU	Graphics Processing Unit	Bộ xử lý đồ họa
DB	Database	Cơ sở dữ liệu
CNN	Convolutional Neural Network	Mạng nơ ron tích chập
AI	Artificial Intelligence	Trí tuệ nhân tạo

Viết tắt	Tên tiếng Anh
Browser	Trình duyệt
Client	Máy khách
Server	Máy chủ
Model	Mô hình học máy
Train	Huấn luyện mô hình
Compiler	Trình biên dịch
Log	Ghi lại

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

Chương 1 sẽ giới thiệu qua lí do chọn đề tài, thành tựu đạt được cũng như các hạn chế gặp phải, trình bày mục tiêu từ đó đưa ra các giải pháp cho bài toán phát hiện tiếng nói và bài toán điểm danh

1.1 Đặt vấn đề

Hiện nay, công nghệ đã đóng vai trò vô cùng quan trọng trong cuộc sống . Từ việc sử dụng điện thoại thông minh để giữ liên lạc, đến việc sử dụng máy tính để làm việc và giải trí, công nghệ đã trở thành một phần không thể thiếu của cuộc sống hàng ngày của chúng ta. Ngoài ra, công nghệ còn được áp dụng rộng rãi trong các lĩnh vực khác như y tế, giáo dục, kinh doanh, sản xuất và nhiều lĩnh vực khác. Các công nghệ mới như trí tuệ nhân tạo, blockchain và internet vạn vật cũng đang mở ra nhiều cơ hội mới cho các nhà nghiên cứu và doanh nghiệp. Đặc biệt chuyển đổi số thúc đẩy một nền giáo dục mở, giúp con người tiếp cận thông tin đa chiều, rút ngắn khoảng cách, thu hẹp mọi không gian, tiết kiệm tối ưu về thời gian. Từ đó con người phát triển nhanh hơn về kiến thức, nhận thức và tư duy [1]. Với công nghệ hiện nay, việc tiếp thu kiến thức không còn là việc phải đến trường, lớp đến dự một tiết học nữa mà giờ đây mọi người có thể chỉ cần một thiết bị điện tử có màn hình hiển thị, kết nối internet để học trực tiếp các bài giảng online, kết nối trực tiếp với giảng viên dù học viên có ở vùng địa lý nào. Vấn đề trên được thể hiện và triển khai mạnh mẽ trong thời kỳ phong tỏa xã hội vì dịch Covid 19 diễn ra với nhu cầu hơn 20 triệu học sinh, sinh viên và 2 triệu giáo viên tham gia vào quá trình dạy và học trực tuyến [2]. Nhưng muốn việc học hiệu quả thì không thể thiếu các bài kiểm tra khách quan để đánh giá kiến thức của người học... Hệ thống học trực tuyến BK-Eleaning Management System, Teams là những ví dụ điển hình khi được xây dựng với mục đích triển khai các khóa học trực tuyến đi kèm các bài kiểm tra đánh giá kiến thức cho sinh viên trong suốt quá trình dịch bệnh.

Tuy nhiên, các hệ thống kể trên vẫn có nhiều hạn chế thiếu sót. Do đó, hệ thống MOOC được trung tâm phát triển EdTech triển khai với mục đích bổ sung các tính năng thiếu sót mà BK- Eleaning Management System chưa có như quản lý tiến độ khóa học, bài thi hay quản lý việc cấp chứng chỉ cũng được phát triển tốt hơn để đáp ứng nhu cầu ngày càng tăng cao của người học. Hiện một số chương trình trên các nền tảng này đã cho phép đổi ra tín chỉ tương đương ở các trường Đại học truyền thống [3]. Nhưng các nền tảng MOOC hiện tại tiêu biểu như Coursera, Udemy hay MOOC của Bách Khoa Hà Nội lại chưa có bất cứ cách giải quyết nào trước tình trạng sinh viên gian lận trong các bài thi trực tuyến.

Năm được vấn đề đó, trung tâm nghiên cứu EdTech đã triển khai thêm daotao.ai :một

hệ thống chống gian lận thi cử của sinh viên. Qua phân tích, việc gian lận của học sinh, sinh viên tổng quan cũng được AI phát hiện chủ yếu thông qua 2 luồng chính: thông qua hình ảnh(phát hiện các vật thể như tai nghe, smart watch, điện thoại.... hoặc xác định hướng mặt so với camera), thông qua âm thanh (phát hiện người thi có đang nói chuyện hay không). Tuy nhiên hệ thống daotao.ai mới chỉ các xử lí các vấn đề bằng hình ảnh chứ chưa xử lí gian lận bằng âm thanh cũng như chưa có bất kì giao diện nào cho học sinh, giảng viên theo dõi kết quả đầu vào đầu ra của hệ thống một cách khách quan nhất.

Nhận thấy những thiếu sót của hệ thống trên, tôi đã đề xuất xây dựng, bổ sung các thiếu sót của hệ thống daotao.ai : triển khai tính năng phát hiện tiếng nói trong quá trình thi của sinh viên và lưu lại tất cả lỗi vi phạm của sinh viên trong quá trình làm bài thi cũng như thể hiện nó trên giao diện để giảng viên có cái nhìn khách quan nhất.

1.2 Mục tiêu và phạm vi đề tài

1.2.1 Mục tiêu

Từ những vấn đề được nêu ở mục [1.1] và được sự phân công của giáo viên, tôi hướng đến giải quyết các bài toán sau:

- Triển khai bài toán phát hiện tiếng nói thời gian thực trên máy khách
- Một luồng điểm danh khuôn mặt cũng như ghi lại các gian lận phát sinh khi sinh viên có những hành vi không đúng

Với 2 bài toán trên, mục tiêu đặt ra là :

- Tính năng phân loại âm thanh dễ triển khai cài đặt, không tốn CPU, GPU của server, sử dụng ít tài nguyên hệ thống nhất.
- Các giao diện triển khai cho người dùng đẹp dễ dàng tương tác, hệ thống lưu trữ dữ liệu.

1.2.2 Phạm vi đề tài

Các tính năng trên được đặt trong một hệ thống triển khai theo từng lớp học với số lượng khoảng 40 sinh viên/ lớp.

1.3 Định hướng giải pháp

Với mỗi mục tiêu đề ra ở mục 1.2 tôi xin được đề ra các giải pháp tương ứng sau:

1.3.1 Giải pháp mục tiêu bài toán phát hiện tiếng nói

- Tự xây dựng bộ dữ liệu phù hợp để huấn luyện mô hình.
- Tiền xử lí dữ liệu.
- Tìm hiểu các mô hình phân loại và chọn ra mô hình phù hợp đối với các tệp âm thanh.

- Triển khai mô hình trên trình duyệt.
- Triển khai trên hệ thống daotao.ai.

1.3.2 Giải pháp mục tiêu bài toán điểm danh

- Thực hiện luồng nghiệp vụ chụp ảnh của người dùng ở máy khách sau đó chuyển về máy chủ AI xử lý và lưu trữ dữ liệu ở cơ sở dữ liệu.
- Phân tích yêu cầu, nghiệp vụ và hướng triển khai cụ thể cho bài toán lưu lại lỗi vi phạm.
- Triển khai trên hệ thống daotao.ai.

1.4 Bố cục đồ án

Phần còn lại của đồ án sẽ được tổ chức thành năm chương với nội dung như sau:

Chương 2 : Chương 2 giới thiệu tổng quan về bối cảnh hiện tại của bài toán, những thành tựu và thách thức được đặt ra và khảo sát phân tích các chức năng của bài toán

Chương 3, Chương 3 trình bày về các công nghệ sử dụng

Chương 4: Đây là nội dung chính của đồ án, chương trình bày về tổng quan của bài toán, cũng như đi sâu vào quá trình thiết kế và triển khai các tính năng trong quá trình thực hiện đồ án và kết quả thực nghiệm của bài toán

Chương 5: Chương 5 trình bày về các đóng góp nổi bật ,những hạn chế của các tính năng.

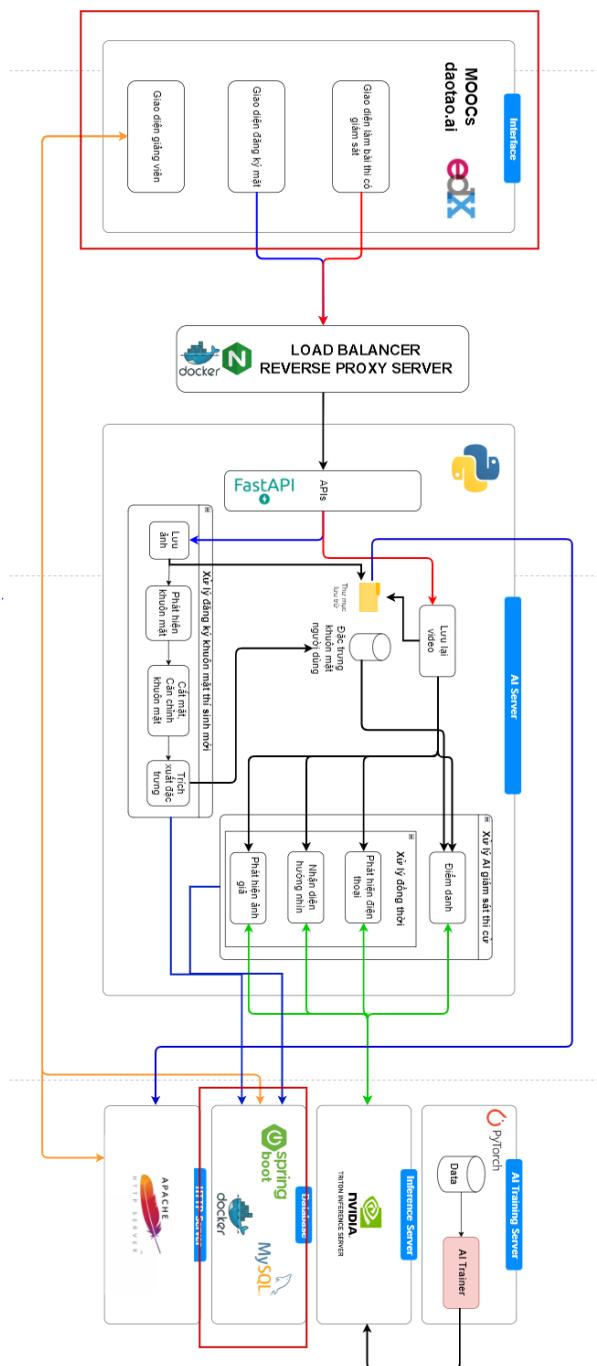
Chương 6: Đây là chương kết luận cũng như trình bày về các hướng phát triển trong tương lai.

CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU

Chương 1 đã giới thiệu qua những bối cảnh thực tế dẫn đến lý do chọn đề tài, bên cạnh đó cũng như tổng quan về giải pháp. Ở chương 2 tôi sẽ đi sâu vào quá trình khảo sát thực trạng của tính năng phát hiện tiếng nói cũng như tổng quan các chức năng của bài toán điểm danh

2.1 Tổng quan về hệ thống

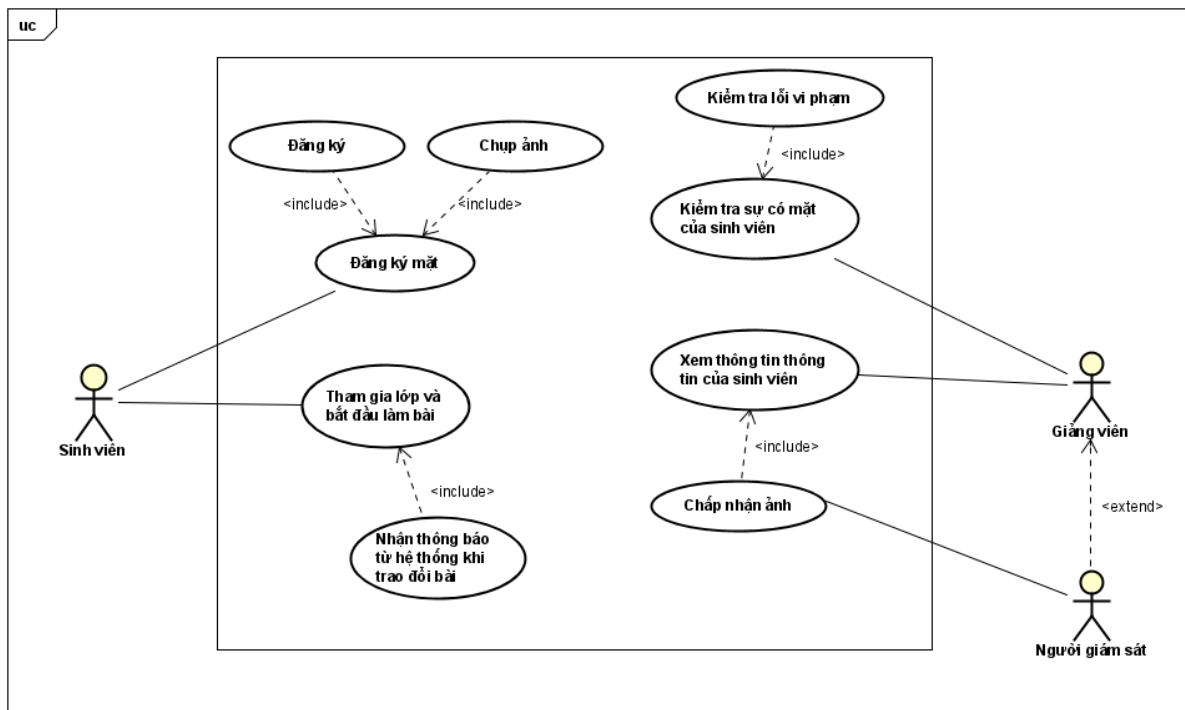
Sau đây tôi xin trình bày tổng quan về hệ thống nhóm tôi đã triển khai trên daotao.ai



Hình 2.1: Hệ thống chống gian lận

Hệ thống gồm nhiều phần công việc và được chia cho mỗi thành viên các nhiệm vụ khác nhau. Hình 2.1 thể hiện phần nhiệm vụ của tôi là việc xây dựng các giao diện tương tác với người dùng với tính năng phát hiện âm thanh trên mỗi máy khách cũng như phần máy chủ lưu trữ dữ liệu sau khi được AI xử lí trả về.

Sau đây là Usecase tổng quan cho phần công việc của tôi.



Hình 2.2: UseCase tổng quan bài toán

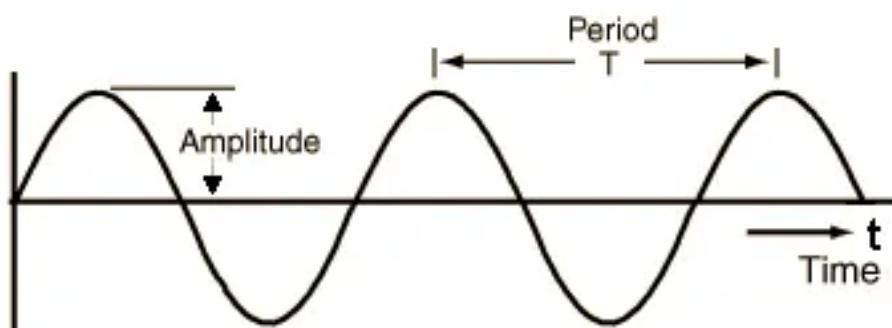
Hình 2.2 mô tả 2 chức năng chính của bài toán của tôi là : Cảnh báo người dùng khi có vi phạm(âm thanh) và bài toán điểm danh. Sau đây tôi xin lần lượt trình bày tổng quan của từng bài toán.

