AI CLASSIFICATION MODEL TO READ SIGN LANGUAGE

Presented to Senior High School Department

Binangonan Catholic College

Binangonan, Rizal

In Partial Fulfillment

For the Subject Practical Research

Of the Strand

Science, Technology, Engineering, and Mathematics

Carlos Miguel M. Suela

Coryl Aikon SD. Cebrero

Rhone Ahprill A. Lobarbio

Carmelo Anthony Gonzal

Chapter 1

The Problem and Its Background

This chapter presents the background of the study, conceptual framework, objectives of the study, hypothesis, scope and limitations, significance of the study and definition of terms.

Introduction

Background of the Study

In today's world there are currently almost 1.5 billion people, nearly 20% of the world's population who are struggling from hearing loss, deafness, and 70 million of the world's population are mute (World Health Organization, 2022). Deafness refers to the person's lack of ability to hear, and mute refers to a person's lack of ability to speak. These two societal and physical problems set boundaries between people who are normal and with this disability. The impact could be, fewer opportunities or lower employment rates due to communication barriers, emotional problems due to a drop of self-esteem, and social withdrawal due to difficulties in socializing or communicating.

Sign Language significantly facilitates communication in the deaf community. Sign language is a language in which communication is based on visual sign patterns to express one's feelings.

The main problem here is that most or the vast majority of the people around the world does not know how to do sign language in fact, of the 48 million people in United States with hearing loss, less than 500,000 or about 1% use sign language, and very minimal of people are willing to learn them, since learning sign language is like learning a whole lot of a new language.

Although some individuals who are deaf since birth or profoundly deaf do have the ability to read English. According to Susan Goldin and Rachel Mayberry, their study entitled “How do profoundly deaf children learn how to read? (2001)” states that Padden and Ramsey, sign language researchers; had figured out a teaching technique called “Chaining”.

Chaining encourages children to see the relation between print and various sign systems the children know. This confirms that the procedure of an individual of mapping sign language into print (English), would be possible.

But in this case, study is still needed to create something that can make the communication gap between normal people and people with communication disabilities recede. In this study a goal is to help the deaf and mute community into having a much easier communication process.

AI or Artificial Intelligence is a study to make computers do things at which, AI is designed to automate simple human tasks up to a high complicated mathematical computation. People often think of Ai would grow into a highly intelligent cyborgs that raises awareness to people, but even if Artificial Intelligence have higher intellectual thinking capacity than humans, the hardware capabilities is still setting us boundaries to get through that.

Through research of intelligent systems, we can try to understand how the human brain works and then afterwards simulating it, turning it into a program

(Wolfgang Ertel, 2018). With the development in areas of deep learning and computer vision there is a way to simulate a human capability to read sign language.

And with that said utilizing one of Artificial Intelligence's capabilities which is computer vision, relatively with the use of hand gesture recognition and hand anchor point tracking automated sign language reader may be achieved. The features extracted are the binary pixels of the images. We make use of Convolutional Neural Network (CNN) for training and to classify the images (Murali, R. S. L., Ramayya, L. D., & Santosh, V. A. 2020).

Computer vision is a combination of image processing and pattern recognition. The output of the Computer Vision process is image understanding. Development of this field is done by adapting the ability of human vision in taking information. Computer vision has been expanded into the vast area of field ranging from recording raw data into the extraction of image pattern and information interpretation. It has a combination of concepts, techniques, and ideas from digital image processing, pattern recognition, artificial intelligence, and computer graphics.

Computer vision works by using an algorithm and optical sensors to stimulate human visualization to automatically extract valuable information from an object. Compared to conventional methods that take a long time and require sophisticated laboratory analysis, computer vision has been expanded into a branch of artificial intelligence and simulated human visualization (Wiley, V., & Lucas, T., 2017). Pattern Recognition as a branch of computer vision focused on the process of object identification through image transformation to get a better image quality and image interpretation. This process aims to extract information to make decisions based on images obtained from sensors. In other words, computer vision seeks to build an intelligent machine to "see."

As stated previously computer vision could be powerful when combined with Artificial Intelligence, but there are many factors to consider as well. First is the learning speed of an Ai when combined with this repository. Second is the software’s optimization, taken into the account we need the program to run, output computations, confidence rate, and results at a normal to a high speed of rate. And last would be the logic of the application of the program into detecting real time images.

TensorFlow could also be a high potential contestant into helping us create the backdoor engineering of computer vision. TensorFlow is an open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. TensorFlow can also be highly flexible in CNN architectures. TensorFlow allows developers to create dataflow graphs—structures that describe how data moves through a graph, or a series of processing nodes. We might also be using PyTorch which is also a library software but is much famous and considered more accurate than TensorFlow.

Even the struggles of having factors to consider and implementing different repositories into a program there are related studies that could help us succeed to the main goal of the product. The feature extraction is automated by using convolutional neural networks (CNNs). An artificial neural network (ANN) is used for classification.

Roel Verschaeren. He proposes a CNN model that recognizes a set of 50 different signs in the Flemish Sign Language with an error of 2.5%, using the Microsoft Kinect. Unfortunately, this work is limited in the sense that it considers only a single person in a fixed environment and does not support real time processing (Pigou, L., 2015).

CNNs are feature extraction models in deep learning that recently have been proven to be very successful at image recognition. It is a type of artificial neural network (ANN), which are nonparametric, machine learning (ML) methods that use interconnected neurons or nodes organized into layers to predict an output, such as classification label, from input data, such as image bands. DL expands the basic ANN framework by incorporating many hidden layers to allow the modeling of more complex patterns than what would be possible with a small number of hidden layers. In CNN classification, the input nodes for the classifier include not just a single pixel, but local groups of adjacent pixels. The CNN learning process incorporates the determination of appropriate convolutional operations and weights associated with multiple kernels or moving windows, allowing the network to model useful spatial context information at multiple spatial scales. (Aaron E. Maxwell, et al., 2021)

CNNs are inspired by the visual cortex of the human brain. The artificial neurons in a CNN will connect to a local region of the visual field, called a receptive field. This is accomplished by performing discrete convolutions on the image with filter values as trainable weights. Multiple filters are applied for each channel, and together with the activation functions of the neurons, they form feature maps. This is followed by a pooling scheme, where only the interesting information of the feature maps are pooled together. These techniques are performed in multiple layers.

Pooling is a key-step in convolutional based systems that reduces the dimensionality of the feature maps. It combines a set of values into a smaller number of values, i.e., the reduction in the dimensionality of the feature map. It transforms the joint feature representation into valuable information by keeping useful information and eliminating irrelevant information. Pooling operators provide a form of spatial transformation invariance as well as reducing the computational complexity for upper layers by eliminating some connections between convolutional layers.

Other than practically assessing derived thematic products, appropriate, consistent, rigorous, and well-documented accuracy assessment methods are key for benchmarking and quantifying improvements resulting from augmentation of existing and development of new DL methods.

