CUỘC THI KHOA HỌC KỸ THUẬT
DÀNH CHO HỌC SINH TRUNG HỌC CẤP QUỐC GIA
NĂM HỌC 2021 - 2022

BẢN BÁO CÁO TÓM TẮT KẾT QUẢ NGHIÊN CỨU

Tên đề tài:

ỨNG DỤNG Look&Tell: GIẢI PHÁP HỖ TRỢ GIAO TIẾP CHO NGƯỜI ĐIẾC BẰNG TRÍ TUỆ NHÂN TẠO

Lĩnh vực: **Phần mềm hệ thống**

Ngày 10 tháng 2 năm 2022

MỤC LỤC

DANH MỤC HÌNH VẼ, SƠ ĐỒ

DANH MỤC BẢNG BIỂU

Tóm tắt **- Mặc dù trên thế giới cũng như ở Việt Nam, các sản phẩm hỗ trợ giao tiếp giữa người Điếc và người Nghe đã có những bước tiến nhất định, tuy thế vẫn còn một số hạn chế như không hỗ trợ tiếng Việt hay yêu cầu các phụ kiện đeo trên người, gây khó khăn trở ngại cho quá trình giao tiếp. Dựa trên nhu cầu về giải pháp giao tiếp tiện dụng và hiệu quả, nhóm nghiên cứu đã phát triển ứng dụng Look&Tell dựa trên công nghệ Trí tuệ nhân tạo và nền tảng điện thoại thông minh. Sản phẩm hướng tới ba tính năng chính (1) dịch Ngữ (dịch ngôn ngữ ký hiệu (NNKH) thành chữ viết), (2) chuyển Ngữ (chuyển âm thanh thành chữ viết) và (3) Từ Điển (từ điển NNKH). Với Dịch Ngữ, bằng việc tạo dựng tập dữ liệu người Việt với 6900 dữ liệu là các video, nhóm đã tiến hành điều chỉnh cấu trúc mạng, các hàm tính toán, đưa ra mô hình LSTM nhận diện đáp ứng được kỳ vọng với tốc độ nhận diện thời gian thực và độ chính xác lên tới 93.33%. Hiện tại, sản phẩm chạy ổn định với cả người Điếc, người Nghe với hai tính năng đầu (Từ Điển đang được hoàn thiện). Đây hứa hẹn là tiềm năng phát triển để hỗ trợ giao tiếp cho người Điếc một cách hiệu quả, tiện dụng với chi phí chấp nhận được.**

DANH SÁCH CHỈ MỤC

PHẦN I: TỔNG QUAN DỰ ÁN

I.1. Lý do chọn dự án

Hiện nay, trên thế giới có khoảng 1,5 tỷ người khiếm thính (KT) [1]. Đến năm 2050 sẽ tăng lên 2,5 tỷ người – tương ứng 4 người sẽ có 1 người KT [2]. Trong đó, theo báo cáo của Liên đoàn Người Điếc Thế giới (WFD), hiện tại có hơn 70 triệu người Điếc toàn thế giới [3]. Riêng với trẻ em, cứ mỗi 1000 em sinh ra lại có 5 em được phát hiện mất khả năng nghe vĩnh viễn [1].

Theo báo cáo của Tổ chức Y tế Thế Giới (WHO) năm 2015, các khoản chi dành cho KT rơi vào 750 - 790 tỷ USD/năm. Với hơn 573 tỷ cho các chi phí xã hội (bao gồm cô lập xã hội, khó khăn trong giao tiếp và định kiến với người KT) [4]. Tổ chức Global Burden of Disease cũng đưa ra báo cáo rằng ngân sách hỗ trợ KT năm 2019 đã vượt mức 981 tỷ USD. Trong đó, 47% liên quan đến suy giảm chất lượng cuộc sống do rào cản xã hội và việc làm, đặc biệt với người Điếc [5].

Tại Việt Nam, trong báo cáo Điều tra quốc gia người khuyết tật lớn nhất năm 2016, tỉ lệ người KT và điếc từ 18 tuổi trở lên ở cả nước là 1,37%; nhóm 5-17 tuổi là 0,25% và 0,13% rơi vào nhóm 2 - 4 tuổi [6].

Với người Điếc, ngôn ngữ ký hiệu (NNKH) là ngôn ngữ giao tiếp chủ yếu trong sinh hoạt và lao động. Trong khi đó, số người phiên dịch NNKH có trình độ tại Việt Nam chỉ chưa tới 20 người [7]. Nhưng các sản phẩm hỗ trợ dù trong quá trình nghiên cứu hay trên thị trường đều gặp phải các vấn đề như có xâm nhập (găng tay phiên dịch) [8], yêu cầu có linh phụ kiện đi kèm khiến giá thành cao [9], và đặc biệt hầu hết không hỗ trợ ngôn ngữ tiếng Việt, …

Nhóm nghiên cứu cũng đã thực hiện khảo sát với 300 người Điếc và người Nghe[1], thu về con số 97,33% mong muốn có sản phẩm hỗ trợ giao tiếp giữa hai cộng đồng, cho thấy nhu cầu rất lớn trong việc tạo cầu nối giao tiếp.

Trong thời đại bùng nổ công nghệ thông tin, viễn thông hiện nay, điện thoại thông minh (Smartphone – ĐTTM) đang dần trở nên phổ biến, là một công cụ rất hữu ích bởi tính tiện lợi, chi phí chấp nhận được, và đặc biệt cũng được ứng dụng trong phân tích, giao tiếp ngôn ngữ như các sản phẩm dịch máy, dịch từ ảnh chụp, … Đó là lý do tại sao chúng ta có thể hướng tới việc xây dựng ứng dụng phần mềm trên ĐTTM để hỗ trợ người Điếc giao tiếp hiệu quả.

I.2. Những nghiên cứu liên quan:

Nghiên cứu tổng quan ở các khía cạnh: đời sống cá nhân và xã hội, kinh tế, ngôn ngữ giao tiếp, sức khỏe thể chất và tinh thần, trình độ học vấn, … liên quan đến người Điếc cho thấy hai phương pháp phổ biến nhất dịch NNKH là dựa vào sensor (bộ cảm biến) và dựa vào nhận dạng hình ảnh.

