SIGNTEGRATE - COMMUNICATION SUPPORT SOLUTION FOR DEAF PEOPLE IN VIETNAM BY SIGN LANGUAGE TRANSLATION APPLICATION USING ARTIFICIAL INTELLIGENCE

Category: **Science and technology**

NGUYEN TAI MINH & PHAM NINH GIANG

UNIT: Hanoi – Amsterdam Highschool for the Gifted

Hanoi Department of Education and Training

Hanoi, February 2022

TABLE OF CONTENTS

LIST OF SCHEMATIC TABLES

PART I: REASONALE OF THE PROJECT

I.1 Reasonale of the project

It was at the secondary school when we had the opportunity to participate in student extracurricular activities at Xa Dan Secondary School - Hanoi, where we experienced and got acquainted with a specialized class for deaf children as well as to be exposed to Sign Language for the first time. One of the problems of the deaf people in Vietnam in particular and around the world in general constantly mentioned by the teachers in our report at that time that was probably not too strange or new but still really a painful and difficult for science and medicine so far that was the language barrier in communication between the deaf people and the community.

In the current social situation, as the trend of modernization and globalization is at the forefront, the need for human connection and communication is increasing. However, when the Covid-19 epidemic situation is still seriously developing and there seems to be no end to make it difficult to communicate via the Online platform. Therefore, we realize that an application technology product that helps break down the language barrier in Gesture-Voice communication with convenience (available anytime and anywhere) will be an effective solution to contribute to bringing the deaf people closer to society.

I.1.1. Current situation of communication and integration of deaf people to community around the world

The proportion of deaf people (including Mute, Deaf, and Hard of Hearing) is increasing significantly compared to the total global population today. Along with that, there is no radical tool or solution for communicating with ordinary people (without affecting the funding). More than 5% of the world's population, including 432 million adults and 34 million children, are experiencing hearing loss and need for hearing rehabilitation. However, as of 2020, only 17% of those who need hearing aids actually use a hearing aid. It is estimated that nearly 80% of deaf people live in low- and middle-income countries (including Vietnam). The prevalence of hearing loss increases with age, and among people over the age of 60, more than 25% are affected by hearing loss.

In that context, most healthcare systems across the globe lack significant human resources to deal with ear diseases. The report said that "Among low-income countries, around 78% have less than one otolaryngologist per million population; 93% have fewer than one audiologist per million people; only 17% have one or more speech therapists per million people; and 50% have one or more teachers for a million deaf people."

According to the latest report by the World Health Organization (WHO), it is estimated that by 2050, nearly 2.5 billion people are predicted to have some degree of hearing loss, and at least 700 million people - one in ten people - will

have to recover their hearing. About 1.1 billion young people (ages 12-35) are at risk of permanent hearing loss due to unsafe listening practices. WHO estimates that unsupported hearing loss causes an annual global cost of US$980 billion. This includes health sector costs (excluding hearing aid), educational support costs, loss of productivity, and social costs. 57% of the costs come from low- and middle-income countries [1].

I.1.2. Current situation of communication and integration of deaf people to community in Vietnam

In Vietnam, according to the statistics of the Vietnam Population and Housing Census in December 2009, there are about 2.5 million deaf people in the country (including deaf people, hearing impaired people, and people with new hearing loss), accounting for nearly 3% of the total population at the time of the survey, mainly concentrated in big cities such as Hanoi and Ho Chi Minh City [2]. In particular, the deaf people are those who are totally unable to "hear and speak" and only use Sign Language as the main means of communication. Meanwhile, people with hearing loss, or as they are called in Cultural Perspectives as People with Hearing Impairments, are those who have limited ability to "hear, speak" and communicate, and use Sign Language as an auxiliary in communication [3]. Most deaf people in Viet Nam are low-middle-income workers, so they do not have the opportunity to be exposed to modern medical technology (such as hearing aids,...) but instead use Sign Language as their primary means of communication. Each region will use its own set of Sign Languages (quite similar and only slightly different due to the different nature of etiquette in each province/city); however, most deaf people in Vietnam use Hanoi Sign Language to popularize daily communication.