As specific examples, Yang et al. compared their proposed hyperspectral band selection method with existing techniques to document improved classification performance while Abdalla et al., assessed their combined DL and k-means method for color calibration. Witharana et al., explored the impact of different data fusion and pansharpening methods on subsequent DL classification performance. DL techniques have been shown to be especially applicable for extracting information from high spatial or spectral resolution datasets. (Aaron E. Maxwell, et al., 2021)

The approach in 3d model uses the Microsoft Kinect to extract appearance-based hand features and track the position in 2D and 3D. The classification results are obtained by comparing a hidden Markov model (HMM) approach with sequential pattern boosting (SP-boosting). This resulted in an accuracy of 99.9% on 20 different isolated gestures on their specifically constructed data set and 85.1% on a more realistic one with 40 gestures (Dieleman, S., Kindermans, P. J., 2015).

Masood, Srivastava, Thuwal and Ahmad (2018) proposed a method to bridge the gap for the people who do not know and want to communicate using sign languages through isolated sign language recognition using methods based on computer vision., the output of the pooling layers was directly fed into the LSTM. The approach gave a result with an accuracy of 95.2%.

In the Philippines every Filipino, regardless of wealth, level of education, or location of residence has seen a shift in their life due to technology. Gadgets are widely used across all socioeconomic classes. Social media has become a part of everyday life for the typical Filipino. Technology has become a common learning tool. Students use their phones, laptops and other gadgets for research and their homework. Most classrooms are equipped with the latest technology for power points presentations that most teachers use for their lessons (Elfren S., 2017).

There are many problems in the Philippines regarding technology. The first is how will they be able to develop a strategy for individuals who will be negatively impacted. The rich will get even richer because they can afford the education required for the new economy. The classes that will be most affected are the lower and middle class. Another problem in the Philippines is the investment in human capital that is necessary to prepare for the next generation for the new economy. Technology in education shouldn't be reserved for the most elite private institutions. It needs to be a fundamental component of the Filipino educational system, especially public institutions. Not everyone can afford or have access to digital tools, which makes it hard for them to cope with the new normal. Many children, especially street children and out-of-school youth, don't have access to these tools. The government needs to provide these children their school needs. But the problem is reaching out to as many children as possible.

The internet speed is also one of the challenges that most Filipinos endure. The average download speed of a fixed internet connection in the Philippines was 81.42 Mbps in November 2022. On the other hand, the average mobile internet connection speed was 24.04 Mbps. The Philippines fell by seven notches to 61st out of 64 in the 2022 edition of the most technologically advanced countries ranking by international magazine Global Finance. The report ranks a country's technological strength across four metrics: internet users as a percentage of a county's population; LTE users as a percentage of the population; IMD World Competitiveness Center's Digital Competitiveness Score; and share of a country's research and development spending to this economic output. Among 11 East and

Southeast Asia countries included in the report, the Philippines ranked second lowest, ahead only of Mongolia (63r).

In the Philippines there is an existing law regarding into the improvisation of local technology and scientifical researches. According to the Republic act 2067 known as “Science Act of 1958” dated June 13, 1958 stated. An act to Integrate, Coordinate, and Intensify Scientific and Technological research and Development and to foster invention; to provide funds therefor; and for other purposes:

Legal Basis

Republic act no. 2067 “Science act of 1958”, *Section 2. In consonance with the provisions of section four, Article XIV of the Constitution, it is hereby declared to be the policy of the state to promote scientific and technological research and development, foster invention, and utilize scientific knowledge as an effective instrument for the promotion of National Progress.*

The researchers have decided to get along with the technological trend going around the world. Further research has led the assembly into the use of Artificial Intelligence, which gave the researchers the opportunity to take advantage of its capabilities. Philippines is a notch below when it comes in ranking with technology, so Philippines was quite indeed getting left behind in the technology industry. Furthermore, the researchers also wanted to improve the quality of technology and the knowledge of people regarding with technology.

The Deaf community is often misunderstood because of how they communicate through sign language, and normal people frequently would have difficulties in comprehending what they are saying when they make certain signals. To be able to communicate with people who are deaf or mute, we must learn sign language, although many people find it challenging to do so. This is why the researchers conducted this study: to help people who are deaf or mute, for us to easily communicate with each other and to break the barriers between the deaf community and the speaking people. With the use of technology, the researchers may look for effective methods of communication that will allow them to express their feelings as well as enable us to comprehend their message as it was intended.

Conceptual Framework

Through the evaluated purpose of developing the recent study, the research patterned by the figure utilizing Coomb's System Approach involving three frames indicating the input, process, output (IPO) model.

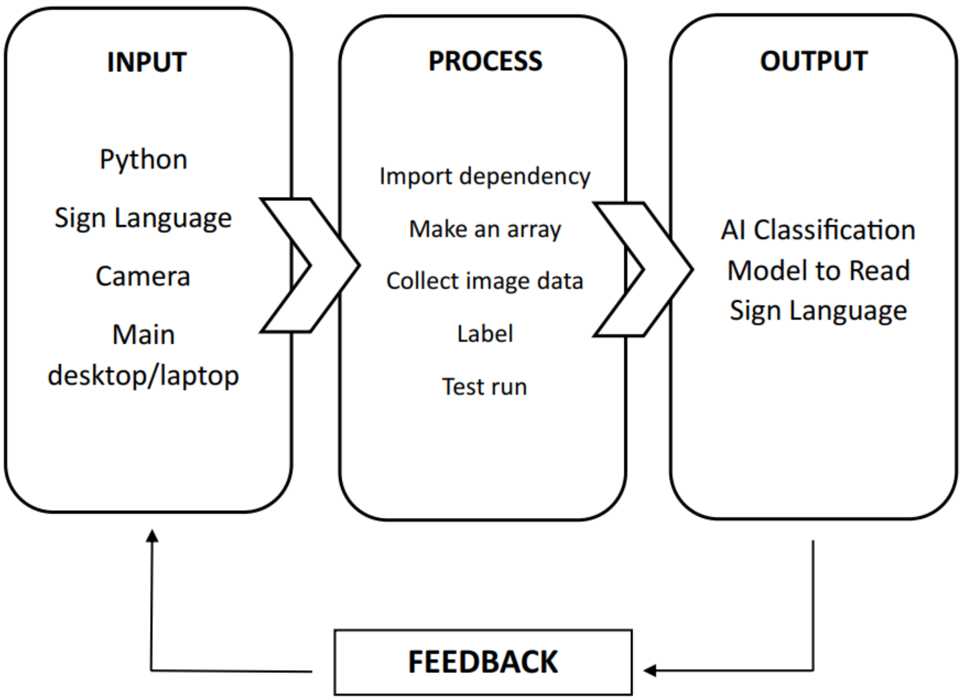


Figure 1.

Conceptual Framework of the study

The input frame contains the preparation of all the materials needed such as importing repositories, setting up local drivers, and the utilization of the programming language. After that, it will proceed to the process stage where it contains all the trials and errors in the development of the product, several tests and experiments that are conducted will be presented and different observation will be taken, in order to proceed on to the output process, which is to have a successful fully responsive and working AI model that can automate reading sign language.

Objectives of the Study

The main purpose of this study is to develop a software that measure the effectiveness of using AI Classification Model to read sign language.

Specifically, the study sought to answer the following:

1. What is the profile of the respondents according to?

    1.1 Age

    1.2 Sex

2. Effectiveness of using AI Classification Model to read sign language

    2.1 Accurateness

    2.2 Scalability

    2.3 Efficacy

Hypothesis

The Study entitled AI Classification Model in Reading Sign Language tested the null hypothesis:

H_0: There is no significant difference in terms of speed output, processes, and learning algorithms, in the epochs returned in automating the process of reading sign language, the main purpose is to automate reading sign language with the means of real time camera feed from a local device.