[1] Người không thuộc cộng đồng người Điếc, người khiếm thính – cách gọi được thống nhất trong các tài liệu giảng dạy đặc biệt và đang được phổ biến trong xã hội

Tiếp cận dựa vào sensor: người dùng sử dụng một cặp găng tay có gắn sensor khi biểu thị NNKH. Sensor sẽ làm giảm các yếu tố nhiễu do môi trường, giúp tiền xử lý hình ảnh đơn giản hơn. Tuy thế, phương pháp tương đối bất tiện do người dùng phải đeo thường xuyên, chi phí đắt đỏ. Các sản phẩm tham khảo: Găng tay phiên dịch ASL (Hoa Kỳ, UCLA [10]); găng tay phiên dịch CSL (Trung Quốc, Wulala Technology [11])

Tiếp cận dựa vào hình ảnh: các cử chỉ của tay, cơ thể được ghi lại dưới dạng hình ảnh cố định hoặc chuỗi hình ảnh bằng camera. Tiện lợi, nhỏ gọn và giá thành là điểm ưu việt, nhưng việc xử lý trong các môi trường phức tạp (nhiều nhiễu: màu da, cấu trúc cơ thể theo sắc tộc, cấu hình camera, …) làm giảm hiệu quả và tốc độ xử lý. Các sản phẩm tham khảo: Microsoft Kinect [12]; VOM [13].

Hướng tiếp cận bằng xử lý hình ảnh hầu hết dựa trên các mô hình Học sâu với hai nhánh chính:

Mạng nơron tích chập (Convolutional Neural Network - CNN): Luyện một tập dữ liệu lớn các hình ảnh về NNKH, trích xuất các đặc trưng tạo thành bản đồ đặc trưng (feature map) sau đó dự đoán các từ. Ưu điểm là tốc độ nhanh, độ chính xác cao, nhưng độ chính xác suy giảm nếu quay ở nhiều góc khác nhau, đồng thời khó dịch các câu gồm chuỗi nhiều hành động.

Mạng nơron hồi tiếp (Recurrent Neural Network - RNN): Thu thập và xây dựng tập dữ liệu lớn các video chứa hành động NNKH. Phát triển hệ thống nhận diện và từ đó trích xuất ra chuỗi các đặc trưng dựa trên khung xương của bàn tay và/hoặc cơ thể. Chuỗi các đặc trưng này được mã hóa thành chuỗi vector làm đầu vào cho mạng nơron hồi quy. RNN rất phù hợp để xử lý các bài toán đầu vào dạng chuỗi và các yếu tố chuỗi có ảnh hưởng tới nhau. Nhược điểm là với việc xử lý dữ liệu tuần tự, tốc độ chạy sẽ tỉ lệ thuận với độ dài dữ liệu.

Cũng đã có một số sản phẩm được cung cấp cho điện thoại thông minh nhưng chưa được khai thác nhiều: MonoVoix [14], Ear Hear [15]

Tại Việt Nam, các nghiên cứu tập trung giải quyết bài toán hệ thống ngôn ngữ và hòa nhập xã hội ở cấp địa phương, khu vực, các sản phẩm hỗ trợ chưa được chú trọng. Sản phẩm có thể kể đến có SYM (ĐH Bách Khoa TP. HCM [9])

Nhìn chung, xu hướng nghiên cứu của thế giới và trong nước là khác nhau, nhưng đều xuất phát và tập trung giải quyết vấn đề cơ bản: “Làm sao để người Nghe hiểu được người Điếc diễn đạt điều gì?”.

Trong dự án của mình, nhóm quyết định phát triển phương án sử dụng mạng nơron hồi tiếp, với phiên bản nâng cấp bộ nhớ dài ngắn (Long Short - Term Memory - LSTM) kết hợp hệ thống nhận diện khung xương từ thư viện Mediapipe, OpenCV. Những điểm cơ thể được nhận diện bằng Mediapipe được xuất ra thành chuỗi véc tơ làm đầu vào cho mạng LSTM. Từ đó có thể tạo ra một trình dịch liên tục và chính xác trong thời gian thực dựa trên các dự đoán.

I.3. Câu hỏi nghiên cứu

Dựa trên nhu cầu về sản phẩm hỗ trợ giao tiếp giữa người Điếc và người Nghe và điều kiện công nghệ, nhóm nghiên cứu đi tới câu hỏi bao trùm: "Làm thế nào để xây dựng một ứng dụng phần mềm hệ thống hỗ trợ giao tiếp giữa người Điếc và người Nghe Việt Nam trên điện thoại thông minh sử dụng trí tuệ nhân tạo?"

Để trả lời được câu hỏi trên, chúng em đặt ra các câu hỏi trọng tâm sau:

1. Người Điếc chủ yếu sử dụng hình thức giao tiếp nào?
2. Để hỗ trợ nhu cầu giao tiếp tối thiểu của người Điếc, qua hình thức giao tiếp chủ yếu của họ, phần mềm ứng dụng cần có những tính năng gì?
3. Phần mềm ứng dụng cần hỗ trợ tối thiểu bao nhiêu từ giao tiếp thông dụng?
4. Để việc hỗ trợ hiệu quả, phần mềm cần hoạt động với tốc độ như thế nào?
5. Phần mềm cần và có thể đạt độ chính xác là bao nhiêu trong việc diễn đạt các từ trong giao tiếp hàng ngày?
6. Phần mềm ứng dụng cần hoạt động trên các nền tảng phần cứng, hệ điều hành và phần mềm khác như thế nào?
7. Công nghệ Trí tuệ nhân tạo sẽ được ứng dụng như thế nào? Ngoài ra còn sử dụng các công nghệ nào khác?
8. Các ngôn ngữ và nền tảng lập trình, thư viện nào là cần thiết?