The message in the 2009 Millennium Development Goals is very clear that Our central goal is to help people with disabilities integrate into the community and this goal is an important factor for us to achieve the Millennium Development Goals as predetermined" *(Message from International Day of Persons with Disabilities 03 December 2009) [2].*

According to the "Survey on the situation of communication between deaf and ordinary people in Vietnam" conducted by the group themselves on the online platform Google Forms, as of December 11th, 2021, there are 247 survey participants, of which 83.4% (i.e. 206) reside in Hanoi and the proportion of Deaf People : Ordinary People who fills out a survey is 1 : 3.

Through the survey data of 54 people who did not have problems with "hearing and speaking" and had contact with deaf people, the number of people who encounter difficulties and obstacles when communicating with deaf people accounts for 96.3% (for example, some responses said that they had to say the same content many times; communication is severely damaged; and they had difficulty grasping and understanding the psychology of the deaf people). 70.4% of people have not known any Sign Language translation software for Vietnamese

people before, while the rest know foreign products. The number of people expecting the project's technology product to be able to translate Sign Language videos in real time in text and voice to optimize communication efficiency is 100%, including some suggestions to add other functions such as translating spoken/written language into Sign Language or teaching Sign Language with lessons according to a variety of topics.

Table 1.1: Information of the survey on 54 people who do not belong to the deaf community but have acquaintances with deaf people in Vietnam

No	Information	Number of respondents	Rate (%)
1	Difficulties and obstacles in communicating with the deaf people	52	96,3
2	If being a deaf person, the biggest difficulty in living, studying and working life is the language barrier in communicating outside the community.	53	98,1
3	Looking forward to a convenient support tool for the deaf people in Vietnam in communicating with ordinary people	52	96,3
4	Do not sign language translation tools or software to support communication between the deaf and ordinary people in Vietnam	38	70,4
5	Expect a video translation of Sign Language manipulation into the spoken text and voice of the product to maximize the opportunities of integration with the community of the deaf people	54	100

Through the survey data in Table 1.2, it can be seen that most people will have difficulty communicating with deaf people and do not know and expect the project's application product to be socially oriented. In the case of communicating with deaf people, the number of people choosing translation support software accounts for the most, followed by inconvenient, unfeasible and non-optimal ways such as writing information to be conveyed on paper, or learning Sign Language, hiring interpreters (according to 2019 statistics, Our country has about
1.5 to 2 million deaf and deaf people, but the number of professional Sign Language interpreters is just over 10, which is a huge difference [5]).

Table 1.2: Information of the survey on 132 people who do not belong to the Deaf community and have never had direct contact with deaf people in Vietnam

No	Information	Number of respondents	Rate (%)
1	If being a deaf person, the biggest difficulty in living, studying and working life is the language barrier in communicating outside the community.	127	96,2
2	Expect a convenient support tool in communicating with the deaf people as the optimal choice if having to communicate with them	65	49,3
3	Do not know sign language translation tools or software to support communication between the deaf and ordinary people in Vietnam	123	93,2
4	Expect a video translation of Sign Language manipulation into the spoken text and voice of the product to maximize the opportunities of integration with the community of the deaf people	125	94,7

A survey of deaf classrooms at Xa Dan Secondary School in Hanoi showed that 91.8% had the greatest difficulty in communicating with the community. In addition, some said that they were inferior and discriminated against, and it seemed that society did not fully understand the deaf peole's culture. Most of them are unaware of a Sign Language translation tool in Vietnam and want and expect a similar product with a meaningful purpose towards the community.