Scope and Limitations of the Study

The study entitled AI Classification Model in Reading Sign Language directly involves the utilization of the technologies' Artificial Intelligence and sign language in the production of a new software model, allowing speech-impaired individuals to converse with hearing individuals who are illiterate in sign language. The researchers' investigatory project covers the process of study and development of a software that can convert sign language into text using Artificial Intelligence. The researchers used English basic sign language, in which is inspired and referenced in ASL (American Sign Language).

This study was conducted by the Grade 11 (STEM) student of Binangonan Catholic College during the school year 2022-2023. The study utilized a quantitative approach. The limit of the study involves only auditory and speech impaired individuals, virtual users, and the within the use of specifically English, general and basic sign language in program related usage. The researchers gathered the data using research-made trial and error experimental tests.

Significance of the Study

This study has the potential to significantly improve the lives of deaf and speech-impaired individuals, as well as those who work with them or provide services to them. The result of this study will benefit the following:

For the Deaf community:

This study may help the deaf community and speech-impaired individuals by using technology. AI that can read sign languages could make it easier for them to communicate with hearing individuals that can't comprehend sign language. This could help to reduce barriers to education, employment, and social interaction.

Hearing People:

AI that can read sign language will also benefit hearing people who interact with deaf individuals on a regular basis. This includes anyone who might need to communicate with deaf people, such as family, friends, teachers, and others. It will make it easier for them to understand what is being communicated, leading to better communication and stronger relationships.

Education:

By offering real – time translation of lectures and classroom discussions, AI that can read sign language might increase the accessibility of education for the deaf community and speech-impaired students.

Emergency Services:

AI that can read sign language could be used by service providers, such as healthcare professionals or emergency responders, to communicate more effectively with the deaf community and speech-impaired individuals during emergencies or in other situations where clear communication is critical.

Definition of Terms

The following terms are defined for a better understanding of the study.

MediaPipe. would be utilized by the researchers for hand anchor point tracking.

AI Modeling. An intelligent system programming concept researchers would use for logical decision-making of the software based on available data.

GPU (Graphics Processing Unit). This hardware is optional in the study, but this can be utilized to accelerate the processing speed of the software program.

Machine Learning. An algorithm concept that would be used for AI Modeling.

OpenCV library. Utilizing OpenCV's library can help the researchers capture/gather data, and leverage the data, that can be used on the AI's learning algorithm.

Python. A high-level, general-purpose, object-oriented programming language, designed and developed by Guido van Rossum that is going to be utilized to script and code the main functions and behaviors of the AI software.

TensorFlow. An open-source library software made by Google, that can be used for big number computations, performance monitoring, and model training.

Chapter 2

Research Methodology

This chapter presents the discussion of the research design, setting of the study, subject of the study, instrumentation/sources of data, and procedure of the study.

Research Design

In this study, the researchers used the quantitative method. Quantitative research methods measure, analyze, and report relationships between variables to understand, describe, and predict a phenomenon (Mertz, 2016). The researchers conducted sets of experiments using the variables in this study which is Artificial Intelligence and Sign language.

Quantitative research is the process of collecting and analyzing numerical data. It can be used to find patterns and averages, make predictions, test causal relationships, and generalize results to wider populations (Pritha Bandari, 2020). The researchers gather sets of numerical data from the manipulation of the independent variable.

The researchers used the Quasi-experimental research, research involves the manipulation of an independent variable without the random assignment of participants to conditions or orders of conditions (ICA Chiang, 2015). The independent variable is manipulated. In this study the researchers will capture different factors, angles, and set of poses of sign language, and will undergo different sets of testing.

Like a true experiment, a quasi-experimental design aims to establish a cause-and-effect relationship between an independent and dependent variable (Lauren Thomas, 2020). The researchers manipulated the independent variable in a non-random criteria to attain the expected result of the product.

Settings of the Study

Barangay Mambog is one of the mainland barangays of Binangonan in the province of Rizal. Its population as determined by the 2020 Census was 11,372. This represented 3.63% of the total population of Binangonan; Mambog is located 2.87 kilometers away from the town proper and 2.58 kilometers away from the municipal hall. It borders Barangay Tatala as its north and west, Batingan as southwest and Mahabang Parang as south. Mambog borders Barangay Looc of Cardona, Rizal as its east. Mambog patron saint is Saint Joseph, it celebrates its patron every May 1, Mambog also celebrates Barangay Macamot's patron Saint Francis de Assisi every October. Mambog has only one public school of Casimiro Ynares Elementary School it also has several private schools such as Elim Christian School of Binangonan and Bridge of Light Grace Christian Academy Inc.