I.4. Mục tiêu nghiên cứu

Trên cơ sở các câu hỏi nghiên cứu, chúng em đặt ra mục tiêu: "Xây dựng ứng dụng phần mềm Look&Tell để hỗ trợ người Điếc giao tiếp với đầy đủ các tiêu chí là hiệu quả, tiện lợi và chi phí chấp nhận được", với các mục tiêu và vấn đề nghiên cứu cụ thể sau:

1. Ứng dụng hỗ trợ nhận dạng và phiên dịch NNKH tay của người Điếc
2. Phần mềm nhận dạng được khoảng 100 câu trong giao tiếp thông thường
3. Phần mềm xử lý nhận dạng thời gian thực (24 – 30 hình ảnh/giây)
4. Phần mềm xử lý nhận dạng chính xác khoảng 90%
5. Phần mềm chạy được trên ĐTTM hệ điều hành Android – API 21 trở lên
6. Phần mềm xây dựng dưa trên công nghệ Trí tuệ nhân tạo (Artificial Intelligence – AI); Học sâu (Deep Learning – DL) – cụ thể là mạng nơ ron LSTM; Thị giác máy tính (Computer Vision – CV);
7. Ngôn ngữ lập trình Python, Java, XML; nền tảng lập trình Visual Studio Code, Android Studio; nền tảng thư viện Tensorflow, OpenCV, Mediapipe.

I.5. Tính mới của dự án

- Đây là ứng dụng phần mềm đầu tiên được xây dựng để hỗ trợ giao tiếp hàng ngày giữa người Điếc và người Nghe tại Việt Nam trên điện thoại Android mà không yêu cầu linh phụ kiện.
- Xây dựng tập dữ liệu huấn luyện về NNKH đã được chuẩn hóa và làm sạch của người Việt Nam trải từ 10 tuổi trở lên.

Trong hai phần sau, nhóm nghiên cứu sẽ lần lượt trình bày về phần *II: Quy trình và phương pháp thực hiện dự án*; phần *III: Kết quả nghiên cứu dự án*

PHẦN II: QUY TRÌNH VÀ PHƯƠNG PHÁP THỰC HIỆN DỰ ÁN

Hình 1. Quy trình thực hiện dự án

Ở phần này, nhóm sẽ tập trung trình bày các công việc được thực hiện tương ứng với các bước trong quy trình. Hai phần cuối gồm: *Chạy thử và kiểm tra sản phẩm, Bảo trì* sẽ được làm rõ trong mục *III.2,3.*

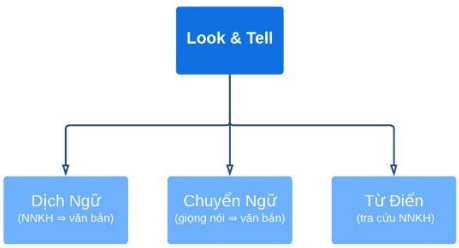
II.1. Điều tra khảo sát và phân tích sơ bộ

Thực hiện nghiên cứu, khảo sát, nhóm còn thu được yêu cầu đặc hữu chính:

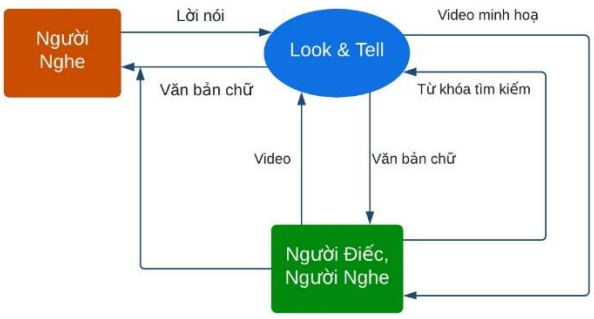
1. Cần có một sản phẩm có thể dịch NNKH thành lời nói/chữ viết
2. Cần có một sản phẩm có thể dịch lời nói/chữ viết thành NNKH
3. Sản phẩm phải có giá thành hợp lý, tiện lợi, hiệu quả
4. Sản phẩm có thể giúp người dùng học NNKH

II.2 Phân tích về chức năng

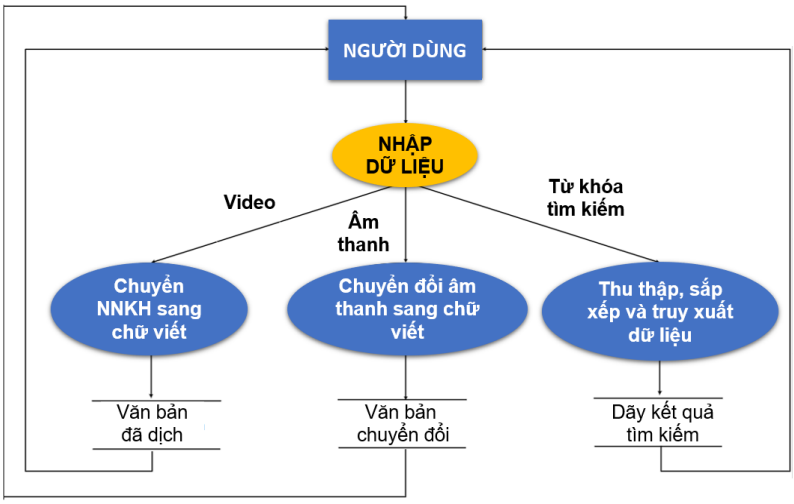
Từ phân tích về yêu cầu và công nghệ, nhóm xác định sản phẩm dự án: Phần mềm ứng dụng tích hợp 3 chức năng: (1) Dịch Ngữ - DN (dịch NNKH sang chữ viết), (2) Chuyển Ngữ - CN (chuyển âm thanh sang chữ viết) và (3) Từ Điển – TĐ (tra cứu NNKH). Phân cấp chức năng ứng dụng thể hiện trong *Hình 2.* Sơ đồ luồng dữ liệu mức ngữ cảnh, ngữ đỉnh ứng dụng lần lượt là *Hình 3,4.*



Hình 2. Sơ đồ phân cấp chức năng ứng dụng Look&Tell



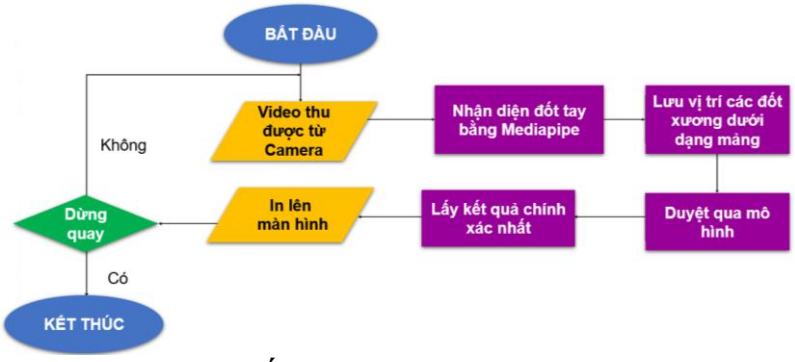
Hình 3. Sơ đồ luồng dữ liệu mức ngữ cảnh ứng dụng Look&Tell