2.2 Khảo sát hiện trạng bài toán phát hiện giọng nói

2.2.1 Lý thuyết

a, Âm thanh

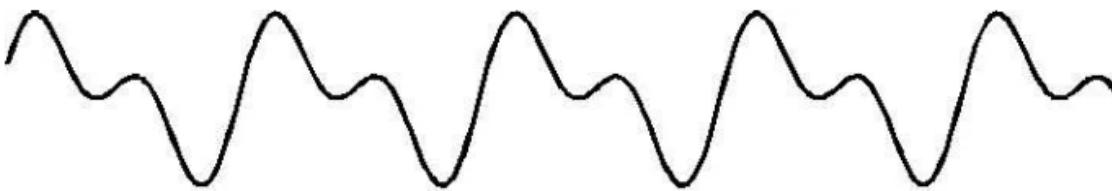
Tất cả chúng ta đều nhớ ở trường rằng tín hiệu âm thanh được tạo ra bởi sự thay đổi áp suất không khí. Chúng ta có thể đo cường độ của sự thay đổi áp suất và vẽ các phép đo đó theo thời gian. Tín hiệu âm thanh thường lặp đi lặp lại đều đặn để mỗi làn sóng có hình dạng giống nhau. Độ cao cho thấy cường độ của âm thanh và được gọi là biên độ. Hình 2.3 thể hiện tần số, biên độ của âm thanh:



Hình 2.3: Tần số và biên độ

Thời gian để tín hiệu hoàn thành một sóng đầy đủ là khoảng thời gian. Số lượng sóng do tín hiệu tạo ra trong một giây được gọi là tần số. Tần số là đối ứng của khoảng thời gian. Đơn vị của tần số là Hertz.

Phân lớn các âm thanh chúng ta gặp có thể không tuân theo các mẫu tuần hoàn đơn giản và đều đặn như vậy. Nhưng các tín hiệu có tần số khác nhau có thể được cộng lại với nhau để tạo ra các tín hiệu tổng hợp với các mẫu lặp lại phức tạp hơn. Tất cả âm thanh mà chúng ta nghe thấy, bao gồm cả giọng nói con người của chúng ta, bao gồm các dạng sóng như sau. Ví dụ, Hình 2.4 có thể là âm thanh của một nhạc cụ:

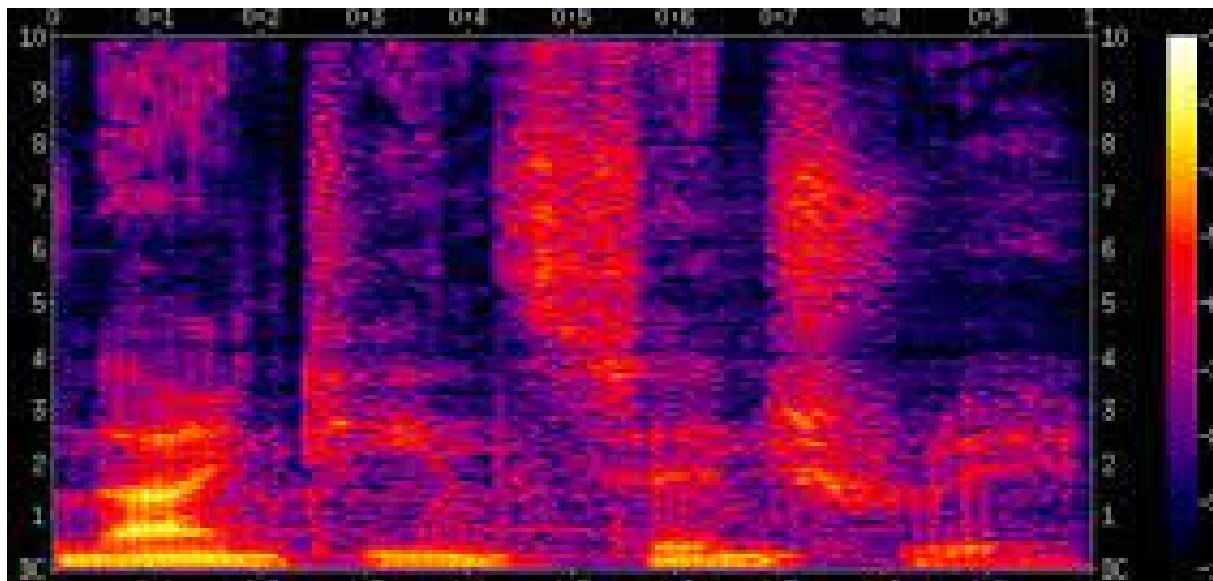


Hình 2.4: Âm sắc

Tai người có thể phân biệt giữa các âm thanh khác nhau dựa trên "chất lượng" của âm thanh, còn được gọi là âm sắc.

b, Mel Spectrogram

Mel Spectrogram của một tín hiệu biểu thị độ lớn tần số của âm thanh theo thời gian và giống như một 'bức ảnh' của tín hiệu. Nó vẽ biểu đồ thời gian trên trục x và giá trị Logarit của tần số trên trục y. Cứ như thế nó đã lấy đi lấy lại quang phổ ở các thời điểm khác nhau, sau đó kết hợp tất cả chúng lại với nhau thành một ô duy nhất [4] Hình 2.5 thể hiện hình ảnh của một mel Spectrogram.



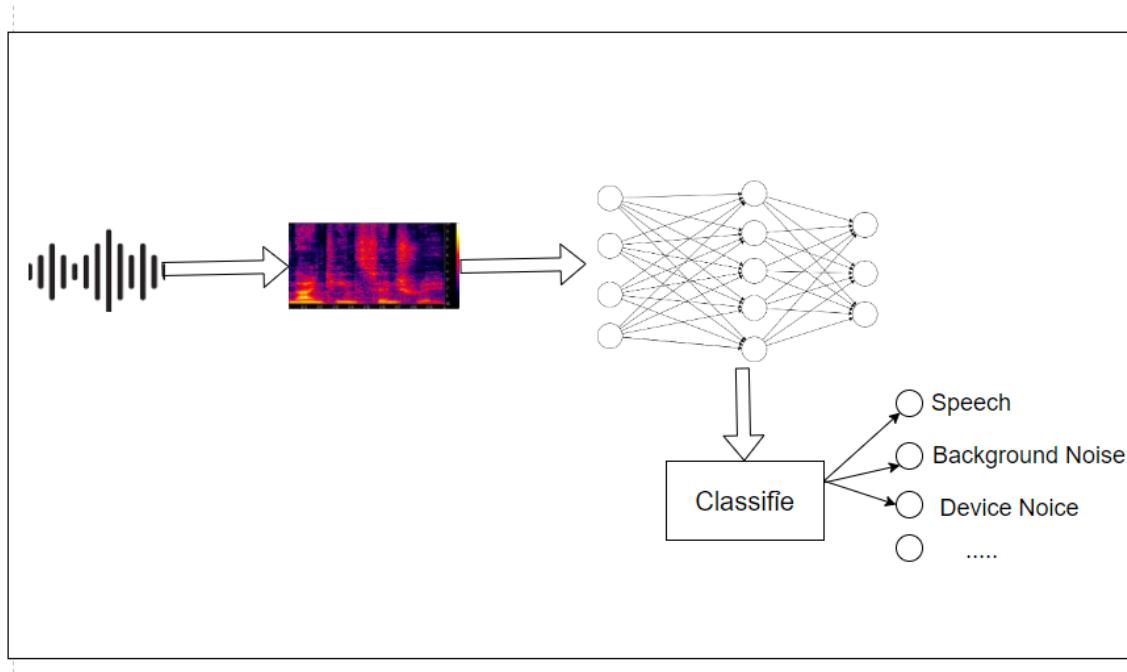
Hình 2.5: Mel Spectrogram

Nó sử dụng các màu khác nhau để biểu thị độ lớn Logarit của từng tần số. Màu càng sáng thì năng lượng của tín hiệu càng cao. Mỗi 'lát cắt' dọc của mel spectrogram. Về cơ bản là phổ của tín hiệu tại thời điểm đó và cho biết cường độ tín hiệu được phân bổ như thế nào ở mọi tần số được tìm thấy trong tín hiệu tại thời điểm đó.

2.2.2 Tổng quan bài toán

Từ khi công nghệ phát triển, việc chống gian lận trong kỳ thi cử đã trở thành một trong những vấn đề quan trọng và cần thiết. Phát hiện giọng nói trong chống gian lận thi cử là một trong các phương pháp đang được sử dụng hiện nay để giám sát và phát hiện các hành vi gian lận liên quan đến âm thanh trong quá trình thi cử. Hiện nay trên thế giới đã triển khai nhiều hệ thống phát hiện các hành vi gian lận liên quan đến phát hiện âm thanh trong thi cử như là Acoustic Fingerprinting Systems (Hệ thống vân tay âm thanh)[5]: phần mềm này được tích hợp vào hệ thống giám sát thi cử để xác định xem giọng nói của thí sinh có khớp với thông tin đã đăng ký hay không Tuy các hệ thống trên đã được phát triển ở nước ngoài nhưng ở Việt Nam ở các trường cấp 3, đại học, tiêu biểu như Đại học Bách Khoa Hà Nội thì chưa có. Nguyên nhân cũng là do chi phí để sở hữu các hệ thống trên để tích hợp vào hệ thống chung là khá lớn, cũng như việc tích hợp các công nghệ AI vào hệ thống thi cử khá khó khăn. Một khác với sự phát triển của đào tạo từ xa như hiện tại, chống gian lận là một vấn đề không thể không có. Do vậy tôi đề xuất tính năng phát hiện giọng nói trong quá trình thi của hệ thống daotao.ai.

Hiện nay, hầu hết các bài toán phân loại âm thanh đều chủ yếu được xử lý theo các bước như Hình 4.6 [6]:



Hình 2.6: Mô hình phân loại âm thanh

Do âm thanh file(.wav) không thể là đầu vào của một mô hình nên tôi dùng thư viện libosa của Python chuyển âm thanh về melspectrogram.

Với tầng mô hình có thể sử dụng các thuật toán CNN, LSTM, VGG-16... với độ chính xác khá cao như Hình 4.7 [7]:

Title	Notes	Accuracy
BEATs: Audio Pre-Training with Acoustic Tokenizers	Transformer model pretrained with acoustic tokenizers	98.10%
HTS-AT: A Hierarchical Token-Semantic Audio Transformer for Sound Classification and Detection	Transformer model with hierarchical structure and token-semantic modules	97.00%
CLAP: Learning Audio Concepts From Natural Language Supervision	CNN model pretrained by natural language supervision	96.70%
AST: Audio Spectrogram Transformer	Pure Attention Model Pretrained on AudioSet	95.70%
Connecting the Dots between Audio and Text without Parallel Data through Visual Knowledge Transfer	A Transformer model pretrained w/ visual image supervision	95.70%
A Sequential Self Teaching Approach for Improving Generalization in Sound Event Recognition	Multi-stage sequential learning with knowledge transfer from Audioset	94.10%
Efficient End-to-End Audio Embeddings Generation for Audio Classification on Target Applications	CNN model pretrained on AudioSet	92.32%
Urban Sound Tagging using Multi-Channel Audio Feature with Convolutional Neural Networks	Pretrained model with multi-channel features	89.50%
An Ensemble of Convolutional Neural Networks for Audio Classification	CNN ensemble with data augmentation	88.65%

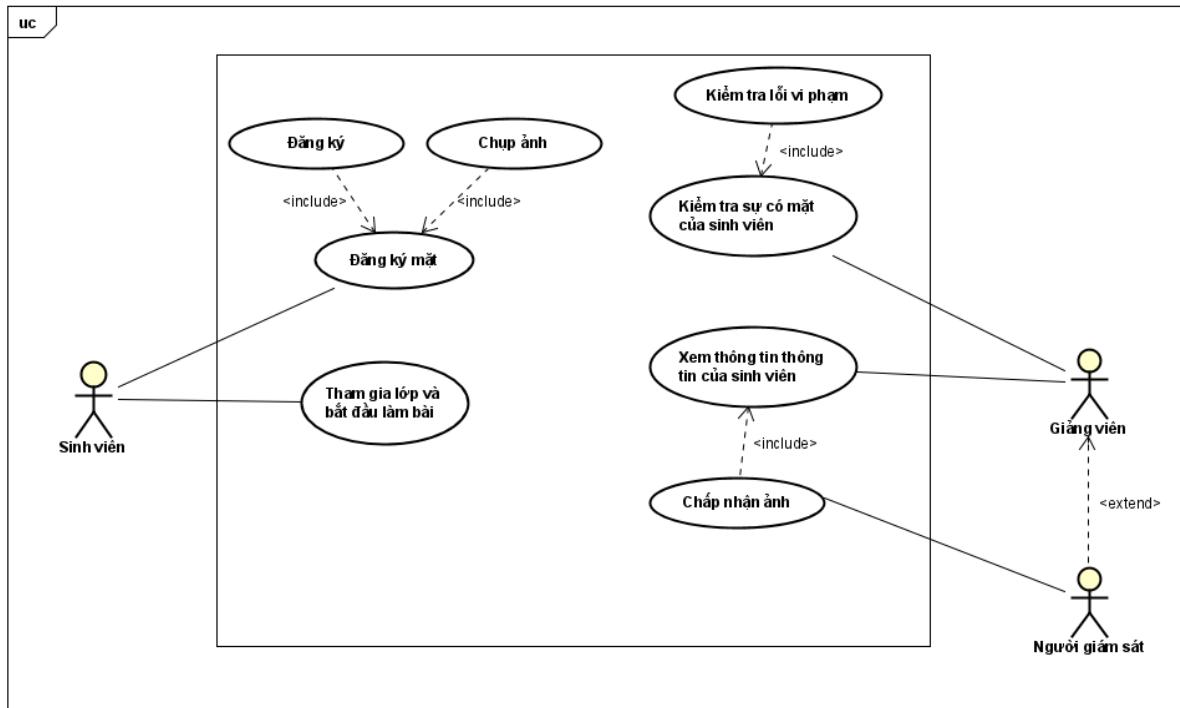
Hình 2.7: Ví dụ các mô hình

Các mô hình trên đều có độ chính xác cao tuy nhiên thời gian huấn luyện lâu, cũng như khá nặng mà yêu cầu nhiệm vụ đề ra là xây dựng vài toán phát hiện tiếng nói trên trình duyệt nên yêu cầu một mô hình nhẹ vs độ chính xác cao để dễ dàng triển khai trên trình duyệt mà không thông qua máy chủ.

2.3 Tổng quan bài toán điểm danh

Phần này tôi xin trình bày tổng quan bài toán điểm danh

2.3.1 Biểu đồ use case bài toán điểm danh



Hình 2.8: UseCase tổng quan

Tính năng bao gồm 3 tác nhân chính là Sinh viên , Giảng viên và Người giám sát.