Table 1.3: Information of the survey on 61 teachers, parents, and students who are deaf at the pilot of Xa Dan Secondary School - Hanoi

No	Information	Number of respondents	Rate (%)
1	The biggest difficulty in living, studying, and working life is the language barrier in communicating outside the community.	56	91,8
2	Look forward to a convenient support tool for gthe deaf people in Vietnam in communicating with the ordinary people	57	93,4
3	Do not know sign language translation tools or software to support communication between the deaf and ordinary people in Vietnam	45	73,8

4	Expect a video translation of Sign Language manipulation into the spoken text and voice of the product to maximize the opportunities of integration with the community of the deaf people	59	96,7

I.1.3. Statistics on some studies on the application of Sign Language translation technology to support communication for people with hearing impairments

In recent years, the number of smartphone users is on the rise significantly. Specifically, in 2021 worldwide there are 6.378 billion smartphone users, and it is expected that this number will continue to increase to more than 7.5 billion by 2026 [4]. At the same time, in the complicated situation of the Covid-19 epidemic, all activities must be organized online, then the demand for mobile electronic devices is increasing. Therefore, communication support software will easily reach a large number of users, especially the Deaf community without affecting their financial resources.

Currently, there are many devices and applications with sign language translation features developed around the world. However, no product has really made a big impact on everyone partly due to limitations in terms of features and language data.

Table 1.4: Statistics of typical Sign Language translation products in the world

No	Product name	Developer	Information		Translation language	Function
			Form	**Operating system**		
1	HAND TALK TRANSLATOR [6]	HAND TALK	Application software	iOS & Android	American English Sign Language (ASL), Brazilian Sign Language (Libras) - English, Latin	Translation of speech and text languages into Sign Language using artificial intelligenc
2	English to Sign Language (ASL) Translator [7]	WeCapable	Website		American English Sign Language (ASL) - English	Translation of text language to sign language images

Through statistics, most of the current products and inventions still do not completely meet the convenience for users. Many products still require supporting

hardware equipment, causing financial difficulties for users. Meanwhile, in Vietnam, there have been many studies around the solution to the problem of language barrier between "the Deaf and Listeners" but currently has not been practically applied in the market. The new studies only revolve around simple letters and words, which are not practical in life.

Table 1.5: Statistics of typical studies in Vietnam and the world on sign language translation technology

No	Research/Invention	Implementing unit	Products - Functions
1	Wearable-tech glove translates sign language into speech in real time [8]	Bioengineer at UCLA - USA	The gloves are equipped with a sensor capable of recognizing American English Sign Language (ASL) operations through hand and finger movements, speaking text, and voice language through the phone's built-in application.
2	Automatic Feature Extraction for Vietnamese Sign Language Recognition using Support Vector Machine [9]	Faculty of Electrical & Electronics, Ho Chi Minh City University of Technology, VNU HCMC HCM, Vietnam	Using Vector Assistive Technology (SVM) to recognize and manipulate letters, marks, and numbers in the Ho Chi Minh City Sign Language table via webcam with the support of Microsoft Kinect devices
3	Deep Learning for Vietnamese Sign Language Recognition in Video Sequence [10]	Ton Duc Thang University Ritsumeikan University	Testing Vector Assistive Technology (SVM) and Deep Vietnamese sign language recognition (DVSL) to recognize 27 simple words in Vietnamese Sign Language (including 12 words of near-static operation and 15 words of operation) with the support of Camera Kinect (Microsoft)

Thereby, we hope to be able to bring Signtegrate to a large number of people, especially the Deaf community, and contribute to breaking down barriers in communicating using Hanoi Sign Language.

I.2. Research Questions

From the above questions, we have the following question: Is it possible to use information technology, especially artificial intelligence, to help solve the

problem of language barriers in communication between the deaf and ordinary people in Vietnam?

I.3. Research objectives

Collect data on familiar and common statements in daily etiquette in Vietnamese and Hanoi Sign Language.

I.4. Novelty and creativity of the project

Based on the above research results, the Signtegrate project team offers new and superior features such as:

- Hanoi Sign Language translation application is the first Vietnamese language processing function developed and tested for deaf people in Vietnam.

- It is Vietnam's first real-time dynamic video recognition processing application.