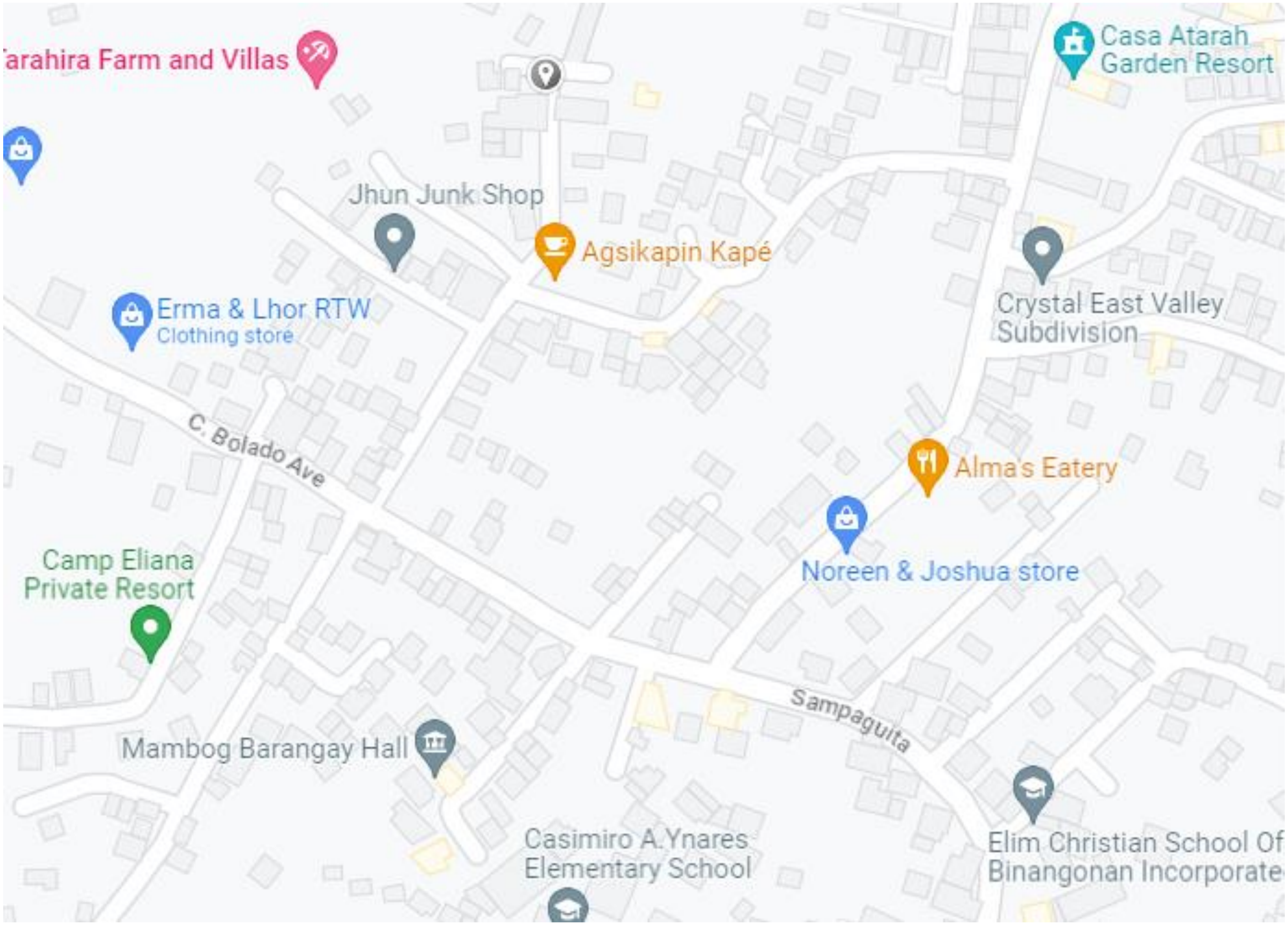


Figure 2.

Location Map of the Study

The image shows the map of vicinity in the settings of the study

Subject of the Study

The data collection method of this sample was through observation that is why the subject of the study was the product itself.

An AI Classification model is a type of machine learning software that uses different forms of learning algorithm that is designed to analyze and categorize data into specific groups based on patterns and features within the data. In this study, the AI classification model that can read sign language is the subject. It is trained using a large dataset of sign language gestures and uses machine learning algorithms to analyze sign language gestures and translate them into text. The model is designed to recognize different sign language gestures and convert them into accurate text translations.

The researchers selected 10 grade 11 students of Binangonan Catholic College to test and evaluate the elected product AI classification model to Read Sign Language.

Materials, Tools, and Equipment

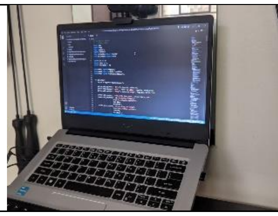
Picture	Name	Description
	Laptop/Desktop	Our main machine on where the product would be made and executed.
	Python 3	The main programming/scripting language that would be utilized for the AI's behavior.
	OpenCv	Would be used to gather images refine images and be used as a training data.
	Webcam	Hardware that would be utilized to gather images for data.
	Sign Language	Main data that would be gathered, and data that the AI should learn

Table 1.

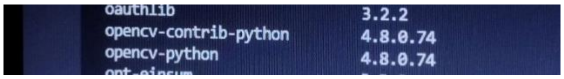
Materials and Equipment for product making

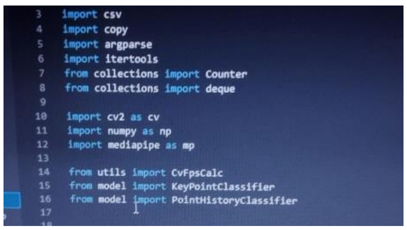
Procedure of the Study

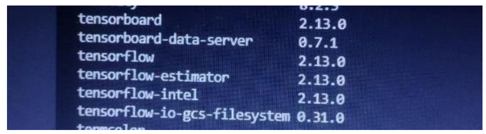
Product making procedures

The process of making the product can be done by the following steps:

1. Import libraries/install dependencies

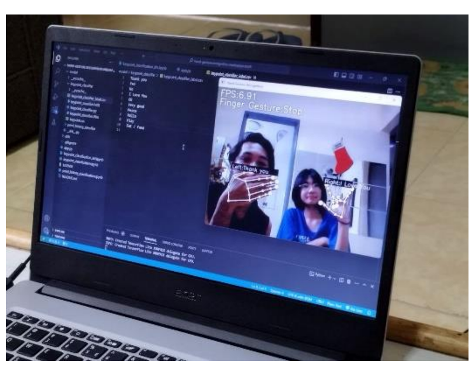




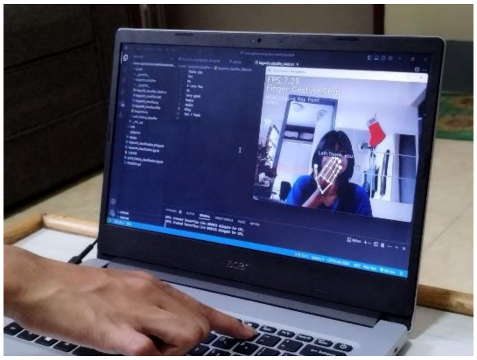


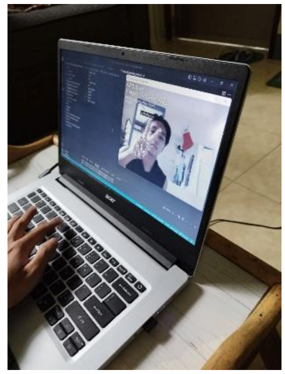
2. Make a comma separated array of sign language gesture labels in python.