Hình 4. Sơ đồ luồng dữ liệu mức ngữ đỉnh ứng dụng Look&Tell

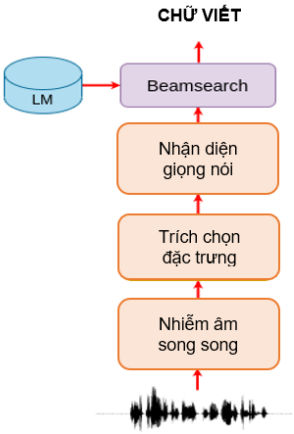
II.3 Thiết kế sản phẩm

II.3.1 Thiết kế về chức năng



Hình 5. Cơ chế hoạt động chức năng Dịch Ngữ

Ban đầu, người dùng bấm quay, ứng dụng sẽ nhận vào video. Video sẽ đi qua Mediapipe, vị trí đốt xương được nhận diện và lưu lại vào véc tơ. Mô hình DL nhận liên tiếp các nhóm 60 véc tơ làm đầu vào và trả về kết quả dịch hiển thị lên màn hình. Quá trình này lặp lại đến khi người dùng bấm dừng quay *(Hình 5)*.



Hình 6. Cơ chế vận hành chuyển âm thanh thành chữ viết – Chuyển Ngữ

Âm thanh thu vào sẽ bị nhiễm các kỹ thuật tăng cường như giản lược thời gian/tần số, thêm nhiễu, vang, … trong khi âm chuyển mà không cần lưu trữ các tín hiệu tăng cường trên đĩa. Sau đó tiến hành trích chọn đặc trưng giọng nói làm đầu vào cho mô hình DL. Mô hình sẽ tính toán và trả ra kết quả cùng với các khả

năng. Kết quả này được duyệt bằng một beamsearcher để khám phá các thay thế và trả đầu ra tốt nhất. Các lựa chọn thay thế có thể được đánh giá lại với một Mô hình ngôn ngữ (Language Model – LM).

II.3.2 Thiết kế giao diện

Toàn bộ phần thiết kế và cài đặt giao diện được nhóm xây dựng theo lược đồ ở *Hình 7*

Hình 7. Quy trình thiết kế và cài đặt giao diện ứng dụng

Phác thảo: Dựa vào các phân tích về chức năng ứng dụng (*mục II.2*), chúng em đưa ra phác thảo ý tưởng giao diện gồm 5 màn hình chính:

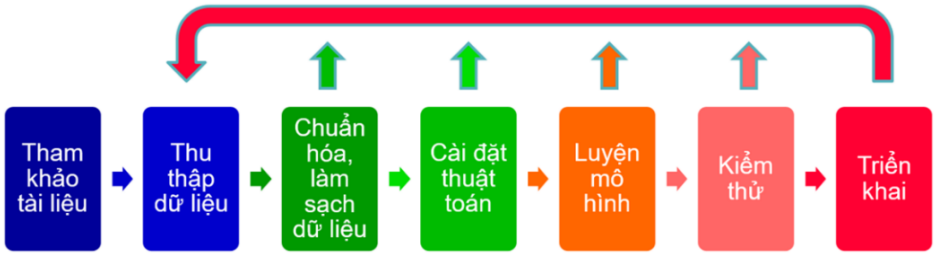
1. Màn hình mở đầu: hiển thị khi ứng dụng được khởi động
2. Màn hình chính: hiển thị thanh điều hướng và thông tin hỗ trợ
3. Màn hình Dịch Ngữ: hiển thị chức năng Dịch Ngữ
4. Màn hình Chuyển Ngữ: hiển thị chức năng Chuyển Ngữ
5. Màn hình Từ Điển: hiển thị chức năng Từ Điển

Tạo giao diện mẫu: Giao diện mẫu gồm các màn hình phác thảo liên kết, tương tác với nhau; yếu tố giao diện người dùng (UI – User Interfaces) hoàn thiện dần.

Hai bước 3 và 4 sẽ tiếp tục được trình bày trong *mục II.4.2*; hai bước cuối cùng sẽ được nêu ra trong *mục III.2.*

II.4 Cài đặt sản phẩm

II.4.1 Cài đặt, phát triển mô hình



Hình 8. Giai đoạn phát triển mô hình

Tham khảo tài liệu:

- Đặt vấn đề dựa trên mục tiêu dự án
- Tham khảo các phương pháp dựng mô hình DL về phát hiện hành động trên các báo khoa học, diễn đàn về AI, trang web mã nguồn mở
- Phân tích ưu nhược điểm các phương pháp, tiếp thu góp ý, thống nhất phương án dựng mô hình

Thu thập dữ liệu: Định dạng và đặc điểm dữ liệu:

- Định dạng: Video
- Số lượng: 90 video/ câu (từ)
- Kích thước các video: giống nhau với từng câu (từ)
- Góc quay: 3 góc (chính diện $\pm 30^{o)}$ ở khung gần: 100 cm ($\pm$ 10 cm), xa: 160 cm ($\pm$ 10 cm) – tính từ camera.

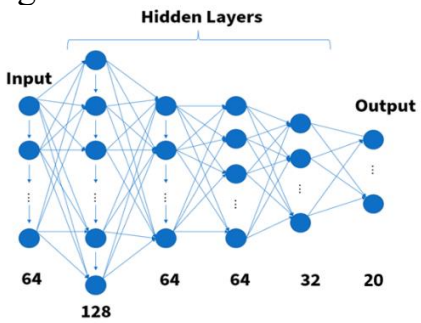
Kết thúc quá trình thu thập, nhóm thu được $\approx$ 10000 video, ứng với 130 câu.

Chuẩn hóa và làm sạch dữ liệu: Dữ liệu sau khi thu thập sẽ được lọc và chỉnh sửa để đạt tiêu chí ban đầu. Hiện tại, nhóm nghiên cứu đã xây dựng được tập dữ liệu gồm 130 câu, tương đương 7800 dữ liệu được xử lý.