- It is the first software to recognize Sign Language translations in sentences, and the accuracy is not affected by background noise such as light, environmen,t or insignificant angle of inclination when shooting.

PART II: PROJECT IMPLEMENTATION PROCESS AND METHODOLOGY

The process will consist of the following steps:

- Survey the needs of Vietnamese people to use software to propose the design requirements of the problem.

- Collect and replicate data of Vietnamese language, a specific database that helps artificial intelligence algorithms well identify Vietnamese language.

- Build language recognition software that includes (1) processing data and generating inputs for recognition algorithms; (2) developing a recognition method that is highly accurate and can be run in real time; and (3) displaying the recognized language .

II.1. Software building process

Step 1: Collect and process input dynamic video data to generate output that is input to the machine learning model

Step 2: Train and test the accuracy of the model

Step 3: Build the software

Step 4: Test and perfect the software

Step 5: Maintain and upgrade the software

II.2. Software tools

II.2.1 Software

Jupyter notebook - a web-based programming platform
Web browser: Chrome

II.2.2 Programming languages

Python 3

II.2.3 Libraries

Opencv - accessing the computer's camera; MediaPipe [12] - using joint extraction; Numpy - arithmetic processing, matrix; and Keras and Tensorflow - building machine learning models.

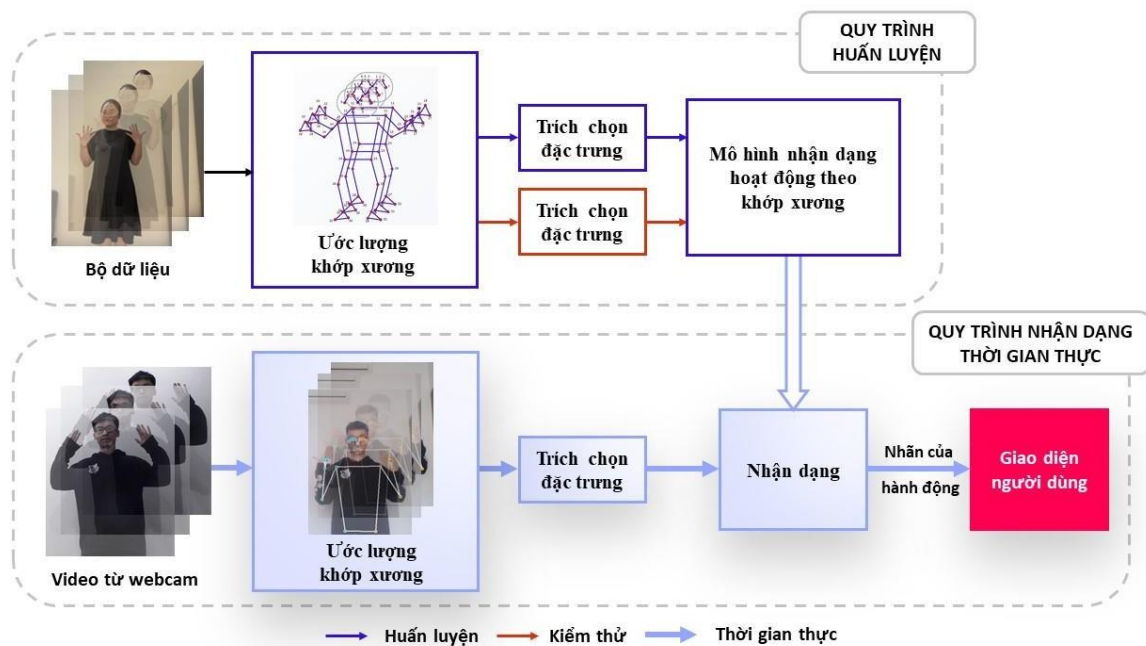### II.2.4 Hand sign recognition processing

Figure 1 - Identification model training process and real-time identification process

II.2.4.1 Identity training process

As Figure 1 depicts, we can see that the training process that creates an identity model consists of three main blocks, which are specifically described as follows:

Block 1: Processing input data

In this block, due to the limited time to collect input sentence samples for the project, at this stage we determined that the goal of the model is to identify the

9

10 most common sentences in daily communication in Vietnamese and Hanoi Sign Language (based on the popular official grammatical structure) including:

1. Xin chào, rất vui được gặp bạn!	1. Hello, nice to meet you!
2. Tạm biệt, hẹn gặp lại!	2. Bye, see you soon!
3. Xin cảm ơn, bạn thật tốt bụng!	3. Thank you, you are so kind!
4. Tôi xin lỗi, bạn có sao không?	4. I'm sorry, are you okay?
5. Tôi yêu gia đình và bạn bè.	5. I love my family and friends.
6. Tôi là học sinh.	6. I'm a student.
7. Tôi thích động vật.	7. I like animals.
8. Tôi ăn cơm.	8. I have rice.
9. Tôi sống ở Việt Nam.	9. I live in Vietnam.
10. Tôi là người Điếc.	10. I'm a Deaf.

(A set of common phrases in life etiquette built by the project has passed the feedback and accreditation of Hanoi Sign Language experts: Ms. Bui Phuong Anh, sign language teacher, Faculty of Special Education, Hanoi University of Education; Ms. Vu Hien, teacher of Deaf class, Xa Dan Secondary School - Hanoi; and Ms. Hoang Thi Anh Hong, teacher of Deaf Children - Hanoi.)

The first step of the model training was to enter the input video data: 394 self-constructed videos of 6 subjects (both deaf and ordinary people to diversify the dataset) were recorded from different angles, brightness, and distance from the camera. The data would be divided into two parts: 80% of the data would be used for training, and the remaining 20% would be used for testing of the model.
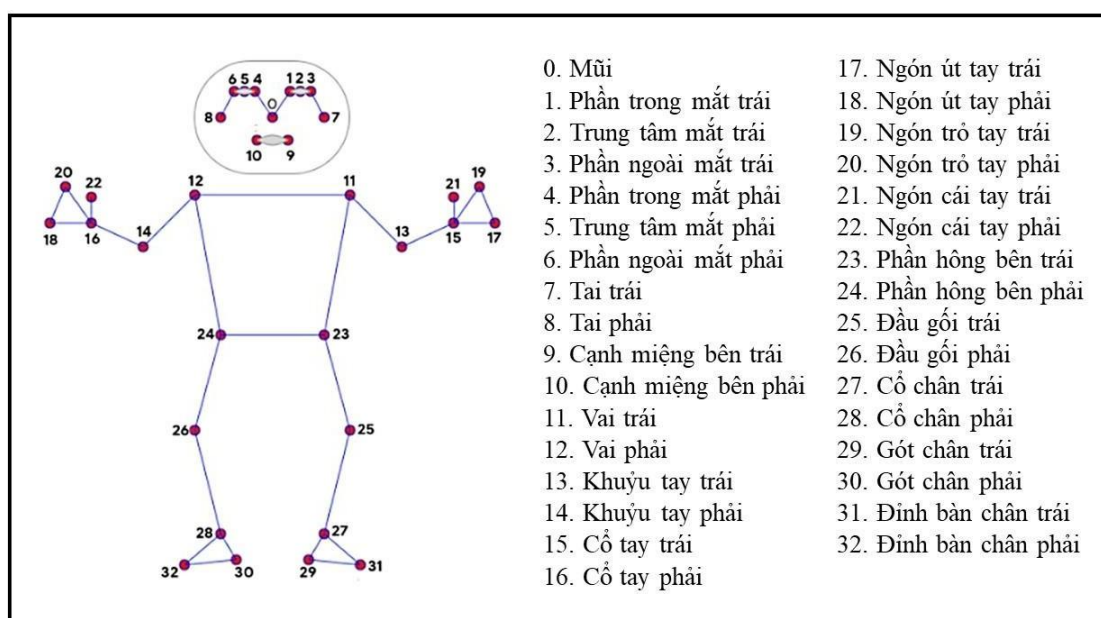
Block 2: Joint extraction

Figure 3: A human model according to the coordinate set of points [12]

To build a model that can process data quickly while ensuring high accuracy, we use the joint extraction algorithm through the MediaPipe library [12]. In each input video, the model will peel the video into 100-120 frames of its own. On each of those frames, the model will use a joint extraction algorithm to

select 33 human joints and save the coordinates of the joints and also the time parameter between the frames. We chose to use a joint extraction algorithm to minimize interference factors from the surroundings such as light, the color of the space behind the user or their clothes, or the person moving behind the user. In addition, the joint extraction algorithm minimizes the number of parameters compared to the original parameters of an initial dynamic video: using only 33 parameters corresponding to 33 joints. Therefore, the number of dimensions of the data set used to train the input decreases sharply, helping the training model to be fast while ensuring accuracy.

Block 3: Building a hand gesture recognition training model
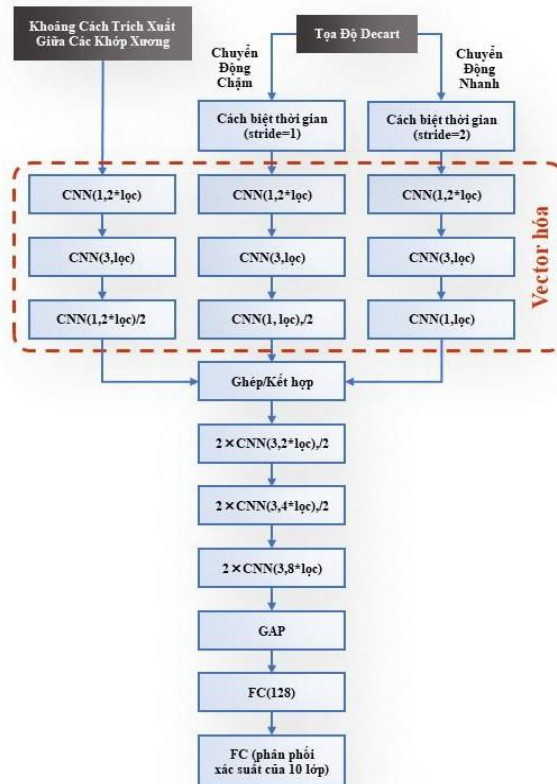

Figure 3: DD-Net machine model[13]

After Block 2, from the problem of the model of recognizing the actions of people according to dynamic videos, the identification problem switched to identification according to the coordinates of the joints. With the identity model, we used the DD-Net deep learning network structure [13] to solve this problem. From the preexisting skeletal coordinate data, the model will run 1000 times each data to be able to "learn" the characteristics of each action. In each run, the DD-Net structure [13] receives input data characterized by space and time. Space is expressed by the distance matrix between each pair of joints, time is considered

11

as the velocity or the change in coordinates of the joint through adjacent frames. The number of parameters is 1,862,410.

After Block 3, the identity model would eventually be rendered and installed into the overall software at the user's clients, for commercial identification applications.

II.2.4.2. Real-time recognition process

For the model to run in real time, the model needs to perform 3 processes. First, the model will import the necessary libraries such as Opencv, MediaPipe [12], Numpy, Tensorflow, Keras. Next, the computer will use the Opencv library to access the webcam for recording and at the same time use MediaPipe [12] to extract and save the coordinates of the joint. Once the client has recorded all 120 frames, it calls the trained recognition model to predict which of the 10 classes of dialogue used for training the action belongs to.

PART III: RESULTS

III.1. Product

- Real-time runtime: displays the contents of the action after 1 second since the model has captured a sufficient number of frames in the webcam.