3. Collect images through webcam using OpenCV.

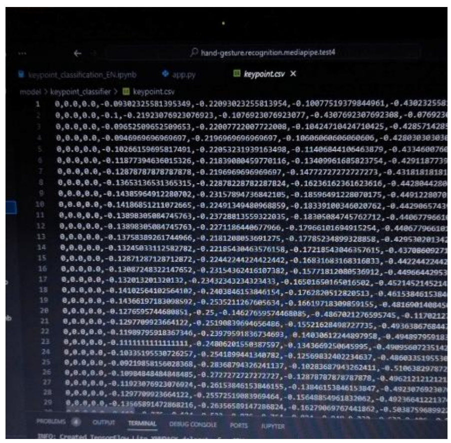


4. Label each image to corresponding sign language gesture using MediaPipe.

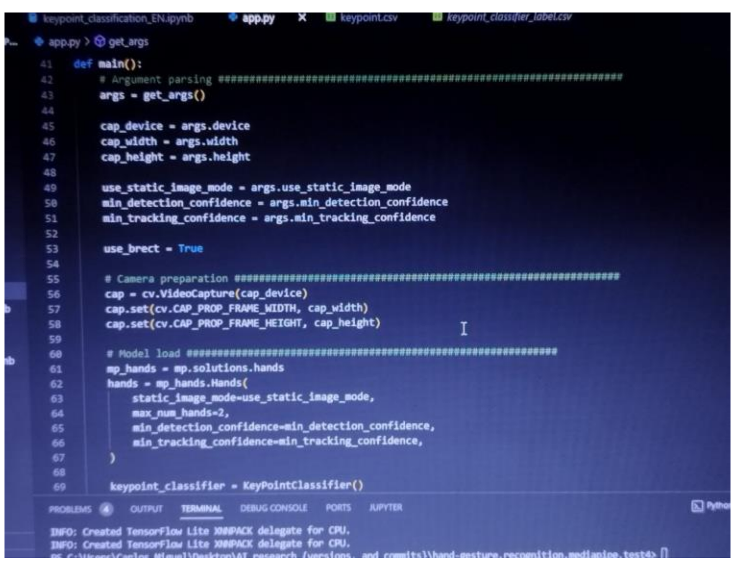




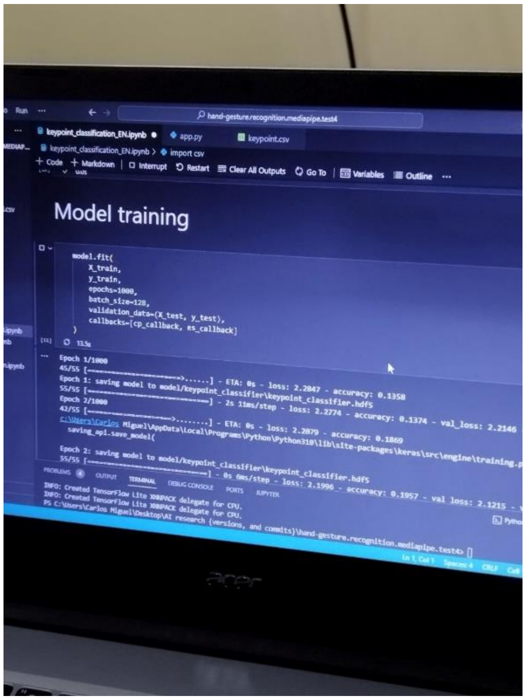
5. Leverage the recently captured images to csv file to be used as training data.

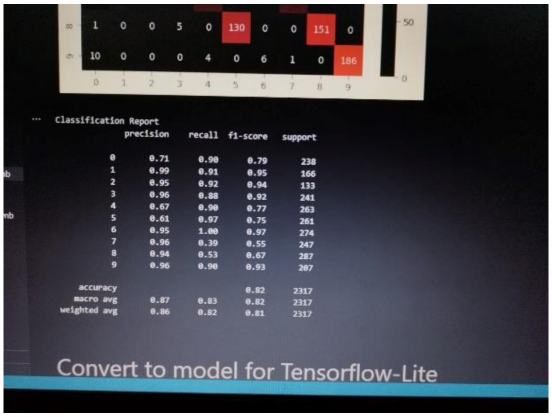


6. Code the behaviors of the AI model



7. If confidence rate is less than 50 (O (n log n) < 50%) collect more images, train another iteration, and test run.





The process of gathering respondent data for quantitative research can be done by the following steps:

1. Randomly select 10 respondents of grade 11 students of Binangonan Catholic College

2. Distribute the product to 10 selected respondents

3. Let the respondents test the product

4. Let the respondents honestly answer the provided questionnaires

5. Summarize gathered data

Figure 3

Gantt chart of the Study

Activity	Time Frame											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1. Preparation of the Title												
2. Title Defense												
3. Preparation of Chapter 1 and 2												
4. Revision of Chapter 1												
5. Submission of Chapter 2												
6. Revision of Chapter 2												
7. Colloquium												
8. Preparation of Chapters 3 and 4												
9. Revisions of Chapter 3 and 4												
10. Final Oral Defense												

Instrumentation

The type of instrumentation which the researchers used was observation and survey questionnaire. There was also a touch of structure observation for further strength of data and information being collected by the researchers.

The researchers planned and decided to have two different points and approaches in gathering the data. First is via the observation, researchers gather data through in such physical observation on the software. This includes user testing, functionality testing, and brute force testing. Second is observation via speed, and time complexities. With the use of the big O notation (O(n)) researchers can gather data of its learning algorithm, speed, and optimization.

A researcher made questionnaire checklist was used to gather the needed information and data. Part I pertains to the profile of the respondents. Part II pertains to the level of satisfaction of the respondents with the regards to the products accuracy, scalibility, and efficacy of the software to the end users. The respondents were asked to rate the items using the four-point scale as follows:

1 – Strongly Agree

2 – Agree

3 – Disagree

4 - Strongly Disagree

Chapter 3

Presentation, Analysis, and Interpretation of Data

This chapter represents the data gathered including the analysis and interpretation of data based on the sub-problems that were set at the outset of the study.

Several Trials and Observation

I. First Trial

METHODS	MATERIALS
Left and right palm recognition (by OpenCV)	**Cvzone (hand tracking module)**
Window app release and video capture	**OpenCV**
Scripting language	**Python**

Table 2.

Qualitative Measurement of materials for the first trial

On the first trial, the researchers tried experimenting on the given materials that they are going to use - OpenCV as the main framework, Cvzone a module for OpenCV and python for scripting. First step was to use the *"HandDetector"* function utilized from the Cvzone hand tracking module, set the minimum detection confidence threshold to 0.8 and maximum hands ("*maxhands*") into 2 and store in a local variable to create a physical bounding box that can be seen physically by the user. The next step was camera utilization, the researchers used python's native function *"input()"* to ask user for a number designated on their camera, and use the *"int()"* function to typecast the input into a number; store the result into a variable named “*cmC”*. Use the previous variable to use in OpenCV’s video capture function *“cv2.videoCapture(cmC)”.* The researchers moved on into creating a while-loop for frame by frame analysis, using the function *“video.read()”* for video detection, *“detector.findHands()”* for hand recognition and *“cv2.imshow”* for physical frame prediction. After scripting and setting up the processes the researchers utilized *“video.release()”* to compile all the scripts into a real time video detection application.

The researchers, after successfully interpreting the application, have begun their observation on the functionality of their product. Based on the observation, the product on first trial was able to successfully utilize the camera of the device used and was also successful in terms of detecting the hands of the user (either left or right), however it cannot detect any forms of sign language. In addition, the product lacked the use of training data and is incredibly reliant on grayscale imaging (to analyze depth) and mathematical computations. Hence, the product on the first trial the researchers has created did not satisfy the research goals.

II. Second Trial

METHODS	MATERIALS
Window app release and video capture	**OpenCV**
Hand landmark detection	**Mediapipe**
Scripting language	**Python**
“h5” data file reader/loader	**Tensorflow**

Table 3.

Qualitative Measurements of materials for the second trial

The second trial adds Tensorflow in the materials to process and unpack h5 type of data files, Mediapipe has also replaced OpenCV for hand detection in the first trial since Mediapipe is much lighter and optimal, this is because Mediapipe creates different points in a hand of the user; tracing it and forming a palm, unlike OpenCV which detects a palm as a whole which makes in heavyweight. Therefore, the components off the product are OpenCV, Mediapipe, Python, Tensorflow.

The first step in the second trial is to import and install Tensorflow, and Mediapipe. When installed via PYPI package manager, the second step is; likewise the first trial use OpenCV’s function *“cv2.VideoCapture()”* to open and utilize the device’s camera. Next is using Mediapipe’s holistic hand module *“mp.solution.hands”.* The final step is creating a while-loop for frame by frame analysis, likewise in the first trial. And loading the h5 datafile set using tensorflow's *"tf.keras.models.load_model"* function.

The product was unsuccessful to run and interpret, this is from the failure of unpacking and opening the datafile using tensorflow's keras. Keras is rensponsible for hdf5 data files and not h5 data files.

III. Third Trial

METHODS	MATERIALS
Window app release and video capture	**OpenCV**
Hand landmark detection	**Mediapipe**
Scripting language	**Python**
Model training and compiling	**Tensorflow**
CSV (comma separated values) file interpreter (Data file)	**Numpy**
Confusion matrix and accuracy report	**Matplotlib and SciKit-learn**

Table 4.