Cài đặt thuật toán:

Qua MediaPipe, video sẽ được trích ra thành 60 khung hình kế tiếp nhau để nhận diện véc tơ hóa vị trí các đốt tay sau đó đẩy qua mạng nơ ron. Mô hình RNN cho phép xử lý thông tin từ những véc tơ vào kế tiếp, tuy nhiên càng đi xa thì lượng thông tin càng mất dần. Trong khi đó, thực tế yêu cầu phải ghi nhớ những thông tin ở xa nhau. Bởi vậy, nhóm sử dụng mạng LSTM cho phép điều chỉnh ghi nhớ thông tin ở gần hay xa theo dữ liệu học.

Tuy nhiên, thay vì dừng ở việc sử dụng mô hình sẵn có, với mô hình LSTM gốc (original model – *Hình 9*), nhóm tiến hành thay đổi cấu trúc mạng và chỉ số học (learning rate) của thuật toán tối ưu Adam (Adam optimizer) để có được phương án tối ưu. Ở các phần ***Luyện mô hình & Kiểm thử*** sẽ trình bày kết quả của các biến thể cấu trúc mạng.



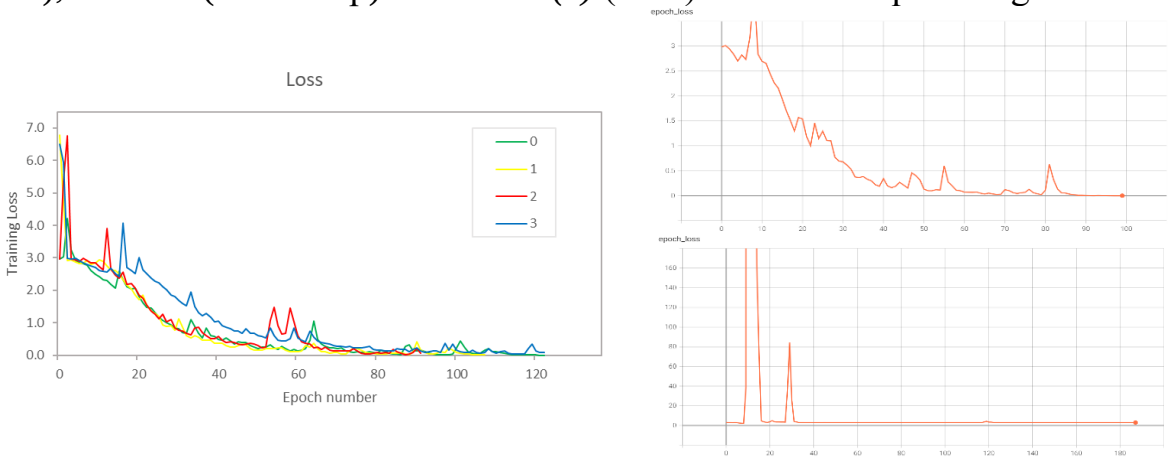
Hình 9. Sơ đồ chi tiết mô hình LSTM gốc

Luyện mô hình: Luyện mô hình đã thiết kế với dữ liệu được chuẩn bị, nhóm thực hiện hai phân đoạn chính:

Chuẩn bị dữ liệu: Tập đầu vào gồm 20 câu, với tổng lượng video là 1200. Dữ liệu cùng với label (nhãn – kết quả cho trước) sẽ được chia làm ba phần: (i) Tập huấn luyện (Training set) để luyện máy học, tối ưu các trọng số; và (ii) Tập kiểm tra (Testing set) để đánh giá mô hình sau luyện, theo tỉ lệ tương ứng 80:20.

Quá trình luyện mô hình: Hình 10. (a) là đồ thị hàm mất mát của cấu trúc mạng gốc và một số biến thể (xem *Bảng 1 – mục Kiểm thử)* trong quá trình luyện.

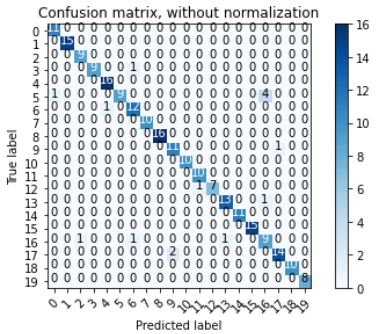
Trong quá trình luyện, phát sinh nhiều trường hợp overfit (quá khớp) như *Hình 10. (b)* (trên) do cấu trúc quá phức tạp để mô phỏng dữ liệu luyện (training data); underfit (dưới khớp) - *Hình 10. (c)* (dưới) do cấu trúc quá đơn giản.



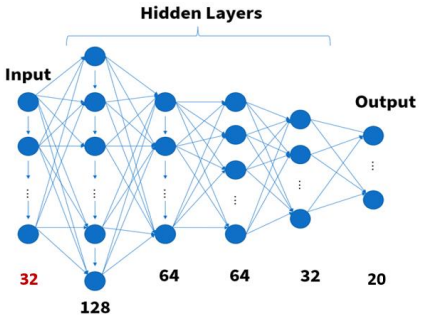
Hình 10. Hàm mất mát trong quá trình luyện mô hình. (a) Kết quả tốt. (b) Overfit. (c) Underfit

Kiểm thử: Với cả mô hình gốc và các biến thể, kiểm thử sau quá trình huấn luyện sử dụng ma trận nhầm lẫn (Confusion matrix), bằng cách cộng tổng đường chéo chính chia cho tổng của ma trận. Kết quả kiểm thử độ chính xác của một số bản đổi được liệt kê trong *Bảng 1*.