Tác nhân Sinh viên có 2 chức năng chính:

- Chụp 5 bức ảnh gồm (thẳng, trái, phải, trên , dưới) để đăng ký khuôn mặt với hệ thống
- Tham gia vào lớp thi và bắt đầu làm bài thi

Tác nhân Giảng viên có 2 chức năng chính:

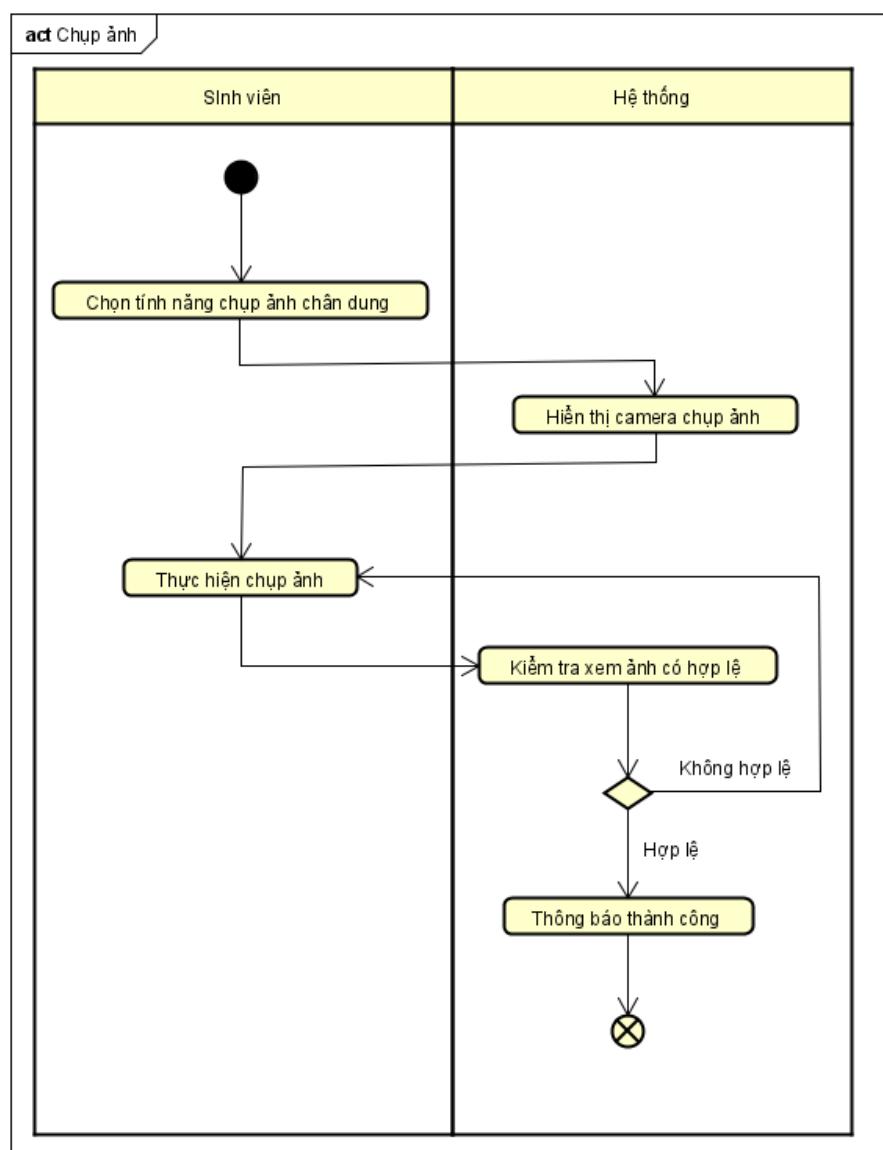
- Xem thông tin chi tiết của từng sinh viên mỗi lớp
- Xác nhận sự có mặt của sinh viên trong từng lớp thi

Tác nhân Người giám sát có 1 chức năng chính:

- Xem và xác nhận thông tin chi tiết của từng sinh viên mỗi lớp

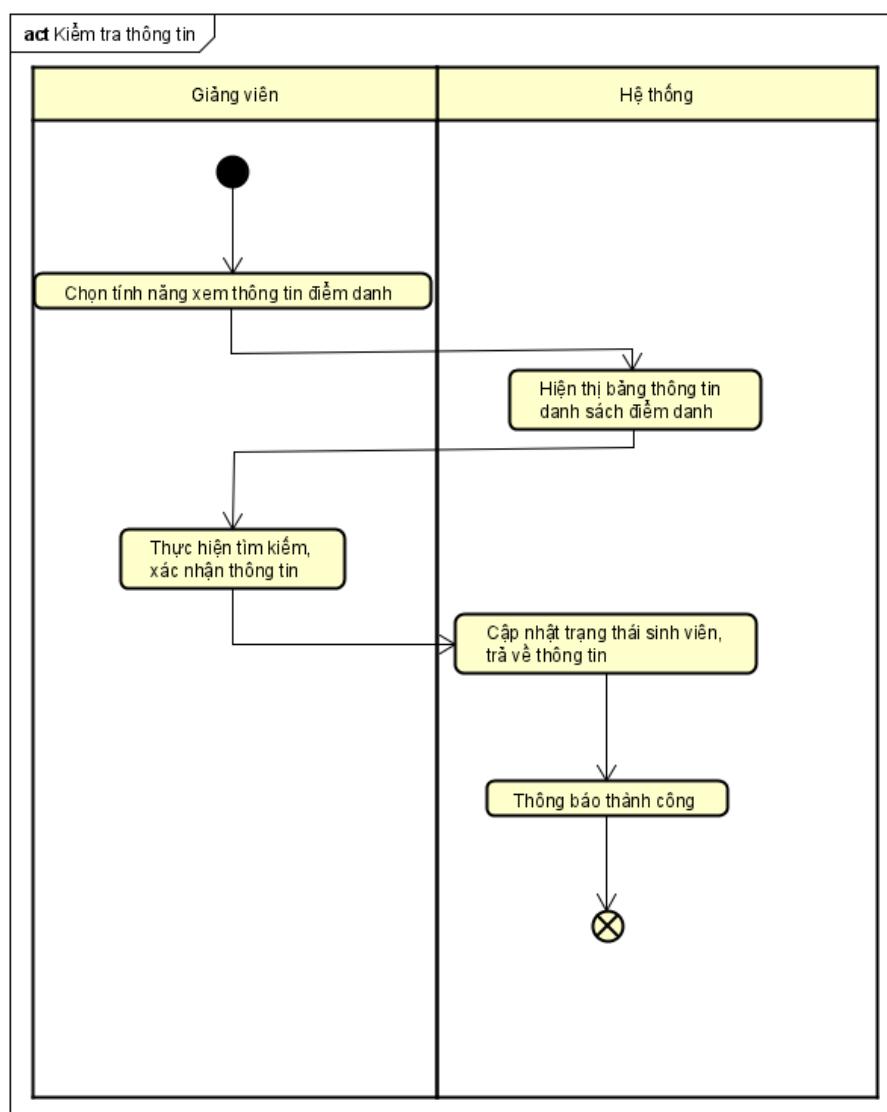
2.3.2 Quy trình nghiệp vụ

a, Quy trình nghiệp vụ chụp chân dung



Hình 2.9: Quy trình nghiệp vụ chụp chân dung

b, Quy trình nghiệp vụ xác nhận thông tin



Hình 2.10: Quy trình nghiệp vụ xác nhận thông tin

2.3.3 Đặc tả usecase

a, Đặc tả ca sử dụng chụp ảnh của sinh viên

Mã usecase	UC001	Tên use case	Chụp ảnh khuôn mặt
Tác Nhân	Sinh viên, Giám sát viên		
Tiền điều kiện	Sinh viên đã đăng ký vào hệ thống EdTech		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Sinh viên	Chụp 5 ảnh các hướng mặt theo hướng dẫn
	2	Hệ thống	Thực hiện huấn luyện mô hình và lưu ảnh
	3	Giám sát viên	Xác nhận ảnh sinh viên chụp hợp lệ
	4	Hệ thống	Thay đổi trạng thái ảnh là Thành công đối với sinh viên
Các luồng sự kiện thay thế	STT	Thực hiện bởi	Hành động
	1	Proctoring	Xác nhận ảnh sinh không đạt về ghi lại lí do
	2	Hệ thống	Lưu lại trạng thái Không đạt cùng ghi chú
	3	Hệ thống	Thông báo lại cho sinh viên
	3	Sinh viên	Chụp lại 5 ảnh các hướng mặt theo hướng dẫn
Điều kiện sau	Không		

b, Đặc tả usecase theo dõi thi của giảng viên

Mã usecase	UC001	Tên use case	Theo dõi quá trình thi
Tác Nhân	Sinh viên, Giảng viên		
Tiền điều kiện	Sinh viên đang chuẩn bị làm bài thi		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Sinh viên	Tham gia lớp thi
	2	Hệ thống	Thực hiện quá trình điểm danh
	3	Giảng viên	Quan sát quá tình trạng có mặt của sinh viên
	4	Giảng viên	Quyết định thời điểm làm bài
	5	Sinh viên	Làm bài
	6	Hệ thống	Ghi lại các lỗi của sinh viên
	7	Giảng viên	Quan sát trực quan các lỗi
Các luồng sự kiện thay thế	Không		
Điều kiện sau	Không		

2.4 Yêu cầu phi chức năng

2.4.1 Yêu cầu trải nghiệm người dùng

Các giao diện phải bắt mắt, thân thiện, màu sắc phải cùng với màu chủ đạo của hệ thống EdTech. Giao diện có hỗ trợ responsive phù hợp với nhiều loại kích thước màn

hình trên các thiết bị khác nhau

2.4.2 Yêu cầu về bảo mật và tải đường truyền

Máy chủ phải đáp ứng được 1 lượng lớn các yêu cầu cùng lúc gửi về và xử lí. Cũng như chỉ những người dùng có quyền mới có quyền xem thông tin sinh viên

2.4.3 Yêu cầu khả năng mở rộng và nâng cấp, bảo trì

Hệ thống gồm các tính năng được đóng gói riêng nên dễ dàng bảo trì và nâng cấp lên phiên bản tốt hơn.

Chương 3, tôi xin trình bày về các công nghệ tôi đã sử dụng.

CHƯƠNG 3. CÔNG NGHỆ SỬ DỤNG

Ở chương này, tôi sẽ trình bày về các công nghệ đã dùng trong đồ án cũng như mục đích, lý do tôi dùng nó đó trong đồ án này.

3.1 JavaScript

JavaScript là một ngôn ngữ lập trình kịch bản được sử dụng để tạo các trang web tương tác. Nó được Brendan Eich tạo ra vào năm 1995 và là một trong ba ngôn ngữ chính được sử dụng để tạo các trang web, cùng với HTML và CSS.

JavaScript được sử dụng để thêm các tính năng động vào các trang web, chẳng hạn như:

- Tương tác với người dùng thông qua các sự kiện như nhấp chuột, di chuột và nhập văn bản.
- Hiển thị nội dung động, chẳng hạn như biểu đồ, đồ họa và video.
- Tạo trò chơi và ứng dụng web.
- JavaScript là một ngôn ngữ dễ học và sử dụng, và có rất nhiều tài nguyên trực tuyến có sẵn để giúp bạn học cách sử dụng nó.

Dưới đây là một số lợi ích của việc sử dụng JavaScript:

- JavaScript là một ngôn ngữ được sử dụng rộng rãi, có nghĩa là có rất nhiều tài nguyên trực tuyến có sẵn để giúp bạn học cách sử dụng nó.
- JavaScript là một ngôn ngữ dễ học và sử dụng, ngay cả đối với những người mới bắt đầu lập trình.
- JavaScript có thể được sử dụng để tạo các trang web tương tác và hấp dẫn.
- JavaScript có thể được sử dụng để tạo trò chơi và ứng dụng web.

3.1.1 Lí do lựa chọn

- Các module x-block của hệ thống MOOC chỉ nhận các tệp js, html, css làm đầu vào
- Có nhiều thư viện hỗ trợ xử lý bài toán được đề ra

3.1.2 Vai trò trong đồ án

Trong đồ án này, JS có vai trò giúp tôi triển khai các logic xử lí trên phía giao diện người dùng ,gửi API cho server cũng như xử lí các logic tính toán trong quá trình làm việc với tệp audio.

3.2 SpringBoot

Spring Boot là một thư viện Java miễn phí và mã nguồn mở được thiết kế để giúp việc phát triển các ứng dụng Spring trở nên dễ dàng hơn. Nó cung cấp một số tính năng giúp giảm bớt gánh nặng của việc phát triển các ứng dụng Spring, bao gồm:

Tự động cấu hình: Spring Boot tự động cấu hình nhiều thành phần Spring dựa trên các thuộc tính bạn cung cấp. Điều này giúp bạn tránh phải viết mã cấu hình thủ công. Tự động khởi động: Spring Boot tự động khởi động các ứng dụng Spring khi bạn chạy chúng. Điều này giúp bạn tránh phải viết mã khởi động thủ công. Cấu hình tích hợp: Spring Boot tích hợp với nhiều công cụ và dịch vụ phổ biến, chẳng hạn như Tomcat, Jetty và WildFly. Điều này giúp bạn tránh phải cài đặt và định cấu hình các công cụ này theo cách thủ công. Các mẫu: Spring Boot cung cấp một số mẫu ứng dụng để giúp bạn bắt đầu nhanh chóng. Các mẫu này bao gồm các ứng dụng web, ứng dụng REST và ứng dụng máy chủ. Spring Boot là một thư viện rất mạnh mẽ và linh hoạt có thể được sử dụng để phát triển nhiều loại ứng dụng Java. Nó là một lựa chọn tuyệt vời cho các nhà phát triển Java muốn phát triển các ứng dụng Spring một cách nhanh chóng và dễ dàng.

Dưới đây là một số lợi ích của việc sử dụng Spring Boot:

- Nhanh chóng và dễ dàng: Spring Boot giúp việc phát triển các ứng dụng Spring trở nên nhanh chóng và dễ dàng hơn. Nó tự động cấu hình nhiều thành phần Spring và tự động khởi động các ứng dụng Spring khi bạn chạy chúng.
- Tích hợp: Spring Boot tích hợp với nhiều công cụ và dịch vụ phổ biến, giúp bạn tránh phải cài đặt và định cấu hình các công cụ này theo cách thủ công.
- Mạnh mẽ và linh hoạt: Spring Boot là một thư viện rất mạnh mẽ và linh hoạt có thể được sử dụng để phát triển nhiều loại ứng dụng Java.
- Được cộng đồng hỗ trợ tốt: Spring Boot là một thư viện được cộng đồng hỗ trợ tốt, với nhiều tài nguyên trực tuyến có sẵn để giúp bạn học cách sử dụng nó.

3.2.1 Lí do lựa chọn

- Được yêu cầu sử dụng
- Có cộng đồng lớn để dễ dàng hỏi đáp những lỗi thường gặp.
- Yêu thích ngôn ngữ Java

3.2.2 Vai trò trong đồ án

Trong đồ án này, Spring Boot giúp tôi xây dựng được một server dùng để xử lý các logic nghiệp vụ cũng như giao tiếp với các cơ sở dữ liệu tương ứng

3.3 MySQL

MySQL là một hệ quản trị cơ sở dữ liệu quan hệ (RDBMS) mã nguồn mở, miễn phí và phổ biến nhất trên thế giới. Nó được phát triển bởi Oracle Corporation và được sử dụng bởi hàng triệu người dùng trên toàn thế giới. MySQL là một lựa chọn tuyệt vời cho các ứng dụng web, ứng dụng di động và các ứng dụng doanh nghiệp.

MySQL có một số tính năng nổi bật, bao gồm:

- Miễn phí và mã nguồn mở
- Dễ sử dụng
- Tốc độ nhanh
- Độ tin cậy cao
- Tính linh hoạt cao
- Được hỗ trợ bởi cộng đồng lớn

3.3.1 Lý do lựa chọn

- Nhẹ, dễ dàng cài đặt, triển khai để lưu trữ truy xuất dữ liệu
- Đã có kiến thức từ trước

3.3.2 Vai trò trong đồ án

MySQL có thể được sử dụng trên nhiều nền tảng, bao gồm Windows, Linux, macOS và Unix. Nó có thể được cài đặt trên máy chủ hoặc trên máy tính để bàn. MySQL cũng có thể được sử dụng trong đám mây.

Trong đồ án này, MySQL là một cơ sở dữ liệu nhẹ, dễ triển khai, sử dụng làm nơi lưu trữ các thông tin của sinh viên cũng như tình trạng học hoặc ảnh và các lỗi vi phạm trong quá trình làm bài.

3.4 Python

Python là một ngôn ngữ lập trình đa năng, được thiết kế để dễ đọc và viết. Nó được Guido van Rossum tạo ra vào năm 1991 và đã trở thành một trong những ngôn ngữ lập trình phổ biến nhất trên thế giới. Python được sử dụng cho nhiều mục đích, bao gồm:

- Phát triển web
- Khoa học dữ liệu
- Học máy
- Tự động hóa
- Phát triển trò chơi

- Nghiên cứu

Python là một ngôn ngữ mạnh mẽ và linh hoạt, có thể được sử dụng để phát triển nhiều loại ứng dụng. Nó là một lựa chọn tuyệt vời cho người mới bắt đầu lập trình, cũng như cho các nhà phát triển có kinh nghiệm.

Dưới đây là một số lợi ích của việc sử dụng Python:

- Dễ học và sử dụng
- Có nhiều tài nguyên trực tuyến
- Được sử dụng rộng rãi trong cộng đồng
- Có nhiều thư viện và công cụ
- Nhanh và hiệu quả
- Linh hoạt và có thể mở rộng

3.4.1 Lí do lựa chọn

- Là một ngôn ngữ dễ dàng xây dựng mô hình
- Có nhiều thư viện hỗ trợ tiền xử lý âm thanh, xử lý ảnh mà ngôn ngữ khác không có

3.4.2 Vai trò trong đồ án

Trong đồ án này, Python cung cấp cho tôi các thư viện như pandas, tensorflow, librosa,... để dễ dàng tiếp cận và xử lý các dữ liệu âm thanh và dễ dàng trong quá trình xây dựng và thực hiện các quá trình xây dựng và huấn luyện các model học máy

3.5 Docker

Docker là một nền tảng mã nguồn mở cho phép bạn đóng gói và chạy các ứng dụng trong các môi trường cách ly. Các môi trường này được gọi là container, và chúng cung cấp một cách để chạy các ứng dụng một cách nhất quán trên bất kỳ máy chủ nào, bất kể hệ điều hành hoặc cấu hình phần cứng.