Qualitative Measurements of materials for the third trial

On the third trial the researchers added Numpy, SciKit-learn, and Matplotlib to the materials being utilized in the process of the product. Due to different testing failures the researchers decided to change the methods in gathering the training data that would be supposedly used by the model. Instead of gathering images and leveraging it as a training data, the researchers decided to log the coordinates of different hand landmarks in an invisible cartesian plane.

The first step for the third trial is camera preparation, researchers again used OpenCV's *"cv.VideoCapture()"* and *"cap.set()"* function to utilize the local camera and automatically adapt and set into local camera's width and height. Next step is creating a while-loop function for frame-by-frame analysis and video recognition, inside the while-loop function would be a *"cv.cvtColor(image, cv.COLOR_BGR2RGB)"* function stored in the "image" variable to pre-process each frame for detection and converting into a colored RGB image. On the third step a hand key-point classifier/labeler would be created using Numpy as CSV interpreter and Tensorflow as model classifier; all of the gathered coordinate data would be stored in a file named "keypoint.csv". Fourth step is scripting the program to log the coordinates of the detected hand landmarks by holistic mediapipe into the csv file using native python "if" and "pass" statement function, here's the code snippet for logging coordinates.

```
def logging_csv(number, mode, landmark_list, point_history_list):
    if mode == 0:
        pass
    if mode == 1 and (0 <= number <= 9):
        csv_path = 'model/keypoint_classifier/keypoint.csv'
        with open(csv_path, 'a', newline="") as f:
            writer = csv.writer(f)
            writer.writerow([number, *landmark_list])
    if mode == 2 and (0 <= number <= 9):
        csv_path = 'model/point_history_classifier/point_history.csv'
        with open(csv_path, 'a', newline="") as f:
            writer = csv.writer(f)
            writer.writerow([number, *point_history_list])
    return
```

Next step is using *"cv.line"* function by OpenCV to physically display the detected hand landmarks by Mediapipe. The final step is model building, using Tensorflow's *"tf.keras.layers()"* the researchers can create lightweight but functional layers of recurrent neural networks and parameters. For model compilation the researchers used *"model.compile()"* function, and for model training *"model.fit()"* function by Tensorflow is used.

```
# Layers
model = tf.keras.models.Sequential([
    tf.keras.layers.Input((21 * 2, )),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(20, activation='relu'),
    tf.keras.layers.Dropout(0.4),
    tf.keras.layers.Dense(10, activation='relu'),
    tf.keras.layers.Dense(NUM_CLASSES, activation='softmax')
])
# Model compilation
model.compile(
    optimizer='adam',
    loss='sparse_categorical_crossentropy',
    metrics=['accuracy']
)
# Model training
model.fit(
    X_train,
    y_train,
    epochs=1000,
    batch_size=128,
    validation_data=(X_test, y_test),
    callbacks=[cp_callback, es_callback]
)
```

The product after training and compiling, successfully detects hand land marks, and classified different sign language gesture based on training data with an accuracy of 83% and time complexity of O(n log n).

Presentation of Data

Distribution of 'Age'

Age	Count of Age
16	8
17	1
18	1
Total	**10**

Distribution of 'Sex'

Sex	Count of Sex
F	6
M	4
Total	**10**

Figure 4.

Profile of the respondents

The researchers tested the product on 4 (40%) male, 6 (60%) female students. Ageing 16 with a count of 8 (80%) respondents, 17 with 1 (10%) respondent, and 18 with 1 (10%) respondent. With a total of 10 (100%) students in Binangonan Catholic College.

SEX	MALE		FEMALE	
ACCURATENESS	MEAN	VERBAL INTERPRETATION	MEAN	VERBAL INTERPREATION
1. The product accurately recognizes different sign language gestures	1	STRONGLY AGREE	1.16	STRONGLY AGREE
2. The product accurately translates different sign language gestures without error.	1.75	AGREE	1.66	AGREE
3. The product accurately classifies sign language despite of having poor video/image quality.	2	AGREE	1.33	STRONGLY AGREE
4. The product accurately identifies different sign language gestures in situations where it is shown from various angles and lighting conditions.	2	AGREE	1.5	STRONGLY AGREE

SCALABILITY	MEAN	VERBAL INTERPRETATION	MEAN	VERBAL INTERPREATION
5. The product can scale up to two palms per user.	2.25	AGREE	1.33	STRONGLY AGREE
6. Constant changes (maintenance) doesn't affect major functionalities of the software.	2	AGREE	1.33	STRONGLY AGREE
7. The product can still function despite the absence of online connectivity.	2.25	AGREE	2	AGREE
8. Rapid adapt in performance (FPS) based on hardware capabilities.	1.75	AGREE	1.66	AGREE

EFFICACY	MEAN	VERBAL INTERPRETATION	MEAN	VERBAL INTERPREATION
9. The product can be used effectively by auditory impaired individuals.	1.5	STRONGLY AGREE	1.33	STRONGLY AGREE
10. The product can improve the access of communication between speech impaired and hearing individuals.	1.5	STRONGLY AGREE	1.16	STRONGLY AGREE
11. The product can be used to reduce the information barrier between speech impaired and hearing individuals	1.25	STRONGLY AGREE	1.33	STRONGLY AGREE
12. The product can be used by hearing individuals.	1.5	STRONGLY AGREE	1.33	STRONGLY AGREE

AGE	16		17	
ACCURATENESS	MEAN	VERBAL INTERPRETATION	MEAN	VERBAL INTERPREATION
1. The product accurately recognizes different sign language gestures	1.12	STRONGLY AGREE	1	STRONGLY AGREE
2. The product accurately translates different sign language gestures without error.	1.75	AGREE	1.5	STRONGLY AGREE
3. The product accurately classifies sign language despite of having poor video/image quality.	1.62	AGREE	1.5	STRONGLY AGREE
4. The product accurately identifies different sign language gestures in situations where it is shown from various angles and lighting conditions.	1.75	AGREE	1.5	STRONGLY AGREE

SCALABILITY	MEAN	VERBAL INTERPRETATION	MEAN	VERBAL INTERPREATION
5. The product can scale up to two palms per user.	1.75	AGREE	1.5	STRONGLY AGREE
6. Constant changes (maintenance) doesn’t affect major functionalities of the software.	1.62	AGREE	1.5	STRONGLY AGREE
7. The product can still function despite the absence of online connectivity.	2	AGREE	2.5	AGREE
8. Rapid adapt in performance (FPS) based on hardware capabilities.	1.75	AGREE	1.5	STRONGLY AGREE

EFFICACY	MEAN	VERBAL INTERPRETATION	MEAN	VERBAL INTERPREATION
9. The product can be used effectively by auditory impaired individuals.	1.37	STRONGLY AGREE	1.5	STRONGLY AGREE
10. The product can improve the access of communication between speech impaired and hearing individuals.	1.25	STRONGLY AGREE	1.5	STRONGLY AGREE
11. The product can be used to reduce the information barrier between speech impaired and hearing individuals	1.25	STRONGLY AGREE	1.5	STRONGLY AGREE
12. The product can be used by hearing individuals.	1.37	STRONGLY AGREE	1.5	STRONGLY AGREE

Table 5.