Như vậy, có thể thấy rằng độ chính xác tốt nhất đạt 93.75% của mô hình (2). Đây được coi là mức xuất sắc theo đánh giá chung về DL. Đồng thời, dựa vào kết quả ở đồ thị hàm mất mát và kết quả kiểm thử mô hình, nhóm quyết định lựa chọn cấu trúc mạng (2) là phương án tối ưu. *Hình 11, 12* lần lượt thể hiện ma trận nhầm lẫn và cấu trúc mạng của mô hình (2):



Hình 11. Ma trận nhầm lẫn khi kiểm thử



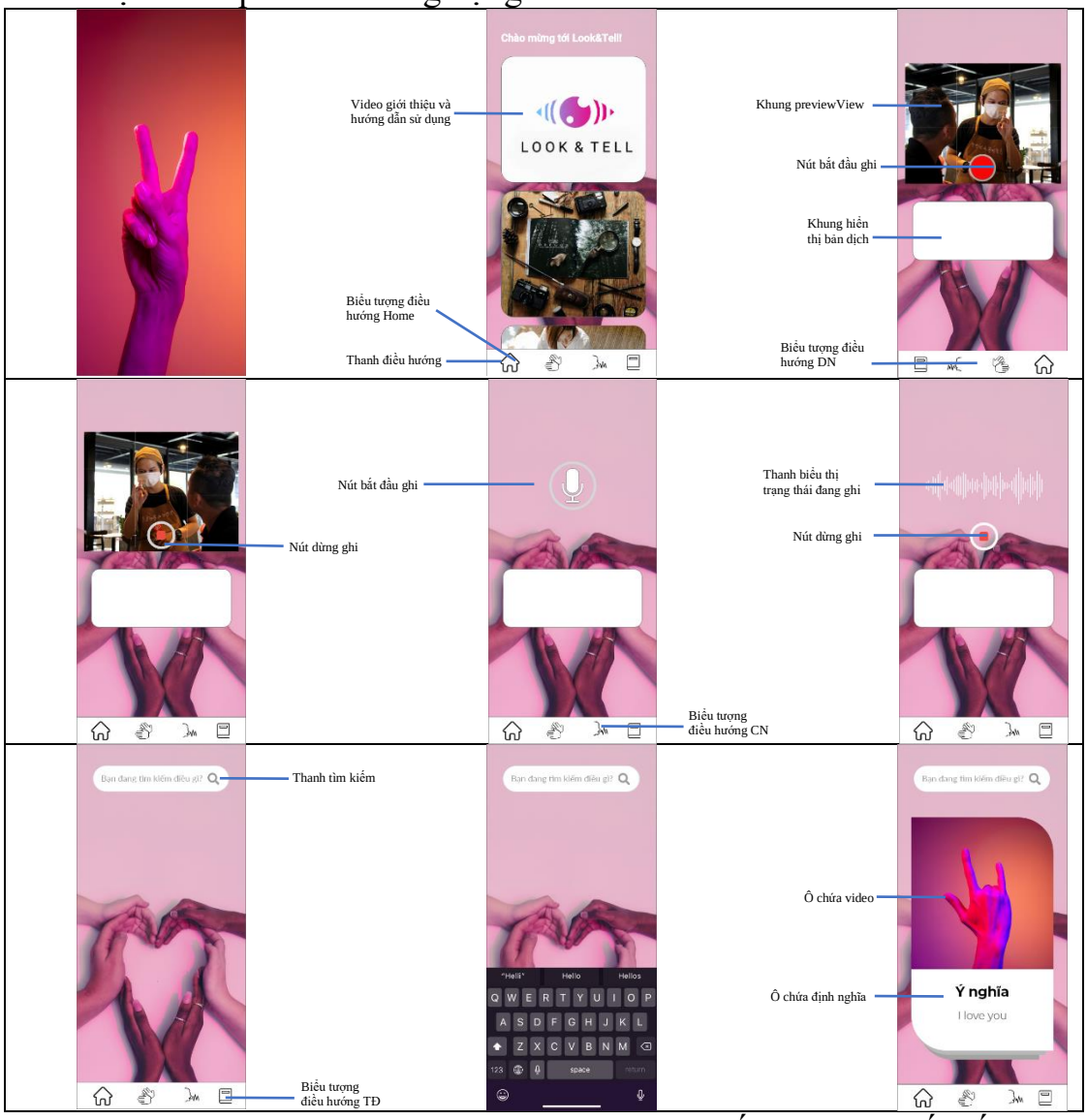
Hình 12. Cấu trúc mạng (2)

Bảng 1. Kết quả kiểm thử các cấu trúc trên cùng một tập dữ liệu

STT	Cấu trúc	Độ chính xác
0	LSTM 64 → LSTM 128 → LSTM 64 → Dense 64 → Dense 32	93,33%
1	LSTM 64 → LSTM 48 → LSTM 64 → Dense 64 → Dense 32	92,917%
2	LSTM 32 → LSTM 128 → LSTM 64 → Dense 64 → Dense 32	93,75%
3	LSTM 48 → LSTM 128 → LSTM 64 → Dense 64 → Dense 32	90,42%
4	LSTM 64 → LSTM 128 → Dense 64 → Dense 32	Quá khớp
5	LSTM 64 → LSTM 32→ LSTM 64 → Dense 64 → Dense 32	Dưới khớp

II.4.2 Cài đặt giao diện

Thiết kế hoàn chỉnh và tạo thông số kỹ thuật thiết kế: Tiến hành hoàn thiện giao diện. Các nội dung thiết kế trực quan và chức năng của từng yếu tố giao diện được diễn tả cụ thể để phát triển ứng dụng.



Hình 13. Giao diện hoàn chỉnh và các thông số kỹ thuật thiết kế. (a)Splash screen. (b)Home. (c) DN (1). (d)DN (2). (e)CN (1). (f)CN (2). (g)TĐ (1). (h)TĐ (2). (i)TĐ (3).

Cài đặt giao diện và chức năng: nhóm sử dụng các thông số kỹ thuật thiết kế và Android Studio - môi trường phát triển tích hợp (IDE) chính thức dành cho phát triển nền tảng Android. Ngôn ngữ lập trình được sử dụng:

- Phần cài đặt giao diện ứng dụng (API): ngôn ngữ đánh dấu mở rộng (eXtended Markup Language – XML).
- Phần cài đặt chức năng: ngôn ngữ nền tảng của bộ công cụ phát triển phần mềm Android (Android SDK) – Java.