Docker hoạt động bằng cách tạo một hình ảnh của ứng dụng của bạn. Hình ảnh là một tập hợp các tệp và thư mục cần thiết để chạy ứng dụng của bạn, bao gồm mã, thư viện và các thành phần phụ thuộc khác. Khi bạn đã tạo hình ảnh, bạn có thể sử dụng nó để tạo một container. Container là một phiên bản của hình ảnh đang chạy.

Container được cách ly với nhau, có nghĩa là chúng không thể truy cập vào tài nguyên của nhau. Điều này giúp ngăn chặn xung đột và đảm bảo rằng các ứng dụng của bạn chạy ổn định. Container cũng có thể được triển khai nhanh chóng và dễ dàng, điều này khiến chúng trở nên lý tưởng cho việc phát triển và triển khai ứng dụng.

Docker là một công cụ mạnh mẽ có thể giúp bạn phát triển và triển khai ứng dụng

một cách nhanh chóng và dễ dàng. Nó là một lựa chọn tuyệt vời cho các nhà phát triển, quản trị viên hệ thống và bất kỳ ai khác cần chạy các ứng dụng trên các máy chủ khác nhau.

Dưới đây là một số lợi ích của việc sử dụng Docker:

- Tính nhất quán: Docker giúp bạn chạy các ứng dụng một cách nhất quán trên bất kỳ máy chủ nào, bất kể hệ điều hành hoặc cấu hình phần cứng.
- Tính di động: Container có thể được triển khai nhanh chóng và dễ dàng, điều này khiến chúng trở nên lý tưởng cho việc phát triển và triển khai ứng dụng.
- Tính cách ly: Container được cách ly với nhau, có nghĩa là chúng không thể truy cập vào tài nguyên của nhau. Điều này giúp ngăn chặn xung đột và đảm bảo rằng các ứng dụng của bạn chạy ổn định.
- Tính khả mở: Docker là một nền tảng mã nguồn mở, có nghĩa là nó có thể được tùy chỉnh và mở rộng theo nhu cầu của bạn.

3.5.1 Lí do lựa chọn

- Phù hợp với hệ thống cần triển khai
- Dễ sử dụng

3.5.2 Vai trò trong đồ án

Trong đồ án lần này, Docker giúp tôi trong việc giúp tôi triển khai máy chủ và cơ sở dữ liệu mà không gặp nhiều vấn đề trong quá trình triển khai.

Ở chương tiếp, tôi xin trình bày về tổng quan chi tiết các bài toán, từ thiết kế đến phần triển khai cũng như kết quả đạt được của các bài toán.

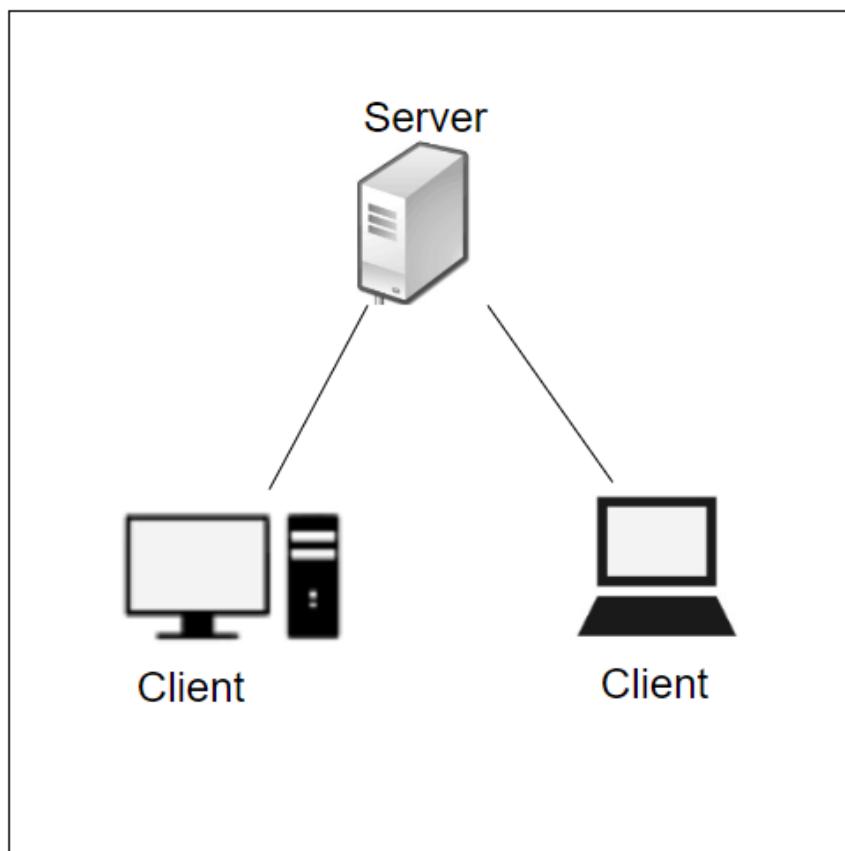
CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Tiếp theo nội dung chương 3, chương 4 tôi xin trình bày về chi tiết thiết kế các tính năng, vẽ các biểu đồ trực quan của bài toán điểm danh sinh viên ở mục 4.1 và thiết kế, cách triển khai bài toán phát hiện tiếng nói cụ thể trong mục 4.2 và kết quả đạt được của cả hai bài toán.

4.1 Thiết kế kiến trúc luồng điểm danh

4.1.1 Lựa chọn kiến trúc phần mềm

Mô hình khách hàng-máy chủ (client-server) là một mô hình mạng trong đó các máy khách yêu cầu các dịch vụ từ các máy chủ. Các máy khách có thể là máy tính, điện thoại thông minh, máy tính bảng hoặc bất kỳ thiết bị nào khác có thể kết nối với mạng. Các máy chủ là các máy tính mạnh hơn được thiết kế để lưu trữ dữ liệu và cung cấp các dịch vụ cho các máy khách. Hình 4.1 thể hiện mô hình khách- chủ như sau:

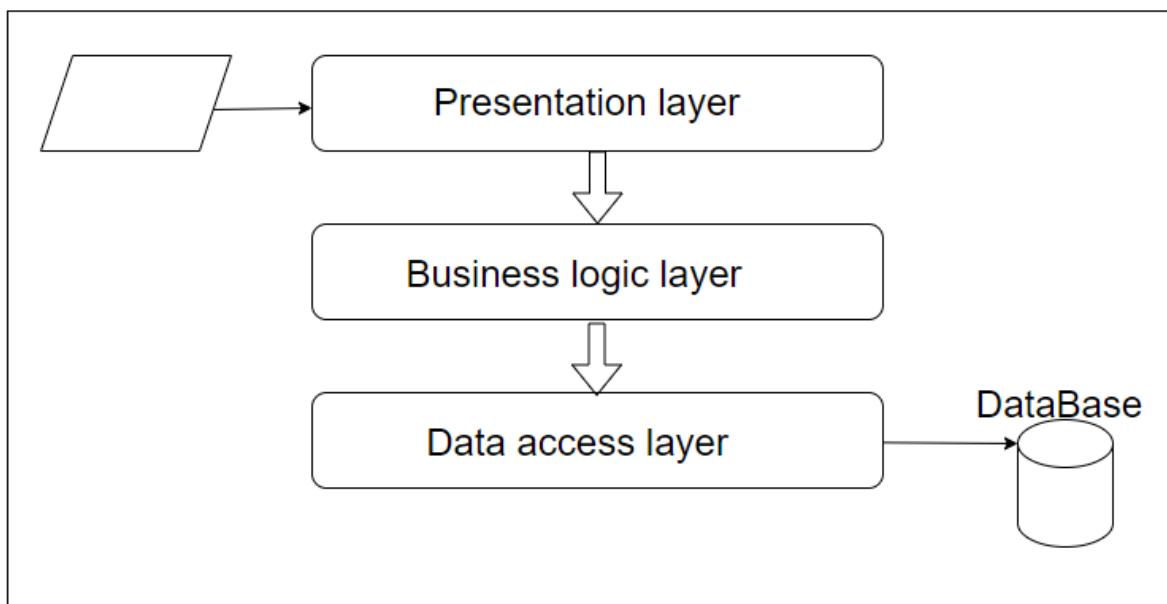


Hình 4.1: Mô hình khách -chủ

Khi có một yêu cầu từ phía máy khách đến máy chủ, bộ phận điều hướng trong máy chủ sẽ tiếp nhận yêu cầu và tìm kiếm dữ liệu trong cơ sở dữ liệu để trả về kết quả phù hợp. Còn ở máy chủ, hệ thống được xây dựng theo mô hình 3 lớp (three-tier). Hình 4.2

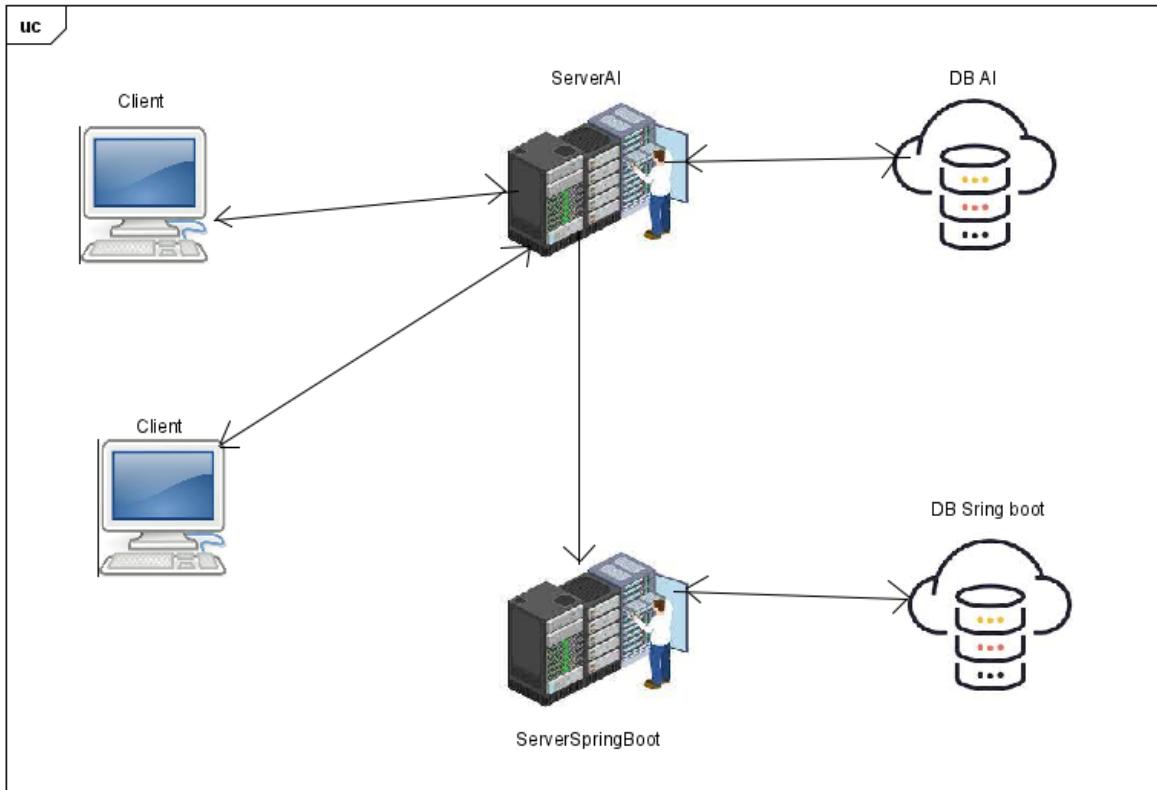
thể hiện mô hình 3 lớp như sau:

- Presentation layer: tầng này tương tác với người dùng bằng API (Controller)
- Business logic layer: chứa các business logic code (Servise)
- Data access layer: Tương tác với database, trả về kết quả cho tầng business logic (Repository)



Hình 4.2: Mô hình MVC trong máy chủ

4.1.2 Thiết kế tổng quan

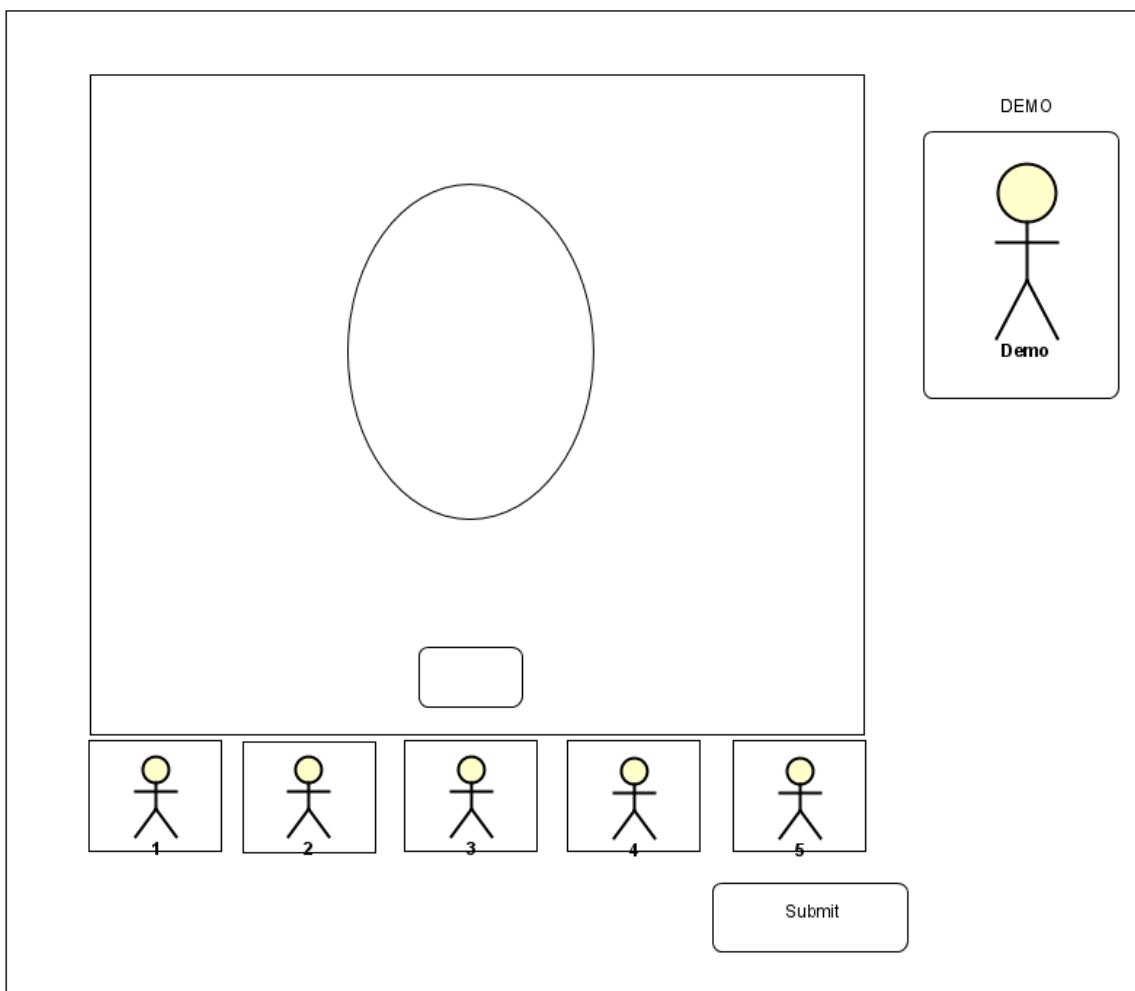


Hình 4.3: Tổng quan luồng tính năng

Mỗi máy khách khi thực hiện chụp ảnh, sẽ gửi 1 request HTTPS với payload gồm các trường thông tin sinh viên và 5 file ảnh tới máy chủ "máy chủ AI" để đăng ký khuôn mặt cho hệ thống, và máy chủ AI sau khi huấn luyện mô hình sẽ ra 1 file weight, sau đó dữ liệu hình ảnh sẽ được lưu lại trên máy chủ AI và lưu lại đường dẫn với máy chủ Spring boot. Tương tự luồng trên, mỗi khi có lớp làm bài kiểm tra, sinh viên sẽ tham gia trực tuyến vào hệ thống, lần lượt thực hiện điểm danh với hệ thống AI, mã trạng thái của sinh viên lúc này sẽ được ghi tạm vào cơ sở dữ liệu của máy chủ Spring boot. Sau đó, trong quá trình làm bài thi, bất cứ khi sinh viên nào có thể hiện gian lận sẽ được máy chủ AI lưu lại video ở cơ sở dữ liệu AI và gửi tất cả dữ liệu cùng đường dẫn Video cho cơ sở dữ liệu của máy chủ Spring boot.

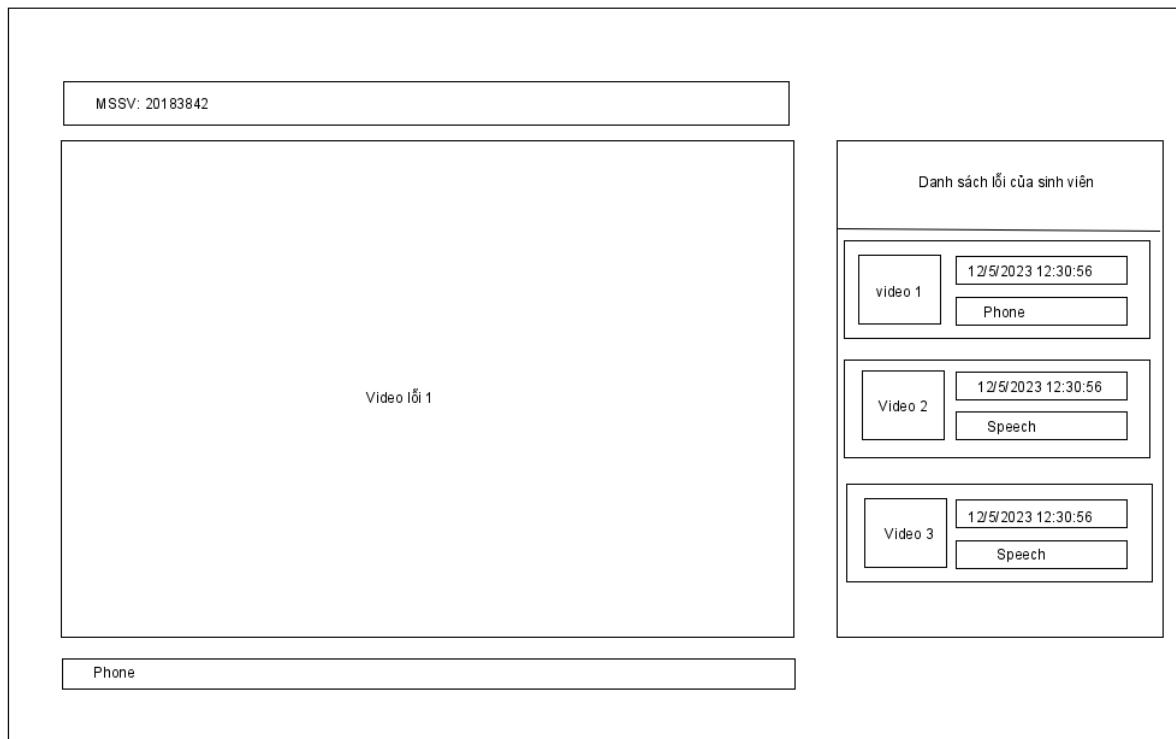
4.1.3 Thiết kế chi tiết giao diện

Giao diện chụp ảnh sinh viên như Hình 4.4:



Hình 4.4: Giao diện chụp ảnh

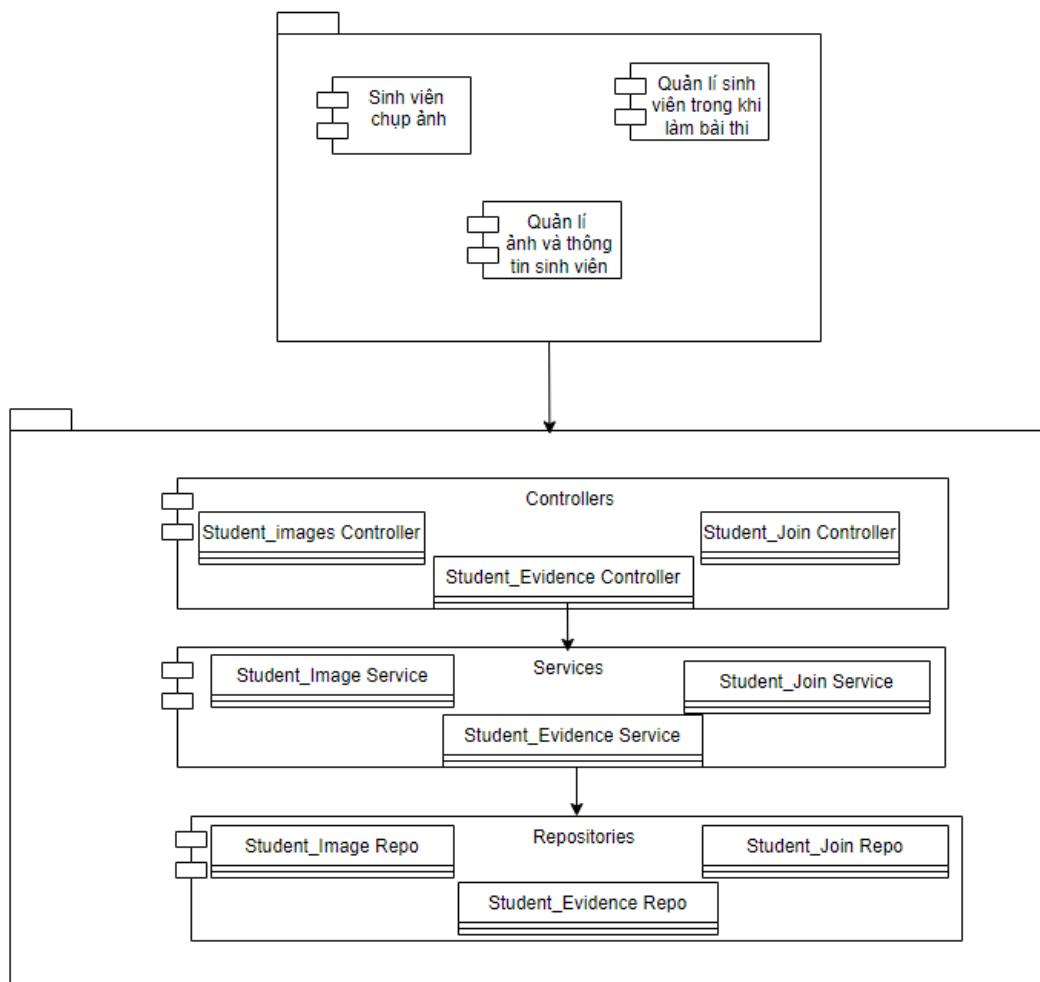
Giao diện xem các lỗi vi phạm của sinh viên như Hình 4.5:



Hình 4.5: Giao diện xem lỗi vi phạm

Giao diện người dùng của hệ thống hướng đến sự linh hoạt và dễ sử dụng đối với người dùng. Các thông tin quan trọng luôn phải đảm bảo phải được hiển thị đầy đủ với kích thước hiển thị đối với từng loại dữ liệu có chiều dài và chiều rộng hợp lý, không vượt quá kích thước màn hình.

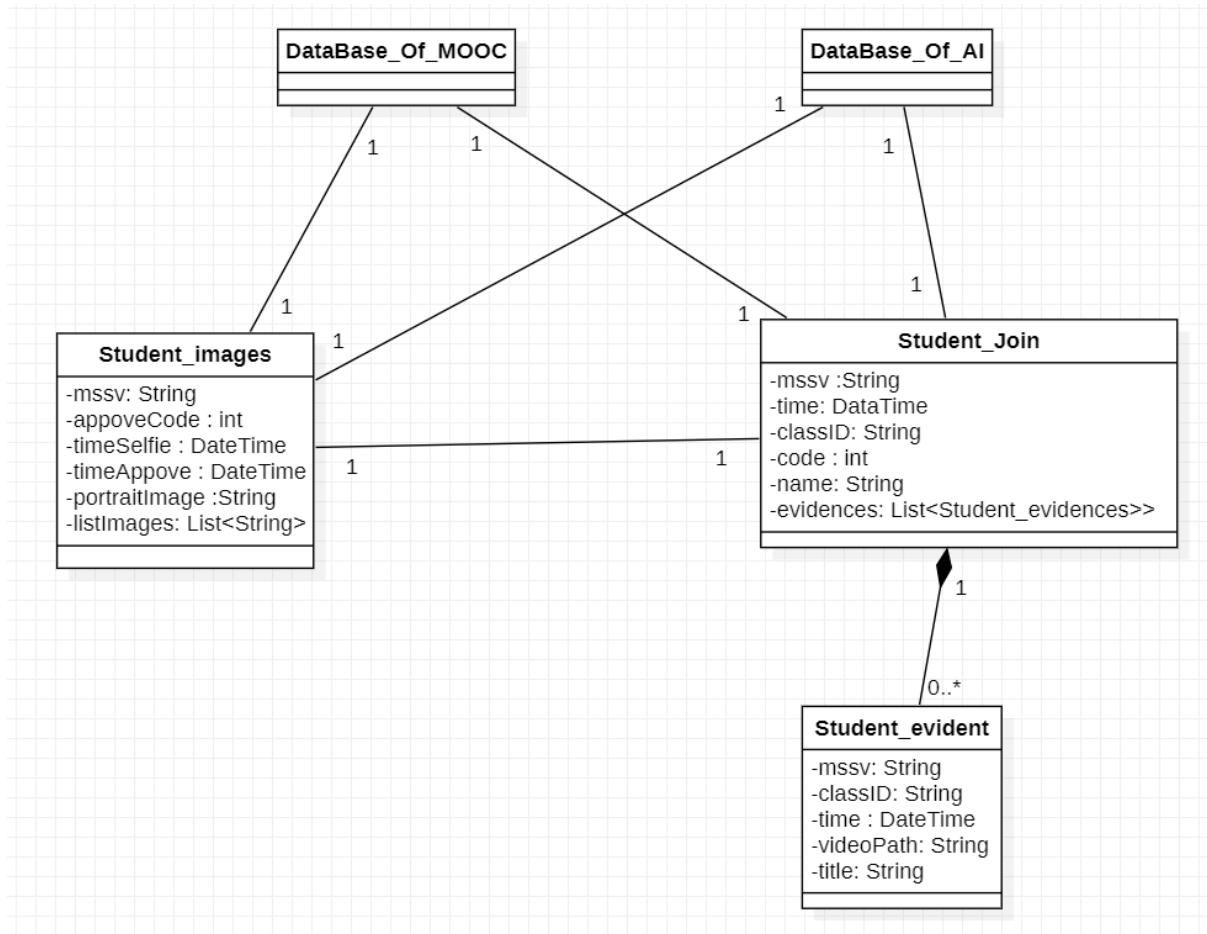
4.1.4 Thiết kế chi tiết gói



Hình 4.6: Thiết kế chi tiết gói

4.1.5 Thiết kế cơ sở dữ liệu

Sơ đồ thiết kế cơ sở dữ liệu



Hình 4.7: Thiết kế cơ sở dữ liệu

Khi sinh viên chụp ảnh, dữ liệu sẽ được gửi cho phía máy chủ AI để tiến hành huấn luyện mô hình nhận diện khuôn mặt. Sau đó dữ liệu sẽ được lưu một phần ở máy chủ AI(gồm các ảnh) và lưu toàn bộ thông tin sinh viên ở máy chủ Spring boot. Cũng như vậy, trong quá trình làm bài thi của sinh viên, luồng dữ liệu video sẽ được đưa về máy chủ AI để xử lí, đánh giá, sau quá trình dự đoán, dữ liệu sẽ được lưu một phần ở máy chủ AI(gồm các video) và lưu toàn bộ thông tin ở máy chủ Spring boot.

a, Thiết kế Bảng " Student-image"

STT	Tên trường	Kiểu dữ liệu	Mô tả	Ràng buộc
1	mssv	String	Mã số sinh viên	Khóa chính
2	timeSelfie	DateTime	Thời gian chụp ảnh	
3	note	String	Ghi chú về ảnh	
4	timeAppove	DateTime	Thời gian đánh giá	
5	Code	Integer	Mã ảnh	
6	PortraitImage	String	Đường dẫn ảnh	
7	ListImages	List<String>	Tập 5 ảnh	

b, Thiết kế Bảng " Student-Join"

STT	Tên trường	Kiểu dữ liệu	Mô tả	Ràng buộc
1	mssv	String	Mã số sinh viên	Khóa chính
2	classID	String	Mã số lớp học	Khóa chính
3	timeJoin	Datetime	Thời gian tham gia	
4	name	String	Tên sinh viên	
5	code	Integer	Mã tham gia	

c, Thiết kế Bảng " Student-Evidence"

STT	Tên trường	Kiểu dữ liệu	Mô tả	Ràng buộc
1	id	Long	Mã bản ghi	Khóa chính
2	mssv	String	Mã số sinh viên	Khóa phụ
3	classID	String	Mã số lớp học	Khóa phụ
4	time	Datetime	Thời gian tạo lỗi	
5	pathVideo	String	Đường dẫn lỗi	
6	title	String	Thông tin lỗi	

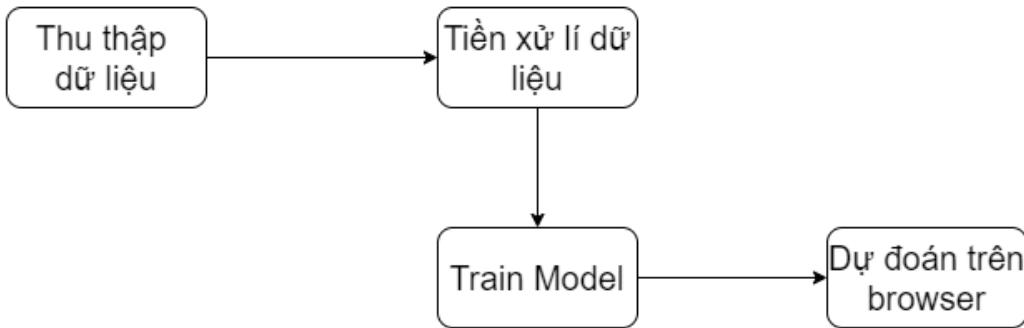
4.2 Bài toán phát hiện tiếng nói

Hiện nay hệ thống phát hiện gian lận của daotao.ai đã được phát triển, tuy nhiên đó chỉ là những tính về xử lý hình ảnh: nhận dạng khuôn mặt, nhận dạng thiết bị.... chưa có bài toán về phát hiện gian lận bằng âm thanh được triển khai trên đó. Vì vậy, yêu cầu được đưa ra là một tính năng có thể phát hiện được tiếng nói trong quá trình thi của sinh viên và đưa ra thông báo cho họ biết họ đang vi phạm.

Đầu tiên, tôi xin nhắc lại khái niệm âm thanh như sau: âm thanh là một trong những thông tin giác quan chính mà chúng ta nhận được để cảm nhận môi trường của mình. Hầu hết mọi hành động hoặc sự kiện xung quanh chúng ta đều có âm thanh độc đáo. Âm thanh có 3 thuộc tính chính giúp chúng ta phân biệt giữa hai âm thanh.

- Biên độ - Độ to của âm thanh
- Tần số - Cao độ của âm thanh
- Âm sắc — Chất lượng âm thanh hoặc bản sắc của âm thanh (ví dụ: Sự khác biệt về âm thanh giữa đàn piano và đàn violon)

Mà hiện nay âm thanh không thể trực tiếp là đầu vào của các mô hình học máy, nên ta cần phải chuyển nó thành các ảnh mel spectrogram(sự thể hiện của thời gian, độ lớn tần số) tương ứng để đưa vào các mô hình. Dưới đây là hình 3.1 trình bày khái quát tổng quan về các khái niệm trong bài toán xử lí, phát hiện tiếng nói:

**Hình 4.8:** Tổng quan các khối

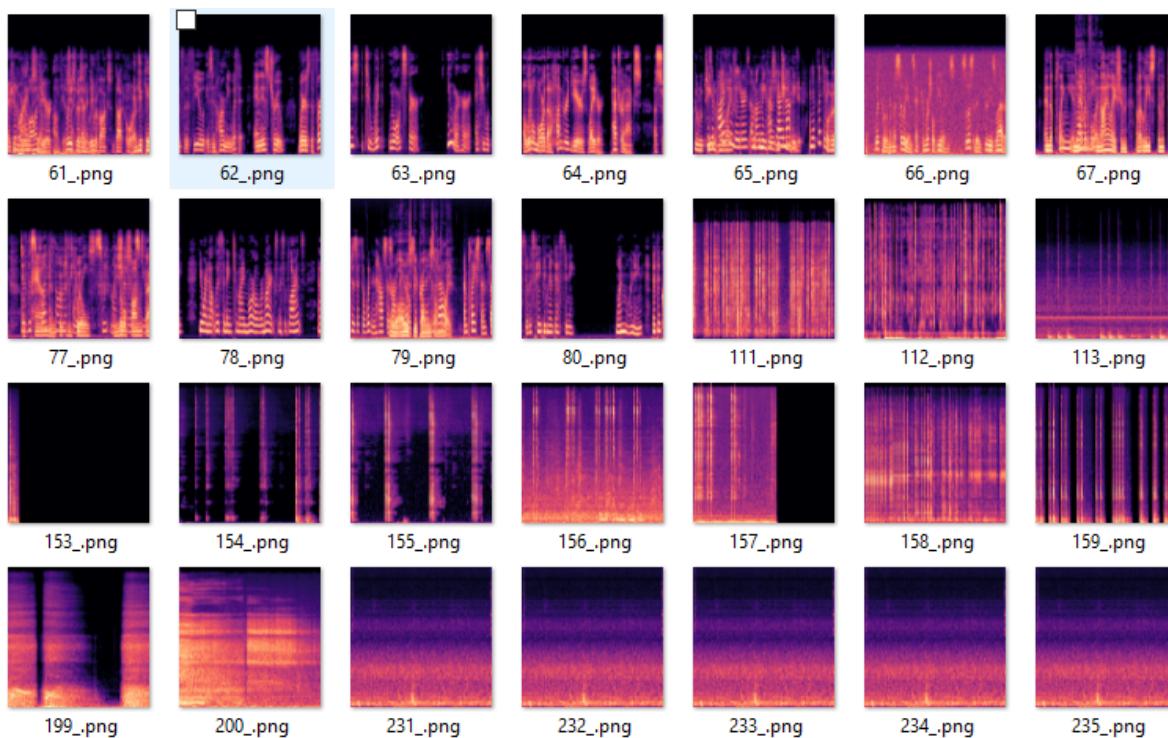
Mô hình được khái quát làm ba khối chính: Khối dữ liệu, Khối huấn luyện và Khối xử lý trên trình duyệt

4.2.1 Khối dữ liệu

Dữ liệu đầu vào của bài toán là các file âm thanh dạng .wav được trích xuất một phần ở tập dataset esc-50 với dữ liệu kiểu MONO (âm thanh 1 kênh). Âm thanh từ tệp được tải vào một mảng hình dạng Numpy (num-channels, num-samples). Hầu hết âm thanh được lấy mẫu ở 44,1kHz và có thời lượng khoảng 4 giây, dẫn đến $44.100 \times 4 = 176.400$ mẫu. Nếu âm thanh có 1 kênh, hình dạng của mảng sẽ là (1, 176,400). Tương tự, âm thanh có thời lượng 4 giây với 2 kênh và được lấy mẫu ở 48kHz sẽ có 192.000 mẫu và hình dạng là (2, 192.000).

Loại âm thanh	Số lượng mẫu
Tiếng nói (Speech)	50
Sự im lặng (Silence)	50
Tiếng gõ bàn phím (Keyboard-typing)	50
Tiếng mở tài liệu (Open-book)	50
Tiếng ho (Coughing)	50
Tiếng click chuột	50

Điều đầu tiên chúng ta cần là đọc và tải tệp âm thanh ở định dạng “.wav”. Sau đó thư viện librosa sẽ được chuyển về dưới các ảnh dạng mel spectrogram làm dữ liệu đầu vào cho Khối huấn luyện mô hình. Hình 4.9 thể hiện ảnh các tệp âm thanh đã được Mel spectrogram hóa



Hình 4.9: Chuyển file audio thành các mel spectrogram

4.2.2 Khối huấn luyện mô hình

Sau khi nhận được các tệp ảnh mel spectrogram cùng các tệp csv đánh nhãn tương ứng. Tôi sẽ bắt đầu quá trình huấn luyện mô hình. Trong khối này, tôi đã tìm hiểu và nghiên cứu theo 3 hướng:

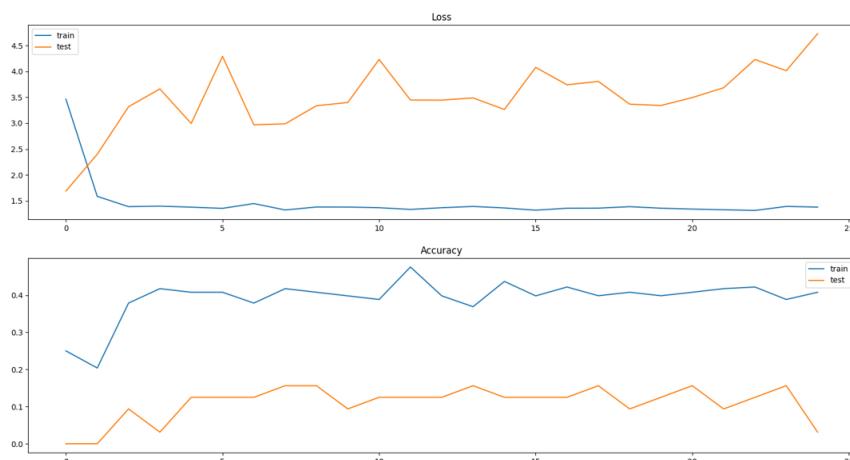
a, Mô hình có sẵn VGG-16

Có nhiều mô hình với độ chính xác cao đã được xây dựng từ trước như LSTM, GVV-16... Hình 4.10 thể hiện cấu trúc mạng của mô hình VGG-16:

conv2d_6 (Conv2D)	(None, 56, 56, 256)	590080
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 256)	0
conv2d_7 (Conv2D)	(None, 28, 28, 512)	1180160
conv2d_8 (Conv2D)	(None, 28, 28, 512)	2359808
conv2d_9 (Conv2D)	(None, 28, 28, 512)	2359808
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 512)	0
conv2d_10 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_11 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_12 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_4 (MaxPooling2D)	(None, 7, 7, 512)	0
dropout (Dropout)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 4096)	102764544
dropout_1 (Dropout)	(None, 4096)	0
dense_1 (Dense)	(None, 4096)	16781312
dropout_2 (Dropout)	(None, 4096)	0
dense_2 (Dense)	(None, 5)	20485
<hr/>		
Total params: 134281029 (512.24 MB)		
Trainable params: 134281029 (512.24 MB)		
Non-trainable params: 0 (0.00 Byte)		

Hình 4.10: Cấu trúc mạng VGG-16

Với kết quả như Hình 4.11:



Hình 4.11: Độ chính xác và mất mát của cấu trúc mạng VGG-16

Từ kết quả trên, tôi thấy dữ liệu bị overfit và số lượng tham số của VGG-16 quá lớn mà yêu cầu đặt ra là một model nhỏ gọn.

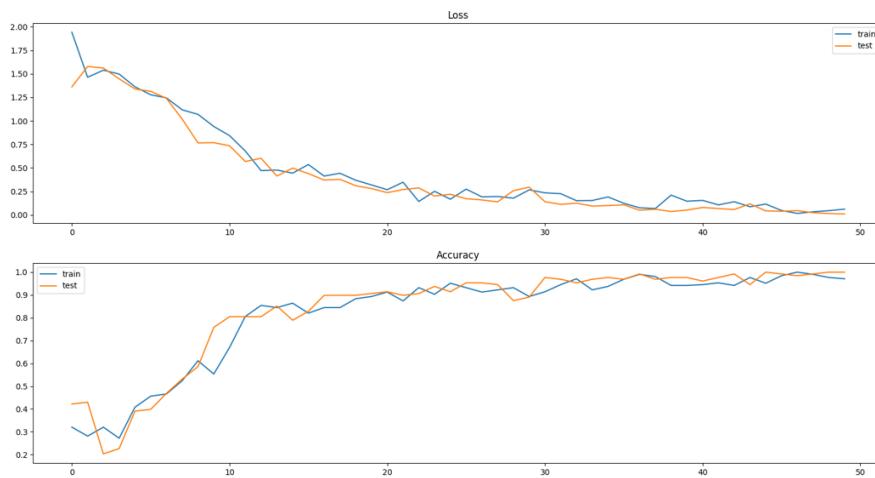
b, Mô hình CNN dựa trên VGG-16

Nhận thấy mô hình trên không phù hợp với bài toán, tôi đã triển khai xây dựng một mô hình với số lượng tham số nhỏ hơn nhiều so với mô hình VGG-16 (2677061 so với 134281029) do số lượng lớp Conv3D ít hơn. Hình 4.12 thể hiện cấu trúc mạng CNN:

max_pooling2d (MaxPooling2D)	(None, 31, 31, 64)	0
dropout (Dropout)	(None, 31, 31, 64)	0
conv2d_2 (Conv2D)	(None, 31, 31, 64)	36928
activation_2 (Activation)	(None, 31, 31, 64)	0
conv2d_3 (Conv2D)	(None, 29, 29, 64)	36928
activation_3 (Activation)	(None, 29, 29, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 64)	0
dropout_1 (Dropout)	(None, 14, 14, 64)	0
conv2d_4 (Conv2D)	(None, 14, 14, 128)	73856
activation_4 (Activation)	(None, 14, 14, 128)	0
conv2d_5 (Conv2D)	(None, 12, 12, 128)	147584
activation_5 (Activation)	(None, 12, 12, 128)	0
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 128)	0
dropout_2 (Dropout)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 512)	2359808
activation_6 (Activation)	(None, 512)	0
dropout_3 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 5)	2565
<hr/>		
Total params:	2677061 (10.21 MB)	
Trainable params:	2677061 (10.21 MB)	
Non-trainable params:	0 (0.00 Byte)	

Hình 4.12: Cấu trúc mạng CNN

Với kết quả thu được khả quan hơn như Hình 4.13: Chi tiết các chỉ số đánh giá sẽ được thể hiện rõ ở [8]:



Hình 4.13: Độ chính xác và mất mát của cấu trúc mạng CNN

c, Mô hình CNN của Google

Một mô hình tiếp cận khác của bài toán trên là dùng thư viện mã nguồn mở của Google [9] với cấu trúc mạng như sau:

CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Layer (type)	Output shape	Param #
=====		
conv2d_1_input (InputLayer)	[null, 43, 232, 1]	0
conv2d_1 (Conv2D)	[null, 42, 225, 8]	136
max_pooling2d_1 (MaxPooling2D)	[null, 21, 112, 8]	0
conv2d_2 (Conv2D)	[null, 20, 109, 32]	2080
max_pooling2d_2 (MaxPooling2D)	[null, 10, 54, 32]	0
conv2d_3 (Conv2D)	[null, 9, 51, 32]	8224
max_pooling2d_3 (MaxPooling2D)	[null, 4, 25, 32]	0
conv2d_4 (Conv2D)	[null, 3, 22, 32]	8224
max_pooling2d_4 (MaxPooling2D)	[null, 2, 11, 32]	0
flatten_1 (Flatten)	[null, 704]	0
dropout_1 (Dropout)	[null, 704]	0
dense_1 (Dense)	[null, 2000]	1410000
sequential_5 (Sequential)	multiple	12006
=====		
Total params: 1440670		
Trainable params: 12006		
Non-trainable params: 1428664		

Hình 4.14: Cấu trúc mạng CNN của Google

Sau khi triển khai thử nghiệm 3 hướng trên tôi thấy mô hình CNN của Google là phù hợp, nhẹ, dễ triển khai nhất trên trình duyệt.

Sau khi tiến hành huấn luyện mô hình phù hợp, tôi đã nhận được một file model.h5 thể hiện các tham số của cấu trúc mạng

4.2.3 Khởi triển khai trên trình duyệt

Để triển khai bài toán phân loại trên trình duyệt trước hết tôi phải dùng thư viện Tensorflowjs của Python hoặc dùng công cụ Onnx để chuyển đổi file model.h5 nhận được sau quá trình huấn luyện sang dạng model.json cùng các file weight tương ứng.

```

    "name": "conv2d_2",
    "class_name": "Conv2D",
    "config": {
        "filters": 32,
        "kernel_initializer": {
            "class_name": "VarianceScaling",
            "config": {
                "scale": 1,
                "mode": "fan_avg",
                "distribution": "uniform",
                "seed": null
            }
        },
        "kernel_regularizer": null,
        "kernel_constraint": null,
        "kernel_size": [
            2,
            4
        ],
        "strides": [
            1,
            1
        ],
        "padding": "valid",
        "data_format": "channels_last",
        "dilation_rate": [
            1,
            1
        ],
        "activation": "relu",
        "use_bias": true,
        "bias_initializer": {
            "class_name": "Zeros",
            "config": {

```

Hình 4.15: File model.json

Hình 4.15 thể hiện cấu trúc file model.json. Sau đó ta cần truy xuất vào audio của trình duyệt thông qua WebAudioAPI, khi có dòng dữ liệu âm thanh thời gian thực, tôi có thể dùng các thư viện SpeechComand.js , M15.js và Tensorflow.js [10] để xử lí dòng dữ liệu trên. Hai thư viện trên sẽ giúp tôi fft cho dữ liệu audio có trong Buffer, cũng như thuật toán MFCCs để trích xuất các đặc trưng của âm thanh và chuyển nó thành tensor có kích thước đầu vào tương ứng kích thước đầu vào của mô hình đã được huấn luyện ở phần trên kết hợp với cửa sổ Hanning để tránh dữ liệu bị gãy. Và lấy tensor đó cho vào mô hình để dự đoán, kết quả sẽ được ghi ra màn hình cùng hình ảnh melspectrogram tương ứng. Chi tiết fft, Hanning window sẽ được thể hiện chi tiết ở [11]

4.3 Thư viện và công cụ sử dụng

Sau đây là bảng liệt kê tất cả các ngôn ngữ lập trình, công cụ IDE tôi dùng trong quá trình thực hiện đồ án lần này:

Mục đích	Công cụ	Địa chỉ URL
IDE lập trình	Visual Studio Code	https://code.visualstudio.com/download
IDE lập trình	IntelliJ IDEA	https://www.jetbrains.com/idea/download
Hệ quản trị cơ sở dữ liệu	MySQL máy chủ	https://dev.mysql.com/downloads/mysql/
Công cụ làm việc với hệ quản trị cơ sở dữ liệu	MySQL Workbench	https://dev.mysql.com/downloads
Ngôn ngữ lập trình	Java	https://www.w3schools.com/java/
Xây dựng giao diện hệ thống	HTML	https://www.w3schools.com/html/
Xây dựng giao diện hệ thống	CSS	https://www.w3schools.com/css/
Ngôn ngữ lập trình Python	Python	https://www.w3schools.com/python/
Framework	Spring Boot	https://spring.io/
Công cụ thử API	Postman	https://www.postman.com/downloads/
Công cụ vẽ biểu đồ	AstarUML	https://astah.net/products/astah-uml/
Công cụ quản lý container	Docker	https://hub.docker.com/
Công cụ quản lý code	Git	https://github.com/

4.4 Minh họa các kết quả đạt được

4.4.1 Bài toán điểm danh

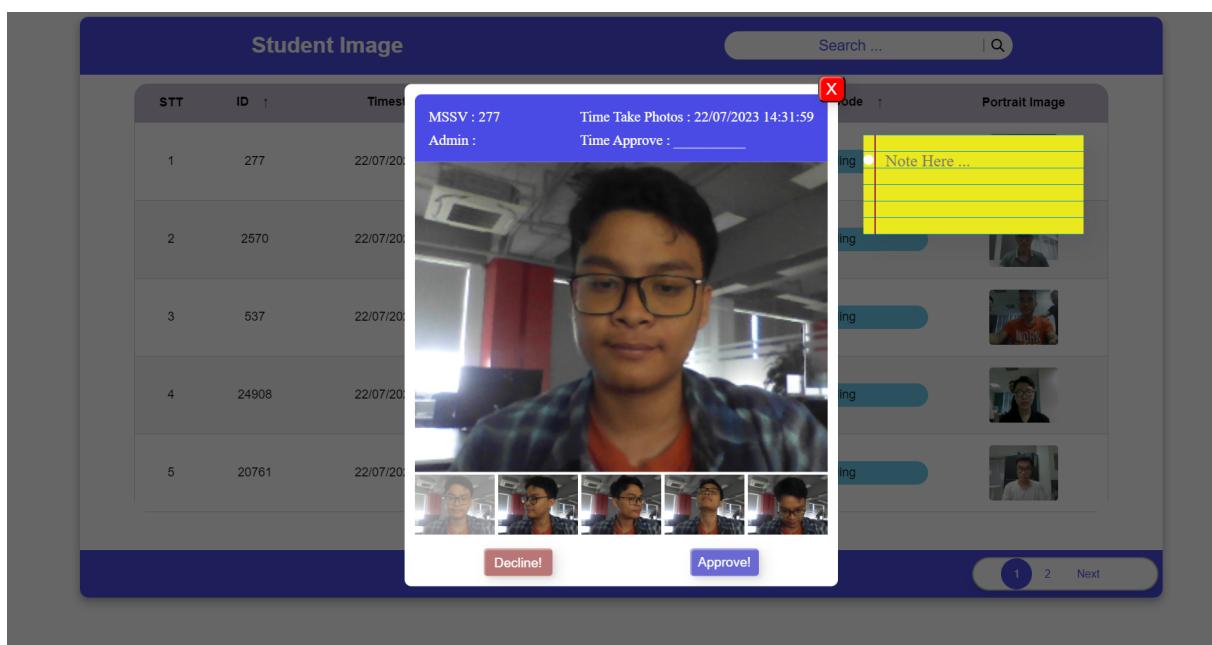
a, Giao diện chụp ảnh của sinh viên



Hình 4.16: Giao diện chụp ảnh của sinh viên

Hình 3.16 mô tả chức năng đăng ký khuôn mặt của sinh viên. Sinh viên sẽ phải chụp 5 hướng của khuôn mặt lần lượt theo hướng: Thẳng, trái, phải, trên, dưới. Hoặc có thể xóa ảnh hướng mặt nào mà bản thân thấy không chụp đúng theo hướng dẫn và chụp lại.

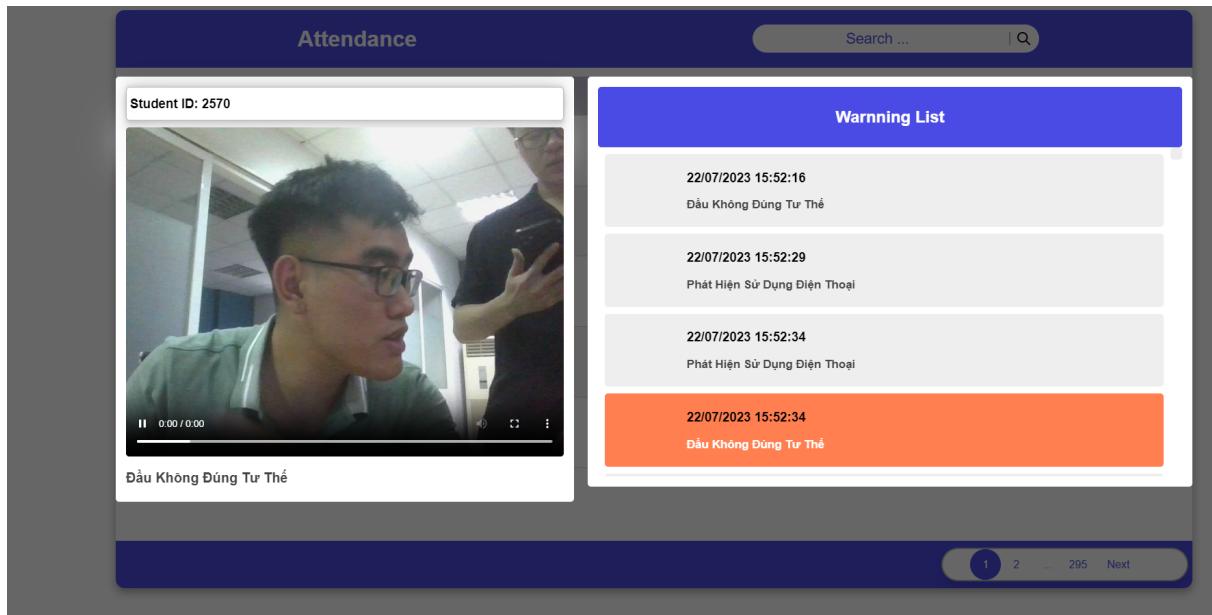
b, Giao diện xem ảnh của sinh viên



Hình 4.17: Giao diện xem ảnh của sinh viên

Hình 3.17 mô tả chức năng xem và xác nhận xem ảnh của sinh viên đăng ký có hợp lệ hay không. Nếu có nhấn Approve để xác nhận và lưu vào trong cơ sở dữ liệu, sai thì ghi chú lại vì sao ảnh không hợp lệ và lưu lại trong hệ thống, hệ thống lúc này sẽ gửi lại thông báo cho sinh viên người chụp ảnh chưa được để yêu cầu chụp lại ảnh.

c, Giao diện xem thông tin tham gia thi và các lỗi vi phạm của sinh viên

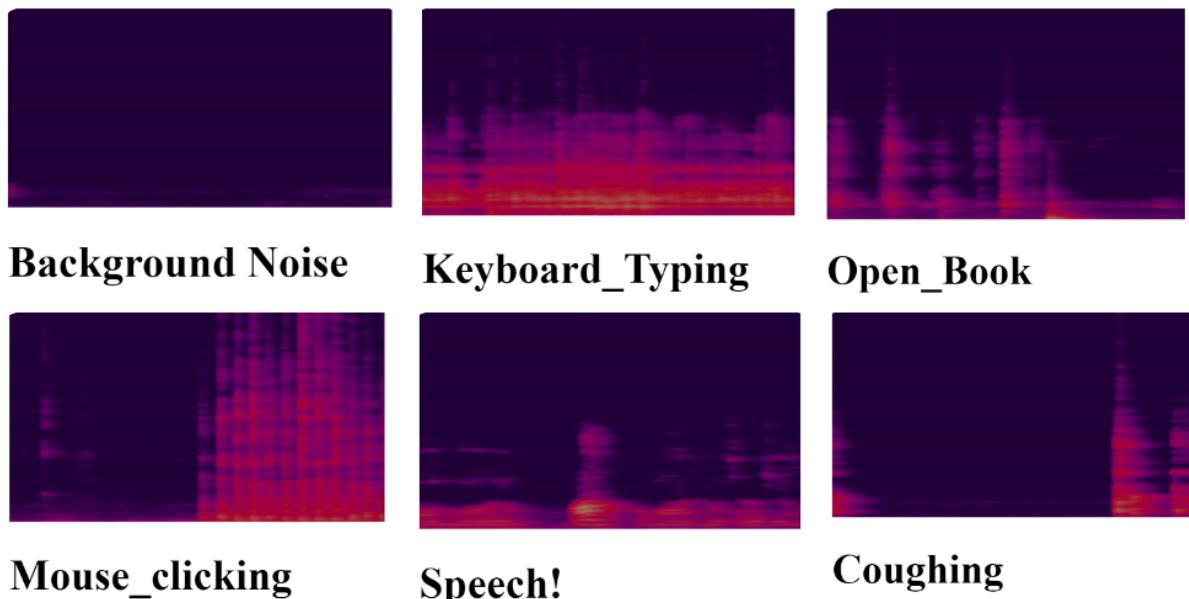


Hình 4.18: Giao diện xem thông tin tham gia thi và các lỗi vi phạm của sinh viên

Hình 3.18 mô tả chức năng điểm danh khi sinh viên có bài thi của Giảng viên. Khi giảng viên ấn nút **Bắt đầu**, yêu cầu điểm danh sẽ được gửi tới hệ thống cứ 5 phút/1 lần cho đến khi giảng viên ấn dừng lại để giảng viên có thể cập nhật liên tục tình trạng có mặt của sinh viên trước lúc chuẩn bị thi. Sau đó, trong quá trình thi, nếu sinh viên có bất kỳ dấu hiệu nào vi phạm sẽ được hệ thống ghi lại và hiển thị khách quan thời gian, lý do và video bằng chứng để giảng viên có thể xem trực quan.

4.4.2 Kết quả đạt được bài toán phát hiện tiếng nói

Hình 4.19 thể hiện kết quả phân loại âm thanh trên trình duyệt như sau:



Hình 4.19: Nhãn và hình ảnh melspectrogram tương ứng

4.5 Kiểm thử

Dưới đây là chi tiết quá trình kiểm thử dựa theo kỹ thuật kiểm tra hộp đen được thực hiện trên thiết bị Laptop Dell Precision 5520 với thông số màn hình 15.6" độ phân giải 4K(3840*2140), RAM 16GB,CPU i7 7820HQ, GPU Quadro 1200M

Sau đây là chi tiết các bảng kiểm thử.

Bảng kiểm thử đối với chức năng chụp ảnh của sinh viên:

Chức năng	Giá trị đầu vào	Giá trị đầu ra	Kết quả
Chụp ảnh	Chụp đủ 5 ảnh	Hiển thị thông báo thành công	Đạt
	Chụp thiếu ảnh	Không hiển thị nút gửi	Đạt
	Chụp sai tư thế	Thông báo lỗi cho người dùng	Đạt
	Gửi ảnh lỗi	Thông báo cho người dùng	Đạt

Bảng kiểm thử đối với chức năng duyệt thông tin, ảnh của giảng viên:

Chức năng	Giá trị đầu vào	Giá trị đầu ra	Kết quả
Kiểm duyệt ảnh	Xem 5 ảnh của từng sinh viên	Hiển thị đầy đủ thông tin, ảnh của sinh viên	Đạt
	Tìm kiếm, sắp xếp	Trả về kết quả khớp với giá trị tìm kiếm, sắp xếp	Đạt
	Từ chối ảnh không hợp lệ và viết ghi chú	Thông báo thành công, đồng thời lưu lại Code cũng như ghi chú Thông báo lại cho người dùng	Đạt
	Đồng ý với 5 ảnh	Thông báo thành công và lưu lại Code vào cơ sở dữ liệu	Đạt

CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Bảng kiểm thử đối với chức năng điểm danh, quan sát lỗi vi phạm của sinh viên:

Chức năng	Giá trị đầu vào	Giá trị đầu ra	Kết quả
Điểm danh và quan sát lỗi vi phạm	Ấn nút bắt đầu điểm danh	Hệ thống bắt đầu gọi API và trả về kết quả 5s/lần và hiển thị thông tin bảng	Đạt
	Ấn nút tạm dừng điểm danh	Hệ thống ngừng quá trình gọi API và giao diện hiển thị kết quả lần gọi gần nhất	Đạt
	Ấn vào sinh viên không có lỗi vi phạm	Thông báo sinh viên hiện không có lỗi vi phạm	Đạt
	Ấn vào sinh viên có lỗi vi phạm	Trả về thông tin gồm video, thời gian, tên lỗi của sinh đó từ lúc bắt đầu thi tới lúc được quan sát	Đạt
	Tìm kiếm, sắp xếp	Trả về thông tin cần tìm kiếm, sắp xếp	Đạt

Bảng kiểm thử đối với chức năng phân loại âm thanh trên trình duyệt của sinh viên:

Chức năng	Giá trị đầu vào	Giá trị đầu ra	Kết quả
Phân loại nhãn âm thanh	Tiếng gõ bàn phím	Hệ thống hiển thị nhãn "Keyboard-typing" trên màn hình	Đạt
	Tiếng click chuột	Hệ thống hiển thị nhãn " Mouse-clicking" trên màn hình	Đạt
	Tiếng nói	Hệ thống hiển thị nhãn " Speech" trên màn hình	Đạt
	Tiếng ho	Hệ thống hiển thị nhãn " Coughing" trên màn hình	Đạt
	Tiếng mở sách	Hệ thống hiển thị nhãn "Open-book" trên màn hình	Đạt
	Im lặng	Hệ thống hiển thị nhãn " Background Noise" trên màn hình	Đạt

Bảng kiểm thử với các yêu cầu phi chức năng:

Chức năng	Giá trị đầu vào	Giá trị đầu ra	Kết quả
Các yêu cầu phi chức năng	Responsive cho các giao diện	Hệ thống hiển thi các phần tử cần thiết	Đạt
	Màu sắc phù hợp với hệ thống	Lấy màu của daotao.ai làm đặc trưng	Đạt
	Hệ thống nhận 1 lượng lớn request	Hệ thống xử lí nhanh chuẩn xác	Đạt
	Chỉ giáo viên vào được giao diện xác nhận	Phân quyền cho các giao diện	Đạt

4.6 Triển khai

4.6.1 Cách triển khai các tính năng

Hệ thống được triển khai với các tính năng khác nhau được triển khai trên các nền tảng khác nhau

Thứ nhất, phần triển khai xây dựng mô hình CNN để huấn luyện dữ liệu âm thanh trong bài toán phân loại âm thanh được triển khai trên môi trường python 3.7.5, pip : 23.2.1 cùng các gói thư viện cần thiết khác được lưu trong file setup.txt. Để huấn luyện lại được mô hình phân loại âm thanh, cần phải dùng lệnh "pip install -r setup.txt" để cài đặt toàn bộ gói thư viện cần thiết.

Thứ hai, phần triển khai các Layout màn hình thực hiện các tính năng tương tác với người dùng được viết bằng JS,HTML,CSS không yêu cầu bất cứ môi trường đặc biệt nào, sẽ được tích hợp vào trong 1 module Open EDX-xbock để triển khai mô hình layout đó trên hệ thống Lms của daotao.ai.

Thứ 3, phần triển khai máy chủ để cung cấp dịch vụ cho người dùng sử dụng. Ở đây tôi dùng thư viện Spring Boot với JDK 8 cùng các gói thư viện với phiên bản tương ứng được ghi lại rõ trong file pom.xml được quản lý bởi Maven. Đi đôi với phần máy chủ là hệ quản trị cơ sở dữ liệu Mysql Server : 8.1.0 để lưu trữ các thông tin. Cả 2 phần trên đều được triển khai trên một docker được cấu hình trong cùng một network để trao đổi dữ liệu với nhau.

4.6.2 Triển khai trên thực tế

Trong quá trình phát triển hệ thống, nhóm đã triển khai thử nghiệm hệ thống trên văn phòng EdTech(tầng 9-tòa nhà B1) với sự hợp tác của các em học sinh đến từ Trường THCS Tạ Quang Bửu. Sau đây là một vài hình ảnh và thống kê sau buổi triển khai hệ thống thử nghiệm thực tế : Hình 4.20 thể hiện : học sinh đang trải nghiệm hệ thống :

CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ



Hình 4.20: Trải nghiệm hệ thống giám sát thi

Với kết quả khảo sát như Hình 4.21:

Timestamp	Bạn đánh giá chương trình phát hiện giả mạo như thế nào? [Tốc độ]	Bạn đánh giá chương trình phát hiện giả mạo như thế nào? [Độ chính xác]
2023/06/24 11:15:20 AM GMT+7	Tốt	Tốt
2023/06/24 11:15:43 AM GMT+7	Tốt	Tốt
2023/06/24 11:15:57 AM GMT+7	Tốt	Tốt
2023/06/24 11:20:45 AM GMT+7	Tốt	Bình thường
2023/06/24 11:22:07 AM GMT+7	Rất tốt	Rất tốt
2023/06/24 11:22:23 AM GMT+7	Rất tốt	Rất tốt
2023/06/24 11:24:55 AM GMT+7	Bình thường	Tốt
2023/06/24 11:27:42 AM GMT+7	Tốt	Rất tốt
2023/06/24 11:28:05 AM GMT+7	Tốt	Tốt
2023/06/24 11:34:04 AM GMT+7	Tốt	Tốt
2023/06/24 11:37:26 AM GMT+7	Bình thường	Tốt
2023/06/24 11:42:00 AM GMT+7	Tốt	Rất tốt
2023/06/24 11:46:16 AM GMT+7	Rất tốt	Rất tốt
2023/06/24 11:49:38 AM GMT+7	Tốt	Tốt
2023/06/24 11:50:13 AM GMT+7	Rất tốt	Tốt
2023/06/24 11:58:07 AM GMT+7	Rất tốt	Rất tốt
2023/06/24 11:59:50 AM GMT+7	Tốt	Bình thường
2023/06/24 12:00:41 PM GMT+7	Rất tốt	Rất tốt
2023/06/24 12:01:14 PM GMT+7	Rất tốt	Rất tốt

Hình 4.21: Kết quả thử nghiệm hệ thống

Vậy chương 4, tôi đã trình bày tổng quan chi tiết, từ thiết kế đến phần triển khai cũng như kết quả đạt được của các bài toán. Ở chương 5, tôi sẽ trình bày về các đóng góp nổi bật, các hạn chế của các bài toán tôi đã triển khai và các hướng phát triển bài toán trong tương lai.

CHƯƠNG 5. CÁC GIẢI PHÁP VÀ ĐÓNG GÓP NỔI BẬT

5.1 Các công hiến

Nhìn chung hệ thống daotao.ai có nhiều tính năng phát hiện, xử lí hình ảnh trong vấn đề chống gian lận thi cử nhưng chưa có bất kì tính năng nào về xử lí âm thanh mà dùng ít tài nguyên hệ thống (CPU, GPU) cũng như hệ thống chưa có bất kì giao diện nào triển khai để người dùng sử dụng một cách khách quan sinh động nên đồ án này tôi đã phát triển 2 bài toán đó như sau:

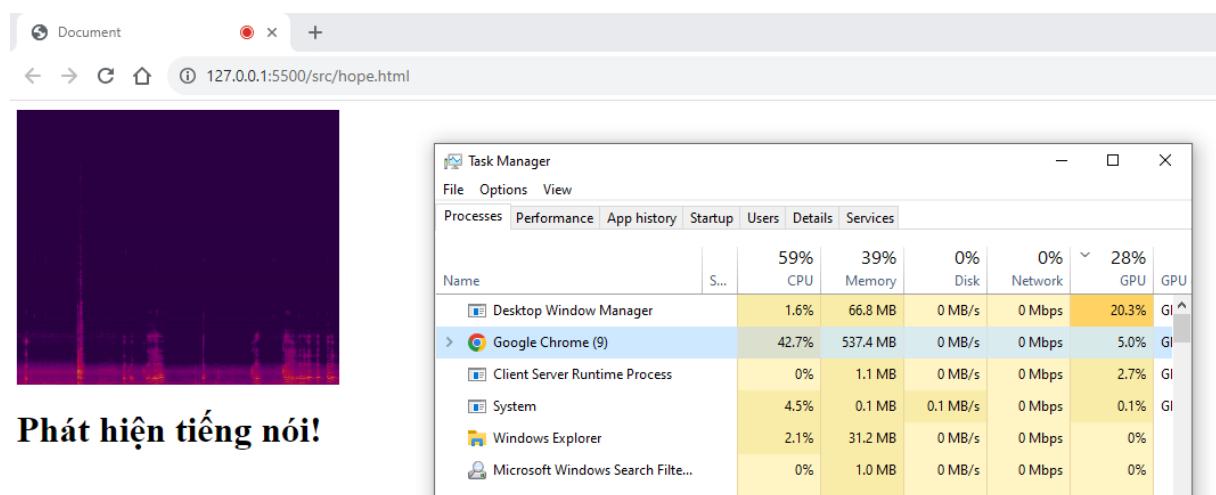
5.1.1 Xây dựng tính năng phát hiện tiếng nói cụ thể áp dụng vào trong trường học

a, Đặt vấn đề

Trong thời điểm hiện tại, có rất nhiều mô hình cũng như bài toán phát hiện tiếng nói nhưng hầu hết các bài toán đó được xử lí bằng ngôn ngữ Python . Tuy nhiên trong quá trình xây dựng hệ thống chống gian lận trong thi cử, tôi đã gặp phải tình trạng hết tài nguyên về CPU, GPU ở phía máy chủ nên yêu cầu đặt ra là phải xây dựng được 1 tính năng phát hiện tiếng nói của sinh viên trong lúc thi mà không tốn nhiều tài nguyên phía máy chủ.

b, Giải pháp đưa ra và kết quả đạt được

Để giải quyết khó khăn trên thì cần một luồng chạy đơn giản: tôi sẽ xây dựng mô hình CNN nhẹ, chuyển nó thành dạng json sao cho JS có thể sử dụng trực tiếp mô hình đó trên client đó mà không cần gửi đến máy chủ. Hình 5.1 thể hiện hiệu năng CPU, GPU tại máy khách mà không liên quan đến máy chủ.



Hình 5.1: Tài nguyên dùng để phát hiện tiếng nói

Sau đó tôi sẽ dùng thuật toán fft để mã hóa âm thanh đầu vào và dự đoán nhãn tương

ứng. Kết quả là tính năng đã chạy ổn với độ chính xác khoảng 70%.

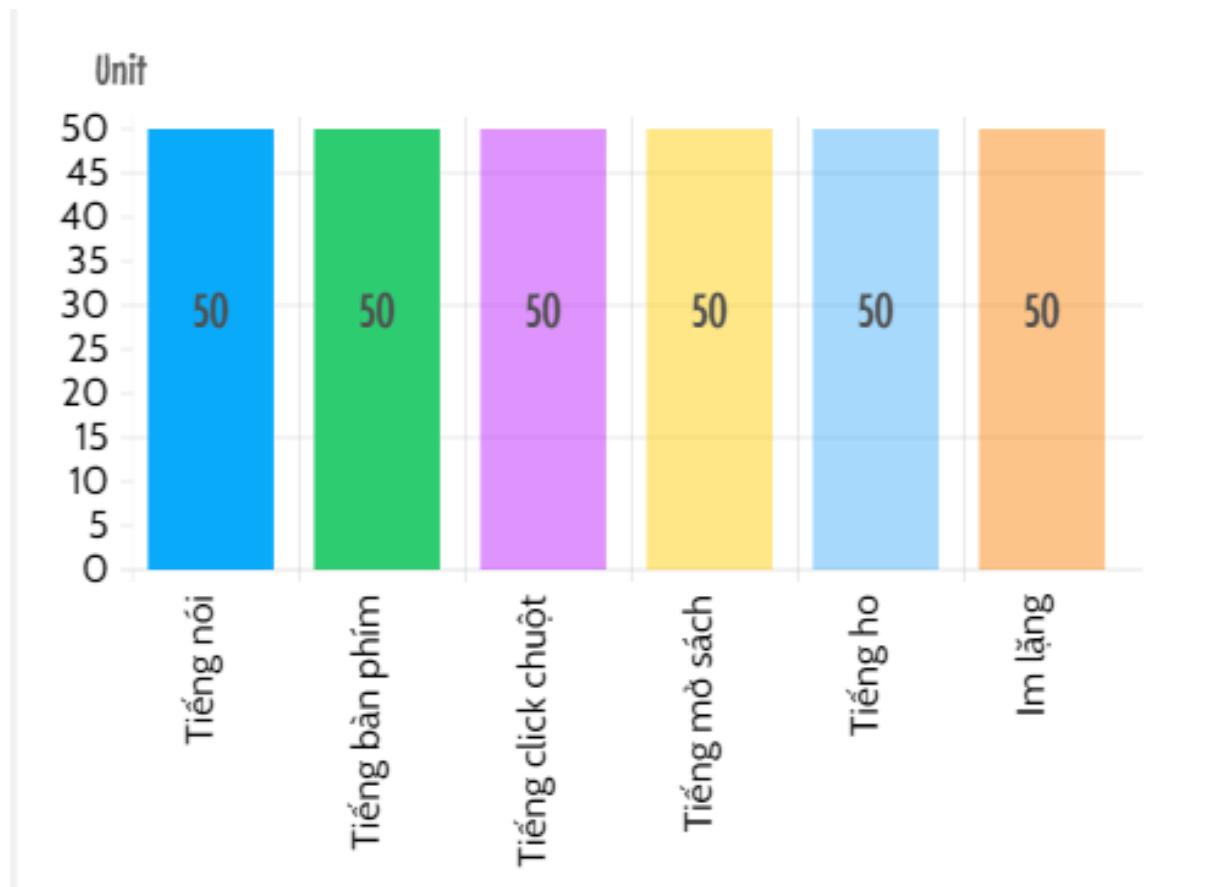
5.1.2 Bộ dữ liệu đánh nhãn các file âm thanh

a, Đặt vấn đề

Sau khi hoàn thiện xong mô hình phân loại âm thanh, để đánh giá được các độ chính xác của mô hình, cần phải thực hiện đánh giá trên bộ dữ liệu phù hợp.Tuy nhiên sau khi tìm kiếm thông tin với từ khóa và trên nhiều trang web khác nhau, tôi vẫn không thể tìm kiếm được bộ dữ liệu nào phù hợp hoàn toàn với các nhãn được đề ra trong bài toán.

b, Giải pháp đưa ra và kết quả đạt được

Để giải quyết vấn đề trên,cụ thể là bộ dữ liệu các file âm thanh dạng .wav để train và kiểm thử mô hình phân loại âm thanh, với số lượng khoảng 300 file .wav chia làm 6 nhãn (tiếng nói, tiếng ho, im lặng, tiếng gõ bàn phím, tiếng click chuột, tiếng mở tài liệu) đã được tự ghi âm lại và đánh nhãn tương ứng và chia làm 2 thư mục train và test. Hình 5.2 thể hiện số tệp mẫu mỗi nhãn:



Hình 5.2: Dữ liệu âm thanh

Bộ dữ liệu này sẽ được đóng góp cho nhà trường và cộng đồng để phục vụ cho mục đích học tập và nghiên cứu những bài toán liên quan. Người dùng có thể tham khảo ở link : <http://surl.li/jrvgi>

5.1.3 Xây dựng tính năng điểm danh cũng như quan sát các lỗi vi phạm

a, Đặt vấn đề

Sau khi triển khai xong các bài toán xử lí AI trên daotao.ai như phát hiện âm thanh, khuôn mặt, hướng mặt, điện thoại... yêu cầu cấp thiết đặt ra là phải xây dựng các giao diện để lấy dữ liệu khuôn mặt sinh viên, lấy dữ liệu âm thanh vào để hệ thống có thể dự đoán cũng như lưu trữ dữ liệu vào cơ sở dữ liệu. Mục đích để giáo viên có thể quan sát, kiểm soát các thông tin và hình ảnh một cách trực quan, sinh động.

b, Giải pháp đưa ra và kết quả đạt được

Để giải quyết các yêu cầu trên, tôi đã xây dựng các giao diện chụp ảnh, thống kê cho giáo viên, sinh viên cũng như xây dựng phía máy chủ, cơ sở dữ liệu để xử lí và lưu trữ dữ liệu. Đồng thời triển khai nó trên Docker cho tiện cho việc triển khai và cài đặt. Quan trọng nhất là triển khai nó trên hệ thống daotao.ai bằng cách xây dựng từng module giao diện , máy chủ và gộp nó vào hệ thống của trường.

5.2 Các hạn chế

Hiện tại do còn nhiều hạn chế nên tính năng phát hiện giọng nói vẫn chưa được chuẩn xác cũng như kết quả chỉ là lời cảnh báo cho sinh viên ở màn hình máy khách.. Vì thế trong tương lai định hướng của tôi sẽ đồng bộ thời gian của audio và video và lưu log vào trong hệ thống.

CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong chương 5, tôi đã trình bày tóm gọn các thành vấn đề và và thành tựu đã đạt được trong suốt quá trình làm đồ án từ đó đưa ra kết luận, hạn chế và các giải pháp trong tương lai

6.1 Kết luận

Thông qua việc nghiên cứu bài toán phân loại âm thanh đã giúp tôi nắm rõ hơn các bước xây dựng và triển khai một vài kỹ thuật tiền xử lí thông tin dạng âm thanh, các mô hình học sâu cho bài toán phân loại ảnh, cũng như cách triển khai một mô hình học máy trên trình duyệt mà không cần gửi dữ liệu về máy chủ dự đoán và trả về kết quả. Bài toán này cũng giúp tôi hiểu hơn về các khái niệm: âm thanh, FFT, Hanning window, Melspectrogram, Spectrogram...

Bên cạnh những khó khăn ở bài toán phát hiện tiếng nói, thì điều tôi học được ở bài toán xây dựng luồng điểm danh là xử lý lỗi CORS giữa các máy chủ với nhau cũng như cũng như cách triển khai một hệ thống lớn như daotao.ai.

Tuy nhiên qua quá trình xây dựng đồ án, tôi đã tích lũy thêm nhiều kiến thức và kinh nghiệm về việc phân tích vấn đề, kỹ năng lập trình và cách làm việc hiệu quả với tất cả các thành viên trong đội ngũ phát triển .Qua đó giúp tôi có thêm hành trang khi bước qua ngưỡng cửa Đại học.

6.2 Hướng phát triển

Vì thời gian phát triển hạn chế nên các tính năng vẫn chưa thực sự hoàn hảo. Các giao diện vẫn tồn tại nhưng lỗi nhỏ cũng như trong quá trình tiền xử lí âm thanh chưa hoàn toàn tối ưu, bộ dữ liệu quá nhỏ. Từ những đánh giá của bản thân cũng như được sự góp ý của thầy cô bạn bè, tôi xin đưa ra một số điểm của các tính năng cần được hoàn thiện như sau:

- Cân thu thập thêm các dữ liệu âm thanh dạng Mono
- Triển khai thêm về hướng âm thanh dạng Stereo
- Thêm các bước tiền xử lí cho các tệp âm thanh như : Tăng cường dữ liệu(Mặt nạ thời gian và tần suất), Lọc nhiễu trước khi chuyển âm thành mel spectrogram
- Xử lí hoàn thiện các vấn đề nghiệp vụ ở máy khách cũng như máy chủ.

TÀI LIỆU THAM KHẢO

- [1] C. T. T. T. C. PHỦ, *Lợi ích từ ứng dụng công nghệ thông tin trong giáo dục*. [Online]. Available: <https://xaydungchinhhsach.chinhphu.vn/loi-ich-tu-ung-dung-cong-nghe-thong-tin-trong-giao-duc>. (visited on 08/01/2023).
- [2] B. T. giáo Trung ương, *Dạy học trực tuyến để ứng phó với dịch covid-19*. [Online]. Available: <https://tuyengiao.vn/khoa-giao/giao-duc/day-hoc-truc-tuyen-de-ung-pho-voi-dich-covid-19-135538> (visited on 08/01/2023).
- [3] MOOC.org, *About moocs*.
- [4] K. Doshi, *Audio deep learning made simple (part 2): Why mel spectrograms perform better*. [Online]. Available: <https://towardsdatascience.com/audio-deep-learning-made-simple-part-2-why-mel-spectrograms-perform-better-aad889a93505> (visited on 08/01/2023).
- [5] VERITONE, *Audiofingerprinting*. [Online]. Available: <https://www.veritone.com/aiware/ai-engines/audio-fingerprinting/> (visited on 08/01/2023).
- [6] K. Doshi, *Audio deep learning made simple: Sound classification*. [Online]. Available: <https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5> (visited on 08/01/2023).
- [7] K. J. Piczak, *ESC: Dataset for Environmental Sound Classification*. [Online]. Available: <https://github.com/karolpiczak/ESC-50> (visited on 08/01/2023).
- [8] N. C. Thắng, *Loss và accuracy*. [Online]. Available: <https://github.com/tensorflow/tfjs-models/tree/master/speech-commands> (visited on 08/01/2023).
- [9] Google, *Speech command recognizer*. [Online]. Available: <https://github.com/tensorflow/tfjs-models/tree/master/speech-commands> (visited on 08/01/2023).
- [10] Google, *Tensorflow*. [Online]. Available: <https://www.tensorflow.org/> (visited on 08/01/2023).
- [11] Wikipedia, *Biến đổi fourier nhanh*. [Online]. Available: https://vi.wikipedia.org/wiki/Bi%e1%ba%bf%e1%91%e1%bb%95i_Fourier_nhanh (visited on 08/01/2023).