Respondents view on product’s quality in terms of sex and age.

Level of agreement:

1.00-1.44 – Strongly Agree, 1.45-2.44 – Agree, 2.45-3.44 – Disagree,

3.45-4.00 – Strongly Disagree

The statements that attained the highest mean based on the respondents' sex are the item no. 1 in which states that the product accurately recognizes different sign language gestures, item no. 10 in which states the product can improve the access of communication between speech impaired and hearing individuals. To sum up the highest mean, the respondents strongly agreed about the accuracy and the efficacy of the AI classification model in detecting and classifying different sign language gestures. The statement that attained the lowest mean is item no. 7 in which the product can still function despite the absence of online connectivity. In this statement, the respondents agreed.

Based on the respondents' age the item that gained the highest mean is item no. 1 in which states that the product accurately recognizes different sign language gestures. To sum up the highest mean, the respondents based on age strongly agreed about the product's ability to accurately identify, classify, and detect sign language gestures. The statement that attained the lowest mean is item no. 7 in which the product can still function despite the absence of online connectivity. In this statement, the respondents agreed.

Presentation of Data

The researchers prepared and gathered all the required prerequisites to make an AI classification model. Setting up IDE, updating Python to its latest version, and installing required libraries/prerequisites via terminal using PyPi (Python's package manager). Once completed, the researchers started off on scripting the main functions of the AI, first is camera utilization and video recognition using OpenCV turning frames into a grayscale to enable the computer to determine the depth of a frame. After that, Holistic Mediapipe is used to enable the computer vision model to detect hand landmarks. Then scripting to log every points of landmarks in an invisible cartesian plane into the desired sign language gesture. After the data was gathered, together with Tensorflow layers are created for the model to enable the AI for the training process based on the gathered data.

Set A: <u>AI classification model using native OpenCV, CNN (Convolutional Neural Networks).</u>

Based on the result of product set A, the classification model using alone OpenCV as a framework, and CNN as a neural network, though functional but was ineffective, the product was able to detect left and right hands however it is struggling to classify and segment different sign language. According to the observation the researchers believe that the lack of training data on different training factors (e.g. Low quality, low exposure images, etc.) is what results for the difficulties on the segmentation. However, knowing the complexity map of Convolutional Neural Networks, collecting huge amounts of the specific kind of data could be extremely resource consuming.

Trial I and II

Accurateness	Scalability	Efficacy
The model couldn't provide the basic functions of video classification. Not allowing it to depict any observations for accuracy.	The model can provide and do functions as it was intended to. However, the method on how the application was built isn't optimized for all types and fields of devices.	The product is ineffective when dealing and compiling in a low-level hardware.

Table 6.

Results of observation based on the accurateness, scalability, and efficacy of the product (Set A-first hand data)

Set B: <u>AI classification model using OpenCV, Mediapipe, Tensorflow – RNN (Recurrent Neural Networks), and LSTM (Long Short-Term Memory) Method.</u>

Based on the observation from the set A, CNN is a resource consuming method since the input and output are independent from each other, this means that if there is a need for another word prediction previous inputs are required. This shows that the memory is required to map back through different neural networks to attain the previous prediction and flag it as an input. Unlike CNN, that is a multi-layer neural network, RNN provides looping nodes, this means that the previous output could be used and fed as an input for the next step, it also showcases its most important feature which is the Memory state. This refers that the previous states and input in a network could be stored and remembered – this reduces the complexities of the parameters making it minimal and optimizable for the device, since if any hidden layer is present in the kernel when previous inputs for a next prediction is needed; all of the previous states/inputs are stored in the hidden layer thus the neuron can thoroughly map back to the kernel which is near the output node.

Trial III

Accurateness	Scalability	Efficacy
The model can provide sign language classification and hand tracking detection at an accuracy of 82%	With the use of Recurrent neural network, much simpler layer minimal layers were attained. Thus, making the model much less resource consuming that it can be run and compiled in a much low-level hardware.	Its ability to recognize different gestures is effective to be used for target audiences and users.

Table 7.

Results of observation based on the accurateness, scalability, and efficacy of the product (Set B-second hand data)

Chapter 4

Summary, Conclusions and Recommendations

This chapter presents the vital findings of the study, the conclusions drawn and recommendations offered based on the specific problems of the study.

Summary of findings

After conducting an observation and test comparing the AI model using CNN (Convolutional Neural Networks) and the AI model using RNN (Recurrent Neural Networks) the researchers found out that there are significant differences between two in terms of accuracy, scalability, and effectiveness. AI model built within RNN's is more accurate than models built within CNN's, RNN's with its recurrent feature is constantly learning data from both inputs and outputs unlike CNN's that has independent input and outputs. Models built within RNN's are much more scalable than CNN's, this is due to the simplicity of RNN's neural maps unlike CNN's which are multi-layered maps. AI models built in RNN's are more effective than CNN's, considering RNN's adaptive performance.

Based on the analysis and integration of the data gathered for this study, the following were found:

1. There is a significant difference in terms of scalability and accuracy between RNN and CNN.

2. An RNN based AI Classification Model to Read sign language is better for faster predictions.

3. CNN based model can still be used when dealing with huge data of raw images.

Conclusion

This study was conducted to make an AI Classification Model to Read Sign Language. The model was in a software form and can be redistributed into devices within proper installation and use of the prerequisites. The experimental method of research was utilized. The descriptive method and survey technique were used for gathering data. The AI Classification Model to Read Sign Language was the basis of the respondents to answer survey questions. The questionnaires served as the instrument for collecting data. The chosen people are grade 11 students of Binangonan Catholic College. The inquiry was conducted during the school year 2023-2024. This section also answers the questions as stated in the objectives, based on the research conducted survey and researchers' experimental method:

Age and Sex

Based on the research conducted survey, students in the age of 16 agreed within the product's accuracy, scalability, and efficacy. Meanwhile, students aged 17 strongly agreed within the product's accuracy, scalability, and efficacy. Referencing to the respondents' sex, male students of grade 11 agreed on the product's accuracy, scalability, and efficacy. According to the scatter of response the female grade 11 students strongly agreed on the product's accuracy, scalability, and efficacy.

Accurateness, Scalability, and Efficacy

According to the product's accuracy. Based on the researchers' experimentation and analysis. The product has provided an accuracy level of 0.82

or 82% on the inference test. According to the researchers conducted survey, the accurateness category has produced an average mean response of 1.41 in female and 1.68 in male, for a total scatter of 1.55. To sum up, the respondents agreed on the product's accuracy. Furthermore, in accordance with the product's scalability. Based on the researchers' experimentation and analysis. The product has shown an improvement in its ability to scale into different instances within the change to RNN. According to the researchers conducted survey, the scalability category has produced an average mean response of 1.58 in female and 2.06 in male, for a total scatter of 1.82. To sum up, the respondents agreed on the product's scalability. Lastly, based on the objectives, the product's efficacy. Onto the researchers' experimentation and analysis. The product can effectively run and provide its functions to benefit the users. According to the researchers conducted survey, the efficacy category has produced an average mean response of 1.29 in female and 1.43 in male, for a total scatter of 1.36. To sum up, the respondents strongly agreed on the product's efficacy.

As hypothesis stated, there is no significant difference in terms of speed, output and processes on every epoch returned. The researchers proves that it is possible to create an automated software using an AI Classification model. Though as shown in the significant difference of CNN and RNN's, different methods have different outcomes in terms of accuracy, scalability, and optimizability. Hence, the researchers reject the null hypothesis of the study.

Recommendations

Based on the findings and conclusions, the following recommendations are set forth.

1. Use Recurrent Neural Networks for minimal layers, but decently high accuracy.
2. Since RNN's training data are number coordinates, there's a need for more accurate training data since it doesn't refer to the image alone. Also, to gather only decent amounts of training data, more data in RNN's means more confusion for the model.

In conducting a new study related to the current investigatory project, the researchers suggested that:

1. Test the products specifically to hearing impaired individuals.
2. Create models that can handle more diverse and larger training data.
3. Make the AI model available for offline usage.

References

Goldin-Meadow, S., & Mayberry, R. I. (2001). How Do Profoundly Deaf Children Learn to Read? https://journals.sagepub.com/doi/abs/10.1111/0938-8982.00022

Ertel, Wolfgang (2018). Introduction To Artificial Intelligence: Second Edition. https://books.google.com.ph/books?hl=en&lr=&id=geFHDwAAQBAJ&oi=fnd&pg=PR5&dq=Through+research+of+intelligent+systems,+we+can+try+to+understand+how+the+human+brain+works+and+then+afterwards+simulating+it,+turning+it+into+a+program+(Wolfgang+Ertel,+2018).+&ots=3G2u3aaF4s&sig=qK4lcqWn_YwXYmvBxVeu9vcYSbI&redir_esc=y#v=onepage&q&f=false

Murali, R. S. L., Ramayya, L. D., & Santosh, V. A. (2020). Sign Language Recognition System Using Convolutional Neural Network and Computer Vision. https://www.ijeiat.com/images/sliders/92dc4915b9487f589cfe29a2d410cc0a.pdf

Wiley, V., & Lucas, T. (2017). Computer Vision and Image Processing: A Paper Review. http://download.garuda.kemdikbud.go.id/article.php?article=624673&val=10290&title=Computer%20Vision%20and%20Image%20Processing%20A%20Paper%20Review

Pigou, L., Dieleman, S., Kindermans, P.-J., & Schrauwen, B. (2015). Sign Language Recognition using Convolutional Neural Networks. https://core.ac.uk/download/pdf/55693048.pdf

Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review.

https://www.mdpi.com/2072-4292/13/13/2450

Masood, S., Srivastava, A., Thuwal, H. C., & Ahmad, M. (2018). Real-Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN and RNN.

https://link.springer.com/chapter/10.1007/978-981-10-7566-7_63

Cruz, E. S. (2017, September 8). Challenges of Technologies. Philstar.com.

https://www.philstar.com/opinion/2017/05/10/1698836/challenges-technologies

Philippines lags in National Technological Strength List. BusinessWorld Online.

(2022, May 29). https://www.bworldonline.com/infographics/2022/05/30/451342/philippines-lags-in-national-technological-strength-list/

“I love you” in sign language. Lingvano ASL. (2022, February 21). https://www.lingvano.com/asl/blog/i-love-you-in-sign-language/#:~:text=The%20sign%20for%20%E2%80%9CI%20love,I%2DL%2DY%20for%20I%20love%20you.

Birley, J. (2020, September 23). 7 helpful ASL signs you should know: Rick Hansen Foundation. https://www.rickhansen.com/news-stories/blog/7-helpful-asl-signs-you-should-know#:~:text=ASL%20for%20Yes%2F%20No,resembling%20a%20mouth%20saying%20no.

Curriculum Vitae

Carlos Miguel M. Suela

Mambog, Binangonan Rizal

09771582004

carlosmiguelsuela@gmail.com

PERSONAL INFORMATION:

Date of birth: October 10, 2005

Place of birth: East Avenue medical center Diliman, Quezon City

Age: 18

Sex: Male

Civil Status: Single

Height: 5’2

Religion: Catholic

Citizenship: Filipino

Father: Charlie S. Suela

Mother: Winnie C. Marquez

SKILLS AND INTEREST:

- Logical analysis
- Creativity
- Adaptability
- Problem solving
- Video editing / graphic designing
- programming

LANGUAGE

- Filipino
- English

EDUCATIONAL BACKGROUND:

Senior High School: Binangonan Catholic College

Adress: Libis, Binangonan Rizal

School Year: 2022-2024

Secondary: Binangonan Catholic College

Adress: Libis, Binangonan Rizal

School Year: 2017-2024

Primary: Bridge of Light Grace Christian Academy

Adress: Mambog, Binangonan Rizal

Rhone Ahprill A. Lobarbio

Bagong bayan Darangan, Binangonan Rizal

09071132015

ahronelobarbio11@gmail.com

PERSONAL INFORMATION:

Date of birth: August 16, 2006

Place of birth: Quezon Province

Age: 17

Sex: Male

Civil Status: Single

Height: 5’3

Religion: Catholic

Citizenship: Filipino

Father: Ronnel E. Lobarbio

Mother: Arnette Aprille A. Alvano

SKILLS AND INTEREST:

- Fast learner

- Time management
- Teamwork & management

LANGUAGE:

- Filipino
- English

EDUCATIONAL BACKGROUND:

Senior High School: Binangonan Catholic College

Adress: Libis, Binangonan Rizal

School year: 2022-2024

Carmelo Anthony Gonzal

495 M. H. Del Pilar St. Libid Binangonan, Rizal

09620750102

carmelogonzal19@gmail.com

PERSONAL INFORMATION:

Date of birth: October 29, 2004

Place of birth: Cardona, Rizal

Age: 19

Sex: Male

Civil Status: Single

Height: 5’5

Religion: Roman Catholic

Citizenship: Filipino

Mother: Olivia S. Gonzal

SKILLS AND INTEREST:

- Interpersonal Communication
- Microsoft Word Proficiency
- Computer Literate

- Time Management
- Teamwork & Management

LANGUAGE:

- Filipino
- English

EDUCATIONAL BACKGROUND:

Senior High School: Binangonan Catholic College

Adress: Libid, Binangonan Rizal

School Year: 2022-2024

Secondary: Binangonan Catholic College

Adress: Libid, Binangonan Rizal

School Year: 2018-2022

Primary: Binangonan Catholic College

Adress: Libid, Binangonan Rizal

School Year: 2013-2018

Coryl Aikon SD. Cebrero

Lunsad, Binangonan Rizal

09362528079

aikoncebrero@gmail.com

PERSONAL INFORMATION:

Date of birth: August 15, 2006

Place of birth: Lying Inn Maternity Clinic

Age: 17

Sex: Female

Civil Status: Single

Height: 5’3

Religion: Catholic

Citizenship: Filipino

Father: Roberto Cebrero

Mother: Conception Cebrero

SKILLS AND INTEREST:

Teamwork

Fast learner

Microsoft Word proficiency

LANGUAGE:

- Filipino
- English

EDUCATIONAL BACKGROUND:

Senior High School: Binangonan Catholic College

Adress: Libis, Binangonan Rizal

School Year: 2022-2024

Secondary: Binangonan Catholic College

Address: Libis, Binangonan Rizal

School Year: 2012-2018

Primary: Lunsad Elementary School

Adress: Lunsad, Binangonan Rizal