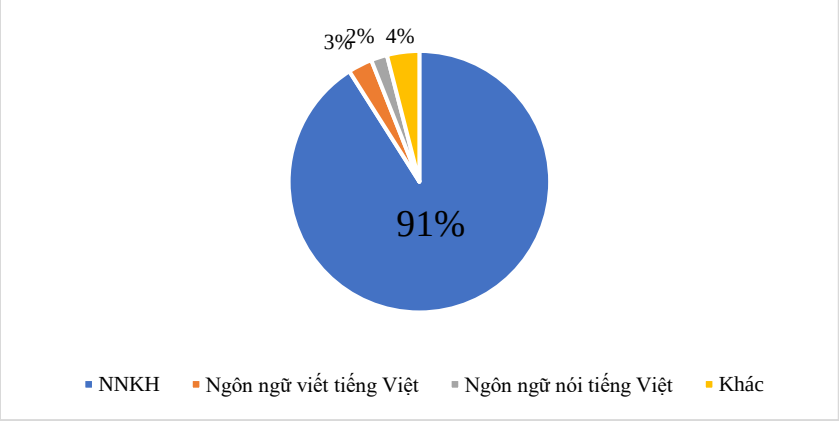
PHẦN III: KẾT QUẢ NGHIÊN CỨU DỰ ÁN

Trong mục này, các kết quả khảo sát, kiểm thử và chạy sản phẩm sẽ được trình bày cụ thể, làm tiền đề và cơ sở cho *Hướng phát triển dự án.*

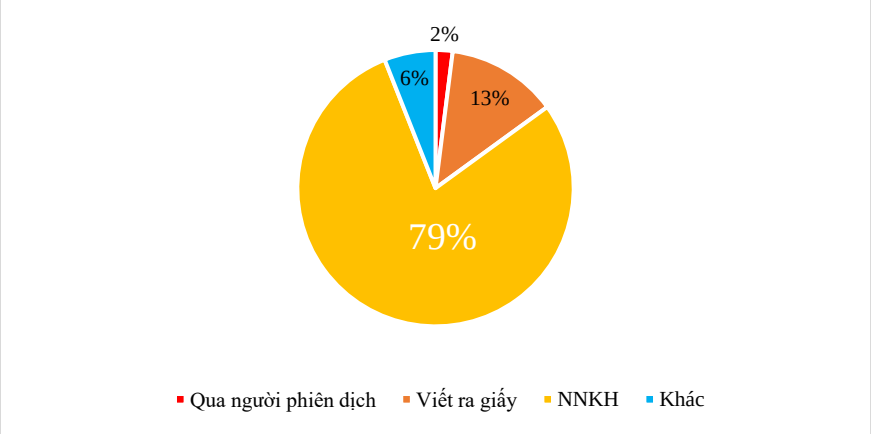
III.1 Kết quả khảo sát thực tế

Hình 14, 15 theo thứ tự cho thấy ngôn ngữ giao tiếp chủ yếu của người Điếc là NNKH; đây cũng là hình thức phổ biến nhất để người Điếc giao tiếp với người Nghe. *(hoàn thành mục 1 – I.3)*

Hình 14. Ý kiến 100 người Điếc về ngôn ngữ giao tiếp chủ yếu



Hình 15. Ý kiến 100 người Điếc về hình thức giao tiếp chủ yếu giữa họ và người Nghe



III.2 Đầu ra sản phẩm

Dựa vào các mục tiêu cụ thể, nhóm đưa ra đánh giá kết quả dự án như sau:

Hoàn thành một số mục tiêu đề ra trong mục I.3

Nghiệm thu ứng dụng Look&Tell ở phía nhà phát triển, nhóm nghiên cứu nhận thấy giao diện thiết kế đã hoàn thiện 100%. Đối với tính năng Dịch Ngữ, sản phẩm đã có thể nhận diện vị trí đốt tay trên ĐTTM; hiển thị label và xác suất lên bảng điều khiển máy tính. Ứng dụng cũng hoạt động ổn định với đặc điểm:

1. Yêu cầu hệ thống:
    - Dung lượng: 425MB
    - Hệ điều hành Android 8.1 - API 27 trở lên *(hoàn thành mục 5 – I.3)*
2. Xây dựng dựa trên các nền tảng thư viện, công nghệ như *mục 6,7 – I.3*

Cụ thể với mô hình DL trong chức năng DN:

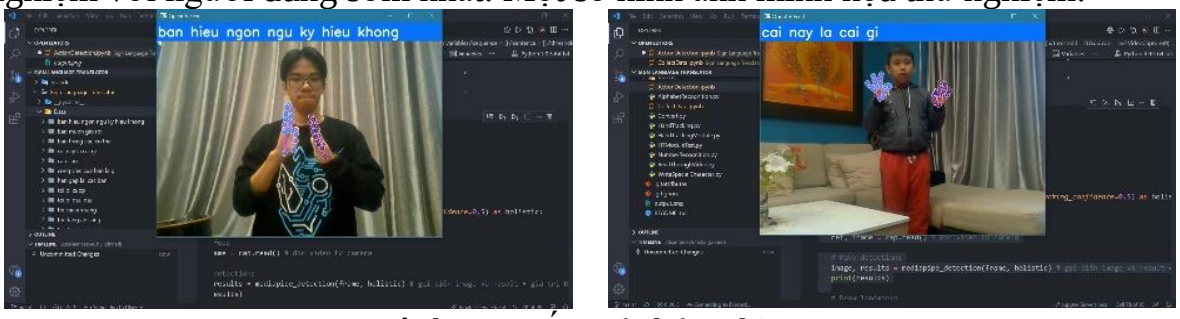
1. Lượng nhận diện: 20 câu *(hoàn thành một phần mục 2 – I.3)*
2. Tốc độ xử lý: 30 khung hình/giây *(hoàn thành mục 3 – I.3)*
3. Độ chính xác: 93,75% *(vượt kỳ vọng mục 4 – I.3)*

Tính khoa học đạt kì vọng: nhóm có đóng góp về mặt khoa học ở các phần:

- Bộ dữ liệu NNKH của người Việt Nam độ tuổi từ 10 tuổi trở lên gồm 7800 dữ liệu đã xử lý.
- Mô hình DL với khả năng nhận diện NNKH của người Việt được huấn luyện bằng bộ dữ liệu trên, thử nghiệm đạt chính xác 93,75%

III.3 Thử nghiệm sản phẩm

Hiện tại, sản phẩm vẫn đang được hoàn thiện trên ĐTTM để có thể thử nghiệm với người dùng sớm nhất. Một số hình ảnh minh họa thử nghiệm:



Hình 16. Kết quả thử nghiệm

PHẦN IV: KẾT LUẬN

Cộng đồng người Điếc Việt Nam là một cộng đồng người khuyết tật lớn, có nhu cầu giao tiếp, hòa nhập cao; mong muốn có những sản phẩm để giúp họ tiến gần hơn với cộng đồng. Đề tài đã kết hợp giữa nghiên cứu, khảo sát và thiết kế, xây dựng; thành công tạo dựng ứng dụng phần mềm Look&Tell như một trợ lý giao tiếp số hóa dành cho người Điếc Việt Nam. Ứng dụng chạy trên hệ điều hành Android, giao diện đẹp mắt, thân thiện với người dùng, rất hiệu quả và hoàn toàn miễn phí. Đồng thời, dự án cũng phát triển mô hình Deep Learning với độ chính xác 93,75%, tốc độ xử lý ổn định thời gian thực 30 khung hình/giây và xây dựng được bộ dữ liệu người Việt Nam với 7800 video.