- Image of the results of running a test of the Signtegrate application using a webcam:
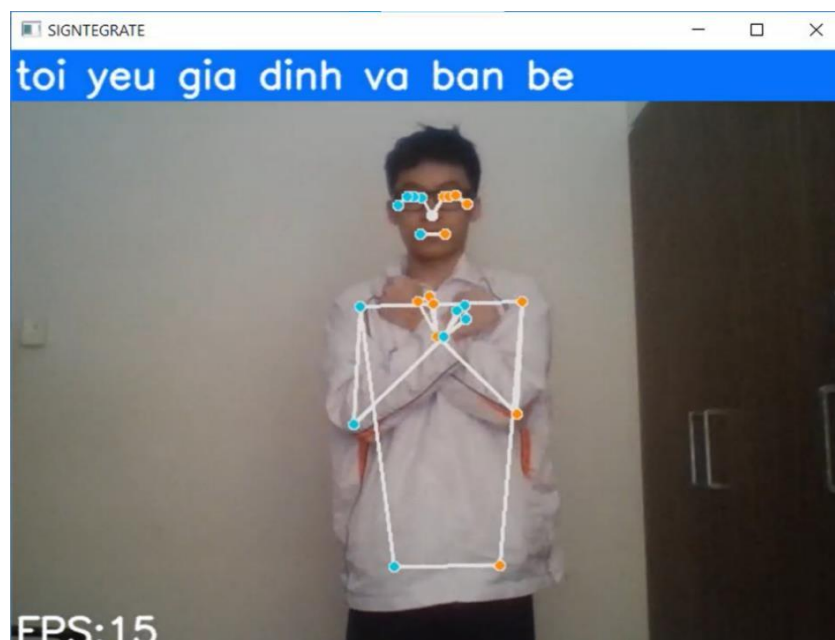
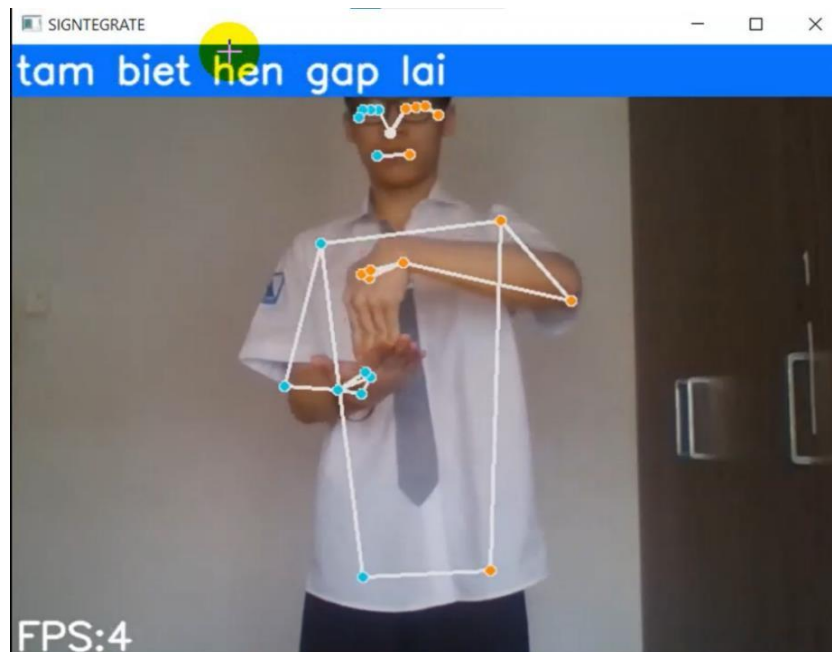*Figure 2 -* Image of the application when the user performs Tôi yêu gia đình và bạn bè"

Figure 2 - Image of the application when the user performs the action: "Tạm biệt. Hẹn gặp lại"

III. 2. Evaluation of the accuracy of the identification model

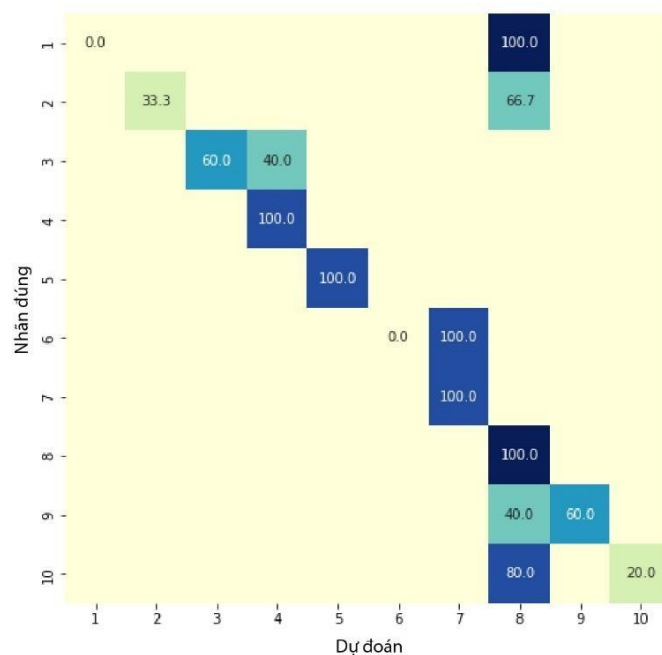The model after being trained was evaluated through a confusion matrix represented as follows:

Figure 3 - Confusion matrix

- Accuracy: 70.1% with data in the test set.

PART IV: CONCLUSION

- Have built a communication support solution for the deaf in Vietnam by applying artificial intelligence.
- The software has run on real time successfully and has good accuracy.
- This is a practical solution to break down the barrier between deaf and ordinary people in Vietnam.

PART V: SIGNIFICANCE

- About society: The Signtegrate application will break down language barriers, minimize difficulties in communication between ordinary people and deaf people, and contribute to helping the deaf community integrate into society. Vietnamese products will make it easier for Vietnamese people to access the application. This product will also be continuously upgraded to provide better applications exclusively for Vietnamese and Vietnamese people.
- About technology: Signtegrate application has contributed new methods to support people with disabilities, enriching the country's scientific and technical treasures.

PART VI: DEVELOPMENT DIRECTION OF THE TOPIC

- About sign language: Expand the input data set to 300 sentences/phrases common in everyday etiquette in Hanoi Sign Language.
- About technology: develop a user-friendly mobile application interface on the iOS platform.

BIBLIOGRAPHY

1. WHO (2021). Deafness and hearing loss.
2. Duong Phuong Hanh (2017). Understand and work with the hearing impaired. *Deaf Education Research Center (CED).*
3. Quynh Trang (2017). Barriers to access to construction by Deaf and Hard of Hearing people. *Column, Architectural Journal.*
4. S. O'Dea (2021). Number of smartphone users from 2016 to 2021. *Technology & Telecommunications, STATISTA.*
5. Hai Anh (2021). Sign Language interpreters, the picky "profession" of first-class learners. *24-hour education, Vietnam Education e-magazine.*
6. Hand Talk App. *HandTalk*
7. English to Sign Language (ASL) Translator. *Braille Tools, WeCapable.*
8. Matthew Chin (2020). Wearable-tech glove translates sign language into speech in real time. *Newsroom, UCLA.*
9. Pham The Hai, Huynh Chau Thinh, Bui Van Phuc, and Ha Hoang Kha (2018). Automatic Feature Extraction for Vietnamese Sign Language Recognition using Support Vector Machine. *2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom).*
10. Anh H. Vo, Van-Huy. Pham, and Bao T. Nguyen (2019). Deep Learning for Vietnamese Sign Language Recognition in Video Sequence.

International Journal of Machine Learning and Computing, Vol. 9, No. 4, August 2019.

11. Duc-Hoang Vo, Huu-Hung Huynh, Phuoc-Mien Doan, Jean Meunier (2017). Dynamic Gesture Classification for Vietnamese Sign Language Recognition. *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 3, 2017.*

12. Google (2020). MediaPipe Pose.

13. Yang, Fan & Sakti, Sakriani & Wu, Yang & Nakamura, Satoshi. (2019). Make Skeleton-based Action Recognition Model Smaller, Faster and Better