PHẦN V: Ý NGHĨA

Về công nghệ Học sâu: Đóng góp dữ liệu về NNKH của người Việt vào bộ dữ liệu quốc tế. Đồng thời, các kết quả thực nghiệm thay đổi mô hình DL sẽ là tiền đề khoa học cho các nghiên cứu sau này.

Ý nghĩa nhân đạo – cộng đồng:

- Giúp người Điếc dễ dàng hòa nhập cộng đồng, không cảm thấy bị bỏ rơi.
- Thúc đẩy người trẻ tạo ra các sản phẩm có ý nghĩa nhân đạo, vì cộng đồng.

Về kinh tế - xã hội: Giúp giảm gánh nặng về chi phí xã hội và tăng hiệu quả lao động, cơ hội việc làm cho người Điếc.

PHẦN VI: HƯỚNG PHÁT TRIỂN

Về mặt sản phẩm:

- Cập nhật, sửa lỗi chức năng Dịch Ngữ, phát triển chức năng Từ Điển
- Phát triển thêm các tính năng mới như:
  1. Nhận diện NNKH qua video tải lên ứng dụng.
  2. Đưa mô hình DL lên server (máy chủ) được tối ưu tốc độ để giảm độ nặng ở ĐTTM, tăng tốc độ xử lý và tiện dụng để phát triển.
  3. Xây dựng ứng dụng trên hệ điều hành iOS.
  4. Phát triển dịch từ chữ viết sang NNKH để giao tiếp được đa chiều.

Về mặt dữ liệu NNKH: Nâng bộ dữ liệu lên 400 câu giao tiếp thông dụng.

Về mặt công nghệ:

- Nâng cao ứng dụng công nghệ xử lý ảnh
- Phát triển mô hình DL hiện tại vượt trội hơn bằng việc thay đổi cấu trúc mô hình và hàm tính toán.
- Nghiên cứu, cài đặt các mô hình DL nhận diện NNKH khác để đóng góp ý nghĩa khoa học và nâng cấp sản phẩm
- Thực hiện kiểm tra với các bộ dữ liệu khác trên mô hình đã xây dựng

TÀI LIỆU THAM KHẢO

[1] WHO. (2021, March 2). *https://www.who.int/news-room.* Retrieved from https://www.who.int:https://www.who.int/news-room/facts-in-pictures/detail/deafness

[2] WHO. (2021, March 2). *https://www.who.int/news.* Retrieved from https://www.who.int: https://www.who.int/news/item/02-03-2021-who-1-in-4-people-projected-to-have-hearing-problems-by-2050

[3] WFD. (2021). *https://wfdeaf.org/our-work/*. Retrieved from https://wfdeaf.org: https://wfdeaf.org/our-work/

[4] WHO. (2017). https://www.who.int/pbd/deafness/world-hearing-day. Retrieved from https://www.who.int: https://www.who.int/pbd/deafness/world-hearing-day/GlobalCostsOfUnaddressedHearingLossExeSum.pdf?ua=1

[5] David McDaid, A-La Park & Shelly Chadha (2021) Estimating the global costs of hearing loss, International Journal of Audiology, 60:3, 162-170, DOI: 10.1080/14992027.2021.1883197

[6] GSO. 2016. The National Survey on People with Disabilities 2016 (VDS2016), Final Report. Ha Noi, Viet Nam: General Statistics Office.

[7] VÂN, H. (2018, 1 5). Mở cánh cửa hi vọng cho người khiếm thính. Hà Nội, Việt Nam.

[8] Minh, T. (2020, 7 7). Găng tay thông minh giúp biến ngôn ngữ ký hiệu thành giọng nói. Thành phố Hồ Chí Minh , Thành phố Hồ Chí Minh, Việt Nam.

[9] Tư, K. (2021, 6 30). Nhóm sinh viên dùng AI chuyển đổi thủ ngữ sang giọng nói và văn bản. Thành phố Hồ Chí Minh, Việt Nam.

[10] Zhou, Z., Chen, K., Li, X. *et al.* Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nat Electron* **3,** 571–578 (2020). https://doi.org/10.1038/s41928-020-0428-6

[11] Coxworth, B. (2021, July 28). *https://newatlas.com.* Retrieved from https://newatlas.com: https://newatlas.com/wearables/sign-language-translation-glove/

[12] F. Soltani, F. Eskandari, and S. Golestan, "Developing a gesture-based game for deaf/mute people Using microsoft kinect," in *Proceedings of the 2012 6th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2012*, pp. 491–495, July 2012.

[13] A. K. Tripathy, D. Jadhav, S. A. Barreto, D. Rasquinha, and S. S. Mathew, "Voice for the mute," in *Proceedings of the 2015 International Conference on Technologies for Sustainable Development, ICTSD 2015*, February 2015.

[14] R. Kamat, A. Danoji, A. Dhage, P. Puranik, and S. Sengupta, "MonVoix-An Android Application for the acoustically challenged people," *Journal of Communications Technology, Electronics and Computer Science*, vol. 8, pp. 24–28, 2016.

[15] G. Subhaashini, S. Divya, S. DivyaSuganya, and T. Vimal, "Ear Hear Android Application for Specially Abled Deaf People," *International Journal of Computer Science and Engineering*, vol. 3, no. 3, pp. 1108–1114, 2015.

PHỤ LỤC

PL – Mã QR bảng câu hỏi khảo sát:

PL – Mã QR video demo sản phẩm: