POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

TOUCHLESS INTERFACE AS AN ALTERNATIVE TO PUBLIC TOUCHSCREEN TECHNOLOGIES

A Thesis
Presented to the Faculty of Computer Engineering
Polytechnic University of the Philippines
Sta.Mesa, Manila

In Partial Fulfilment of the Requirements for the Degree
Bachelor of Science in Computer Engineering

by

Nombrado, Paul Joseph A.
Esguerra, Scott John V.

February 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

THE PROBLEM AND ITS SETTING

This chapter discusses the overview of the research problem. This includes the Introduction, Background of the Study, Theoretical Framework, Conceptual Framework, Statement of the Problem, Scope and Limitations, Significance of the Study, and Definition of Terms.

Introduction

The COVID-19 pandemic has led to a dilemma in human society and caused a huge loss to the lives of many people worldwide. The rapid spread of the devastating pandemic caused disruption worldwide making tens of millions of people suffer to the negative symptoms and effects of the virus, currently estimated nearly 181 million confirmed cases recorded by the World Health Organization (WHO) as of June of 2021 and increasing every day. Most of the commercial places like the mall, supermarket and other enterprises are facing an existential crisis due to lockdowns and community quarantines. Furthermore, continuing this effect could put our workforce at risk of losing their livelihood. Economy workers and consumers are particularly vulnerable because the majority lack social protection quality health care especially in the Philippines. According to WHO, the Philippines already have 1.38 million reported cases of COVID-19 and around 24 thousand death in the said month.

Even though the number of cases in the Philippines seem quite small compared to the overall population which have around 110 million, the main problem is that the number of cases continues to grow on about four percent weekly due to the rapid spread of the virus mostly in concentrated areas like the mall, supermarkets and commuting sites. The number of cases can grow massively over the years if ignored or not prevented.

The COVID-19 is a type of contagious disease that can spread quickly airborne. Based on Hong Zhou, a microbiologist at University of California, Los Angeles (UCLA), basically the corona virus enters a cell and multiply and sometimes a millions of times. All of the copies then spill out into the patient’s airways and some of the copies end up in tiny droplets of fluid and when the patient cough or sneezes, those droplets that can store thousands of virus particles spray out onto tabletops, railings, bus seats, food, other people, etc. The virus can easily jumped from person to person and also can stick to surfaces that the infected individual touched. Study shows that the corona virus can live on surfaces for days so this can be devastating mostly on commercial areas where most people gather. The transmission occurs when the person touches the contaminated surface and then touches their mouth, nose or eyes. Therefore, it is possible to get infected by making contact with the contaminated object or through the air.

Since most of the used objects that people always make contact of are from the touchscreen devices like the navigation application where people get directions to go certain places and locate their way inside the mall or the ticket vending machine in Light Rail Transit (LRT) and Metro Rail Transit (MRT), these are some of the dangerous and high risk places where transmission of corona virus could occur. The researchers came up with a solution to prevent contact between the said devices and still give service to the consumers by harnessing touchless interface through image recognition and machine learning algorithms. The system will serve as an alternative to touchscreen interfaces currently found within public and private establishments. The system would use hand and movement tracking technologies with additional advancements for the system to cater towards its purpose as an alternative and/or replacement that supports touchscreen. The project aims to establish a safer and risk-free interactions as healthcare would be a big priority in order to lessen the spread of COVID-19 and prevent such pandemic from ever happening again.

Theoretical Framework

Machine Learning. In order to recognize a person's hand in the interface, the researchers would place a camera to capture the typical dynamic hand gestures like swiping, scrolling or pressing a button that the user makes when using the interface of the device. This study would use machine learning algorithms in order for the machine to analyze the different hand patterns and understand it real time when the user operates the interface. The way the machine learning algorithm captures the "pattern" in practice is by computing a function that takes an image as input and produces the appropriate label as output. (Kubát, 2017).Machine learning is a technique used for analyzing data and creates automatic analytical model. It is a branch of artificial intelligence that identifies different patterns and make decision for itself with less human interaction.

Image Processing. Image recognition or processing is crucial in this study because the device needs computer vision to be able to understand images or videos in the outside world. The task of recognizing what the image or video clip represent rely on computer vision in order for it to work. The recorded dynamic hand patterns is necessary to be processed first to make them more suitable as input data for the machine learning algorithm. The specific shade, color and/or opacity of every pixel of the image or pixel within frame of the video clip is needed to be understand before analyzing the input data to be put on the neural network for training.

Convolutional Neural Network. Convolutional neural network is a technique within machine learning with its effectiveness in pattern recognition. One of the most popular deep neural networks is the Convolutional Neural Network (CNN). It take this name from mathematical linear operation between matrixes called convolution. CNN have multiple layers; including convolutional layer, non-linearity layer, pooling layer and fully-connected layer. The CNN has an excellent performance in machine learning problems. Especially the applications that deal with image data, such as largest image classification data set (Image Net), computer vision, and in natural language processing (NLP) and the results achieved were very amazing. (Albawi, Mohammed, & Al-Zawi, 2017)

Conceptual Framework

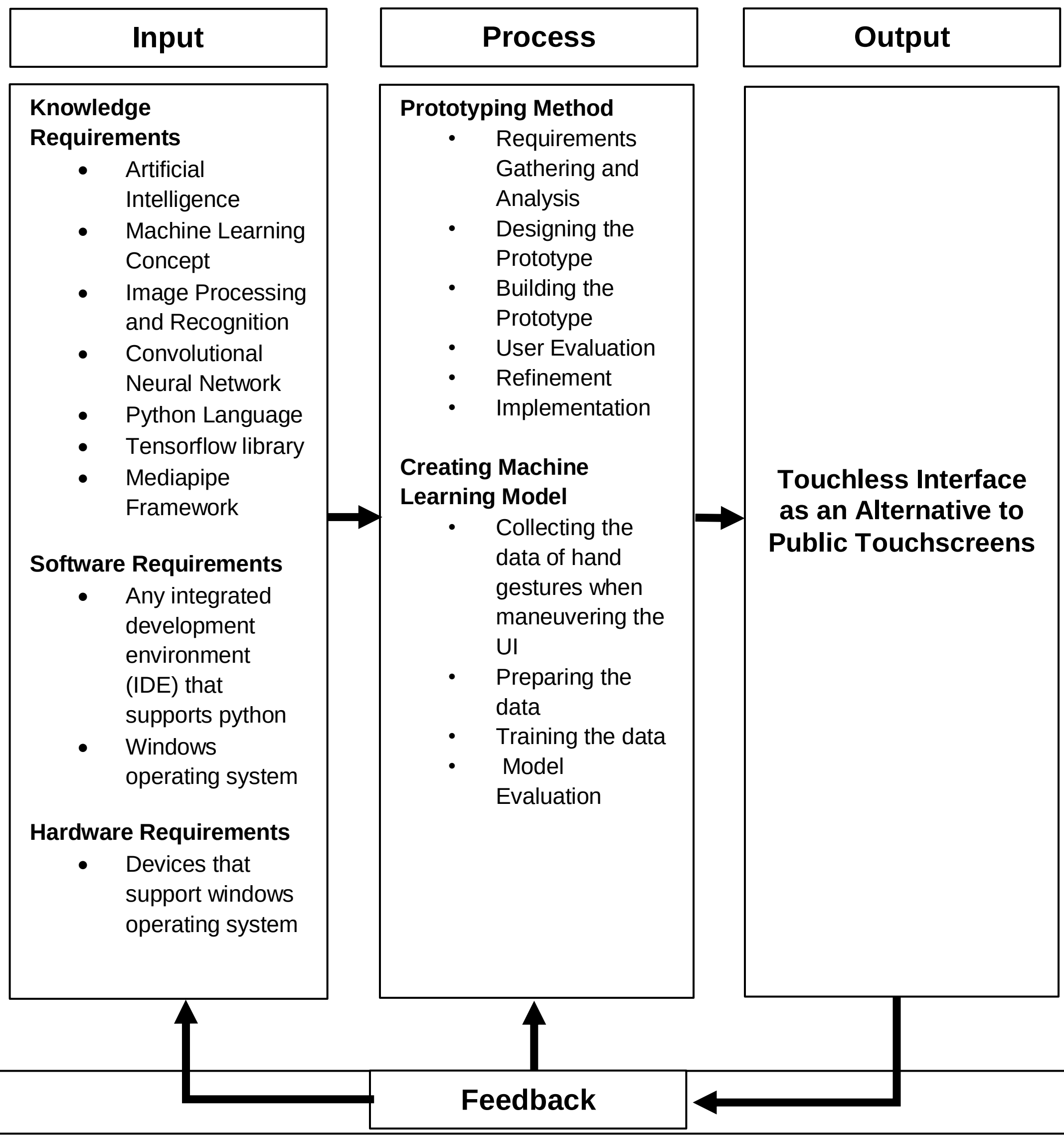
The researcher applied the research paradigm presented in Figure 1 for the development of touchless interface as an alternative to public touchscreens. In conceptualizing the whole process of this research, the researcher adopted the Input-Process-Output (IPO) template.

The important data which served as requirements for the development of the study were stated in the input block and were categorized as follows: (1) Knowledge Requirements which are the theories and ideas that served as pre-requisites for the further development of the study; (2) Software Requirements which are the computer programs or applications that would be used in the development of the system; and (3)

Hardware Requirements which include the list of necessary components and modules that would be integrated for the whole function of the system to be developed.

The steps to be followed in developing the study are presented in the process block of the research paradigm. These steps would lead to the title of this research which is “Touchless Interface as an Alternative to Public Touchscreens” as stated in the output block of the research paradigm. Also, a feedback block was included in the research paradigm which is interpreted as a closed system for the research paradigm of the study.

Figure 1. Research Paradigm of the Study



Statement of the Problem

This study aims to develop an application that will help substitute the public touchscreen used in malls, LRT, MRT, etc. to eliminate the need of direct contact with a touchscreen which would prevent surface spread of pathogens.

1. What are the stages in development of Touchless Interface as an alternative to touchscreens using developmental and prototyping methods?

2. What will the application use as the user's input in terms of:

   1.1 Hand Controlled Pointer; and,

   1.2 Hand Signs/Hand Gestures?

3. How effective will the application be in reading the user's input in terms of:

   2.1 Accuracy; and,

   2.2 Speed?

4. Which recognition system will be suited and accurate while being successful in interpreting:

   3.1 Real time image recognition; or

   3.2 Hand mesh approach?

Scope and Limitations of the Study

The Application is capable of recognizing hand-signs/gestures and hand-arm movement using a camera and reading them as input commands to a computer similar to a keyboard key press or a touchscreen tap. The program receives constant data from the camera in the form of a single frame and interprets if a person is in viewing range, then puts multiple nodes corresponding to specific parts of the person in view i.e., Shoulders, fingertips, and etc., the program will then use the position of the nodes as its data to which the AI model will interpret its actions using several frames from the camera.

The AI model is created through a different program associated with the program that will be added in the computer. It is created using python language and other dependencies/libraries. The data used for training the AI model is gathered by the researchers, the same way data is gathered in the main application by enacting the gestures in front of a camera, which is then fed to the AI to train it.

A Holistic model from Media pipe is used utilizing the pose and hand recognition which would find the torso, arm and hand of the body. Due to the AI created using data points in the Torso, arm and hands of the user, unfortunately the program would not be able to utilize some of its functions for users who have impairments mainly in the arm and hands.

Significance of the Study

According to an update by WHO on the Covid-19 pandemic, surfaces can hold on the virus for a significant amount of time ranging from 2 to 5 days depending on the material of the surface, and thus needs constant cleaning to keep the surface clean and lessen the spread of the virus. (World Health Organization, 2020)

Health Conscious. As the pandemic continues, a change of focus to health and disease prevention have risen and would likely continue even after the pandemic thus issuing technologies which will have this in mind in its development.

Concerned Agencies/Private groups/individuals. Others who have in mind the public's safety and concern who are aware of the risks of infection spreading through surfaces can utilize this technology for systems such as kiosk machines in malls, public services, food establishments, and others, lessening the need of constant cleaning since no surface are needed to touch which in turn increases the prevention of surface infections.

Future Researchers/Developers. This research will be available for other researchers to improve upon and for developers to use as a guide in the applications of machine learning.

Definition of Terms

Accuracy is the degree to which the result of a calculation, measurement, or specification conforms to the correct value or a standard.

Mediapipe is a framework for building multimodal (eg. video, audio, any time series data), cross platform (i.e Android, iOS, web, edge devices) applied ML pipelines. With MediaPipe, a perception pipeline can be built as a graph of modular components, including, for instance, inference models (e.g., TensorFlow, TFLite) and media processing functions. (Mediapipe.dev, n.d.)

Tensorflow is an end-to-end open-source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications. (Tensoflow, n.d.)

AI Model is a model created as an artificial intelligence in which it can predict through a selected list of answers which it was trained to recognize given the data required for it to predict.

Holostic Model is a feature of Mediapipe that integrates the face, hand, and pose models together to be used in different applications through mediapipe

Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. (Python (programming language), n.d.)

Nodes are points created by mediapipe to show the predicted position of different parts of the body in its face, hand, and pose models

Chapter 2

REVIEW OF LITERATURE AND STUDIES

This part of the research specifies the different studies and related works pertaining Touchless Interface as an Alternative to Public Touchscreen Technologies using machine learning algorithms, as well as image processing using open cv library.

Touch Screen Technology

In a systematic review of research into touchscreen across settings, populations, and implementations, the touchscreen interface is a combination of display and input device. The user's physical point of contact to the screen of the device is translated as an input to the interface and displays another graphical interface. Touchscreen interfaces was first commercialized in 1983 and after that, touchscreens have been implemented in a variety of uses and was installed in many devices and did a breakthrough in the world of computing. In the last decade, touchscreens can be found across industrial, commercial, and consumer applications, in devices as diverse as industrial controls, medical equipment, photocopiers, refrigerators, home thermostats, desktop computers, and, of course, smartphones. (Orphanides & Nam, 2017)

American Journal of Infection Control had an article about the Evaluation of an automated ultraviolet-C light disinfection device and patient hand hygiene for reduction of pathogen transfer from interactive touchscreen computer kiosks that touchscreens are a potential source of pathogen transmission. In their facility, patients and visitors rarely perform hand hygiene after using interactive touchscreen computer kiosks. The said computer kiosk have been contaminated by the transferring of viruses from the infected touchscreen and transferred to the fingertips. (Alhmidi, Cadnum, Piedrahita BS, John MD, & Donskey MD, 2018)

The upcoming paragraphs discusses the example uses and innovations using touchscreen technology. The innovation of robotic touchscreen totem for two-dimensional haptic force display states the Current touchscreen-based haptic systems that is used on contact friction to provide kinesthetic force feedback to the user. This paper presents the mechatronic design for a novel haptic interface which uses a steered wheel to provide kinesthetic force feedback on a large-format touchscreen. The goal is to display haptic constraints to a touchscreen user in the form of boundaries, area-of-effect fields, or paths. The device can exert up to 13.7 N of static friction force and penetrates a haptic wall by 6.0 mm in a head-on collision, while offering minimal resistance to motion along the wall. The magnitude of interaction force between the user and device is sensed with 11 mN resolution through displacement measurements of 3D-printed nylon flexures, designed to have uniform 7.0 N/mm stiffness in all directions parallel to the surface plane. (Price & Sup, 2016)

Touchscreen technologies have been widely adopted nowadays in mobile phone industries specifically smart phones with touchscreen capability that have become the focus of commercial competitions among manufacturers. The mobile phones sold in the market right now are mainly larger sized and gradually occupied the market of traditional keypad phones in the industry. (Zhu & Li, 2016)

Touchscreen input for cockpit flight displays in commercial aircraft offers advantages, including ease of use and reduced weight to aircraft manufacturers, airlines, and pilots. Commercial aircraft cockpits are replete with physical controls, including many forms of switches, knobs, levers, dials, keypads and wheels. While computer-based flight instrument displays are becoming increasingly prevalent (i.e., the 'glass cockpit'), these displays are almost exclusively used for data output, and user input to displayed objects is dependent on a separate indirect device. When a cursor is incorporated in the display, a trackball is typically used for item selection. (Cockburn, et al., 2017)

Mobile technologies, such as tablet devices, also open up new possibilities for diagnosis of health-related issues, monitoring, and intervention for older adults and healthcare practitioners. This is a study on administering cognitive test through touch screen tablet devices. Current evaluations of cognitive integrity typically occur within clinical settings, such as memory clinics, using pen and paper or computer-based tests. In the present study, we investigate the challenges associated with transferring such tests to touch-based, mobile technology platforms from an older adult perspective. (Jekins, Lindsay, Eslambolchilar, Thornton, & Tales, 2016)

Computer Vision

Computer vision process image acquired from the electronic camera, which is like a human vision system where the brain processes image that they get from the eye. Computer vision is rich and rewarding topic for the study and research for engineers, scientist and many others. Now that the computing technologies are more capable and are memory are vast, computer vision is found in many places. There are now many computer vision systems found in industrial use; cameras inspect mechanical parts to check parts, foods is inspected for quality and images are used in astronomy that they get from computer vision techniques. (Nixon & Aguado, 2019).

Image Processing

Images are an important medium which humans observe the majority of information they received from the real world. Image processing has been developed for many years. In general, image processing contains a number of techniques for treating images to obtain information from the image. With the latest technology, image processing techniques are expanding over wider and wider application areas. Because of the accumulation of solid research results and improvement of electronic technology, many new theories have been proposed, many techniques have been exploited, and many new applications have been created. (Zhang, 2021)

Machine Learning

Machine learning is a tool to tackle problems for which classical method are incapable of. Machine learning algorithms critically evaluate some problems that broadly aims to enable computers to “learn” without being directly programmed. It has origins in the artificial intelligence movement of the 1950s and emphasizes practical objectives and applications, particularly prediction and optimization. Computers “learn” in machine learning by improving their performance at tasks through “experience”. In practice, “experience” usually means fitting to data; hence, there is not a clear boundary between machine learning and statistical approaches. Indeed, whether a given methodology is considered “machine learning” or “statistical” often reflects its history as much as genuine differences, and many algorithms (e.g., least absolute shrinkage and selection operator (LASSO), stepwise regression) may or may not be considered machine learning depending on who you ask. (Bi, Goodman, Kaminsky, & Lessler, 2019)

Computational methods using machine learning algorithms can help improve the performance of a solution to the problem or to make accurate predictions. Term experience in machine learning refers to the past information available to the learners. It typically takes the form of electronic data collected and available for analysis. The data could be in the form of digitized human-labeled training sets, or other types of information obtained via interaction with the environment. The quality and size of the data are crucial to the success of the prediction made by the machine. (Mohri, Rostamizadeh, & Talwalkar, 2018)

There are a vast amount of problems that this algorithm can solve in this era advance technology. This are some uses of machine learning algorithm that can replace manual labor or make predictions.

In this study, we present a comprehensive review of the application of ML in agriculture. A number of relevant papers are presented that emphasize key and unique features of popular ML models. Typically, ML methodologies involves a learning process with the objective to learn from “experience” (training data) to perform a task. Yield prediction, one of the most significant topics in precision agriculture, is of high importance for yield mapping, yield estimation, matching of crop supply with demand, and crop management to increase productivity. Disease detection, one of the most significant concerns in agriculture is pest and disease control in open-air (arable farming) and greenhouse conditions. The most widely used practice in pest and disease control is to uniformly spray pesticides over the cropping area. Weed detection and management is another significant problem in agriculture. Many producers indicate weeds as the most important threat to crop production. The accurate detection of weeds is of high importance to sustainable agriculture, because weeds are difficult to detect and discriminate from crops. Again, ML algorithms in conjunction with sensors can lead to accurate detection and discrimination of weeds with low cost and with no environmental issues and side effects. ML for weed detection can enable the development of tools and robots to destroy weeds, which minimize the need for herbicides. (Liakos, Busato, Moshou, Pearson, & Bochtis, 2018)

There are many opportunities for further breakthroughs in machine learning to provide even greater advances in the automated design and discovery of molecules and materials. The field of computational chemistry has become increasingly predictive in the twenty first century, with activity in applications as wide ranging as catalyst development for greenhouse gas conversion, materials discovery for energy harvesting and storage, and computer-assisted drug design. With machine learning, given enough data and a rule discovery algorithm, a computer has the ability to determine all known physical laws (and potentially those that are currently unknown) without human input. In traditional computational approaches, the computer is little more than a calculator, employing a hard-coded algorithm provided by a human expert. The standard description of chemical reactions, in terms of composition, structure and properties, has been optimized for human learning. Most machine-learning approaches for chemical reactions or properties use molecular or atomic descriptors to build models, the success of which is determined by the validity and relevance of these descriptors. A good descriptor must be simpler to obtain than the target property and of as low dimensionality as possible. Automatic discovery of scientific laws and principles by inspection of the weights of trained machine-learning systems is a potentially transformational development in science. As scientists embrace the inclusion of machine learning with statistically driven design in their research programs, the number of reported applications is growing at an extraordinary rate. This new generation of computational science, supported by a platform of open-source tools and data sharing, has the potential to revolutionize molecular and materials discovery. (Butler, Davies, Cartwright, Isayev, & Walsh, 2018)

In the process that identifies geologic structures such as faults, salt domes or other petroleum element systems in general, this kind of structural interpretation heavily depends on the vast knowledge of the interpreters as well as the geologic structures such as texture and geometry. By treating seismic data as images rather than signal traces, researchers have been able to utilize advanced image-processing and machine-learning techniques to assist `interpretation directly. In this study, the proponents mainly focus on the interpretation of two important geologic structures, faults and salt domes, and summarize interpretation workflows based on typical or advanced image-processing and machine-learning algorithms. In recent years, increasing computational power and the massive amount of available data have led to the rise of deep learning. The convolutional neural network, a form of deep-learning model that is effective in analyzing visual imagery has been applied in fault and salt dome interpretation. (Wang, Di, Shatif, Aluadah, & AlRegib, 2018)

Dermatological diseases are the most prevalent diseases worldwide. Despite being common, its diagnosis is extremely difficult and requires extensive experience in the domain. In this research paper, we provide an approach to detect various kinds of these diseases. We use a dual stage approach which effectively combines Computer Vision and Machine Learning on clinically evaluated histopathological attributes to accurately identify the disease. In the first stage, the image of the skin disease is subject to various kinds of pre-processing techniques followed by feature extraction. The second stage involves the use of Machine learning algorithms to identify diseases based on the histopathological attributes observed on analyzing of the skin. Upon training and testing for the six diseases, the system produced an accuracy of up to 95 percent. (Kumar, Kumar, & Saboo, 2016)

Tomato a nutritious and nourishment fruit is one of the top-grown agricultural produce in the world. With large-scale production and the need for high-quality tomatoes to meet consumer and market standards criteria, have led to the need for an inline, accurate, reliable grading system during the post-harvest process. This study introduced a tomato grading machine vision system based on RGB images. The proposed system performed calyx and stalk scar detection at an average accuracy of 0.9515 for both defected and healthy tomatoes by histogram thresholding based on the mean g-r value of these regions of interest. Defected regions were detected by an RBF-SVM classifier using the LAB color-space pixel values. The model achieved an overall accuracy of 0.989 upon validation. Four grading categories recognition models were developed based on color and texture features. The RBF-SVM outperformed all the explored models with the highest accuracy of 0.9709 for healthy and defected category. However, the grading accuracy decreased as the number of grading categories increased. A combination of color and texture features achieved the highest accuracy in all the grading categories in image features evaluation. (Ireri, Belal, Okinda, Makange, & Ji, 2019)

Skin diseases are more common than other diseases. Skin diseases may be caused by fungal infection, bacteria, allergy, or viruses, etc. The advancement of lasers and Photonics based medical technology has made it possible to diagnose the skin diseases much more quickly and accurately. But the cost of such diagnosis is still limited and very expensive. So, image processing techniques help to build automated screening system for dermatology at an initial stage. The extraction of features plays a key role in helping to classify skin diseases. Computer vision has a role in the detection of skin diseases in a variety of techniques. Due to deserts and hot weather, skin diseases are common in Saudi Arabia. This work contributes in the research of skin disease detection. We proposed an image processing-based method to detect skin diseases. This method takes the digital image of disease effect skin area, then use image analysis to identify the type of disease. Our proposed approach is simple, fast and does not require expensive equipment other than a camera and a computer. The approach works on the inputs of a color image. Then resize the image to extract features using pre-trained convolutional neural network. The system successfully detects 3 different types of skin diseases with an accuracy rate of 100%. (Soliman & ALEnezi, 2019)

Convolutional Neural Network

According to the book “Recent Trends and Advances in Artificial Intelligence and Internet of Things” Convolutional neural network (CNN) is a special type of multilayer network architecture which was inspired by how living beings see. The CNN is versatile in multiple fields of computer vision and language processing.

Convolutional Neural Network (CNN), also called ConvNet, is a type of Artificial Neural Net-work (ANN), which has deep feed-forward architecture and has amazing generalizing ability as compared to other networks with FC layers, it can learn highly abstracted features of objects especially spatial data and can identify them more efficiently. A deep CNN model consists of a finite set of processing layers that can learn various features of input data (e.g. image) with multiple level of abstraction. The initiatory layers learn and extract the high level features (with lower abstraction), and the deeper layers learns and ex-tracts the low level features (with higher abstraction). The basic conceptual model of Canadas shown in figure 2, different types of layers described in subsequent sections.

CNN has a deep feed-forward architecture, having great generalizing capabilities when compared to other networks, it has the capability to learn highly abstracted features of objects and identify them in an efficient manner. (Ghosh, Sufian, Sultana, Chakrabarti, & De, 2020)

In recent years, computer vision which is one of the fastest growing artificial intelligence disciplines, has become increasingly important in our society due to its wide range applications in different areas such as health care and medicine (algorithms that can diagnose medical images for diseases), vision-based robotics, self-driving cars (that can see and drive safely). Convolutional neural networks are biologically inspired architectures and represent the core of deep learning algorithms in computer vision. In this paper, we represent the fundamental building blocks of convolutional neural networks and the most popular convolutional neural network architectures in the history, including those that have achieved the state-of-the-art performance on standard recognition datasets and tasks such as ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). (Bezdan & Bacanin, 2019)

This is a note that describes how a Convolutional Neural Network (CNN) operates from a mathematical perspective. This note is self-contained, and the focus is to make it comprehensible to beginners in the CNN field. The Convolutional Neural

Network (CNN) has shown excellent performance in many computer vision and machine learning problems. Many solid papers have been published on this topic, and quite some high quality open source CNN software packages have been made available. There are also well-written CNN tutorials or CNN software manuals. However, I believe that an introductory CNN material specifically prepared for beginners is still needed. Research papers are usually very terse and lack details. It might be difficult for beginners to read such papers. A tutorial targeting experienced researchers may not cover all the necessary details to understand how a CNN runs. This note tries to present a document that is self-contained. It is expected that all required mathematical background knowledge are introduced in this note itself (or in other notes for this course) has details for all the derivations. This note tries to explain all the necessary math in details. We try not to ignore an important step in a derivation. Thus, it should be possible for a beginner to follow (although an expert may feel this note tautological) ignores implementation details. The purpose is for a reader to understand how a CNN runs at the mathematical level. We will ignore those implementation details. In CNN, making correct choices for various implementation details is one of the keys to its high accuracy (that is, “the devil is in the details”). However, we intentionally left this part out in order for the reader to focus on the mathematics. After understanding the mathematical principles and details, it is more advantageous to learn these implementation and design details with hands-on experience by playing with CNN programming. CNN is useful in a lot of applications, especially in image related tasks. Applications of CNN include image classification, image semantic segmentation, object detection in images, etc. We will focus on image classification (or categorization) in this note. In image categorization, every image has a major object which occupies a large portion of the image. An image is classified into one of the classes based on the identity of its main object, e.g., dog, airplane, bird, etc. (Wu, 2017)

Sequential CNN

Since each layer connects only to the previous and following layers, we refer to this conventional CNN as a “sequential CNN model”. This model is trained from scratch using the spine images. The proposed model encompasses five convolutional and three fully-connected layers. Input images of dimension 227 × 227 × 3 are fed into the input layer. There are 96 filters in the first convolutional layer, each of dimension 7×7 and stride 2. A Rectified Linear unit (ReLU) activation follows each convolutional layer to enhance learning. All the other convolutional layers have filters of dimension 3×3. Weights are initialized from a zero-mean Gaussian distribution. A local response normalization (LRN) layer is included after the first and second convolutional layers to aid in generalization, motivated by the lateral inhibition process of biological neural networks. Max-pooling layers with a pooling window of 3×3 and stride 2 follow the LRN layers and the fifth convolutional layer. There are three fully-connected layers, the first two fully connected layers having 4096 neurons each; the third fully connected neurons has two neurons which feed into the Softmax classifier. Dropout regularization is achieved by dropping 50% of the neurons in the first and second fully-connected layers during the process of training, to alleviate over-fitting issues. The proposed model is trained by optimizing the multinomial logistic regression objective using stochastic gradient descent (SGD) with momentum. The model is optimized for its hyper-parameters by a randomized grid search method. L2-regularization is used with a weight penalty of 5×10-4. The learning rate is initialized to 0.001 and is reduced three times before convergence. The mini-batch size is 10 and the training is stopped after 60 epochs. The proposed model achieves faster convergence due to implicit regularization imposed by smaller convolutional filter dimensions, greater depth, usage of L2-regularization parameter, and dropouts in the fully-connected layers.

Long-Short Term Memory (LSTM)

In the past years, traditional pattern recognition methods have made great progress. However, these methods rely heavily on manual feature extraction, which may hinder the generalization model performance. With the increasing popularity and success of deep learning methods, using these techniques to recognize human actions in mobile and wearable computing scenarios has attracted widespread attention. In this paper, a deep neural network that combines convolutional layers with long short-term memory (LSTM) was proposed. This model could extract activity features automatically and classify them with a few model parameters. LSTM is a variant of the recurrent neural network (RNN), which is more suitable for processing temporal sequences. In the proposed architecture, the raw data collected by mobile sensors was fed into a two-layer LSTM followed by convolutional layers. In addition, a global average pooling layer (GAP) was applied to replace the fully connected layer after convolution for reducing model parameters. Moreover, a batch normalization layer (BN) was added after the GAP layer to speed up the convergence, and obvious results were achieved. The model performance was evaluated on three public datasets (UCI, WISDM, and OPPORTUNITY). Finally, the overall accuracy of the model in the UCI-HAR dataset is 95.78%, in the WISDM dataset is 95.85%, and in the OPPORTUNITY dataset is 92.63%. The results show that the proposed model has higher robustness and better activity detection capability than some of the reported results. It can not only adaptively extract activity features, but also has fewer parameters and higher accuracy.

Human activity recognition (HAR) plays an important role in people's daily lives because it has the ability to learn profound advanced knowledge about human activities from raw sensor data. People could automatically classify the type of human motion and obtain the information that the human body needs to convey by extracting features from daily activities, which in turn provides a basis for other intelligent applications. Hitherto, this technology has been widely used in the fields of home behavior analysis, video surveillance, gait analysis, and gesture recognition, etc. (Xia, Huang, & Wang, 2020)

Synthesis of the Reviewed Literature and Studies

The related literature and studies found in Chapter 2 were used as a foundation in order to develop this research. These also served as the backbone of the study for making Touchless Interface as an Alternative to Public Touchscreens become feasible by exploiting the latest technology of computer vision and applying the different techniques that are associated with image processing. The advancement of technology and newly proposed theories in computer algorithms had been helpful especially in machine learning for data analysis that are used in building the researchers' system. With the use of data the proponents' got after processing the different gestures of hand images when using the interface, the researchers took advantage convolutional neural networks which is a machine learning algorithm in order to analyze the visual image data of the hand gestures like swiping or pressing a button in the interface.

Chapter 3

METHODOLOGY

This chapter explains the different research methods used in gathering and analyzing the data, and the processes involved in development of the device and answering the research problems.

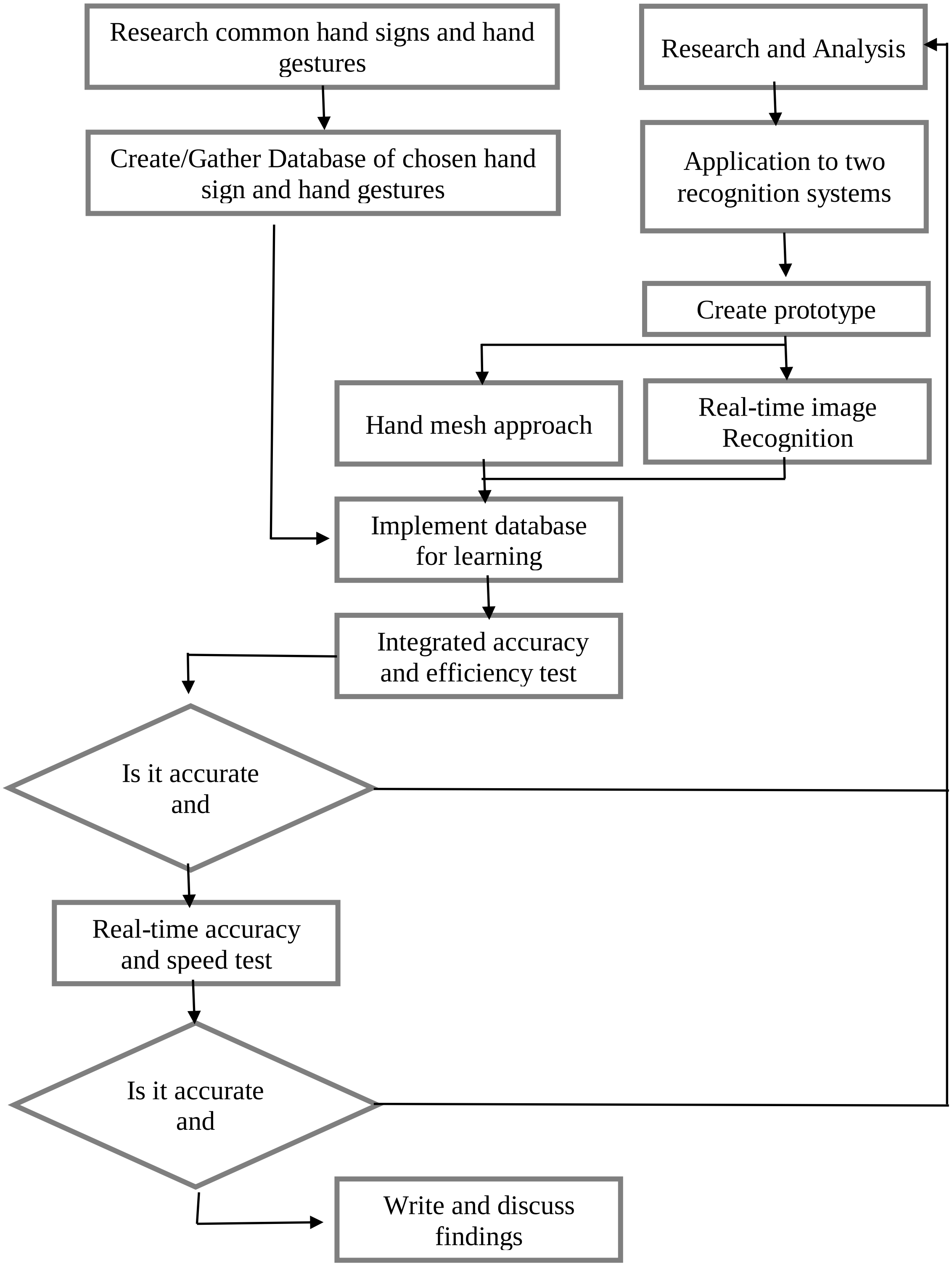
Research Design

The proposed project is used as a background application or is interlaced with another application that uses touch-based inputs like those in touch screens and acts as a substitute for getting inputs. A developmental method will be used to tackle this research. Prototyping method will be used to assess problem number 1 and 3, as this will give a more tangible answer. An experimental method will provide sufficient data for problem number 1 and 2, as tests both through the learning accuracy and the actual real-time accuracy will be implemented.

Flowchart of Research Design/Process Flowchart

Flowchart is a diagram that shows the process or flow of a system. They are widely used in multiple fields which converts complex processes to an easy-to-understand diagram. Flowchart uses multiple shapes to define steps in conjunction with arrows to show the flow or sequence of the process. Figure 2 below depicts the flow of the process which the researchers will follow in the creating, prototyping, and experimenting of the system.

Figure 2. Research Design/Process Flowchart



Description of Research Instrument Used

Experiments will be used as our research instrument. Two experiments are conducted, Tensorflow's categorical accuracy and loss test will be the initial test and real-time prototype testing will be the second test. Tensorflow's categorical accuracy and loss test will be conducted on the model training which will show the accuracy and loss of the model during its training over epochs. The real-time prototype testing will involve the model created that passed the initial test, the model will then be tested by the researchers using recorded videos and/or actual demonstrations of the hand signs and hand gestures)

Statistical Treatment

Problem number 2 require the computation of accuracy, this is achieved by using Tensorflow. The accuracy and loss are calculated by Tensorflow using the following formulas:

$$Accuracy = \frac{Number\ of\ Test\ Values}{Number\ of\ correct\ Prediction\ Values}$$

$$Loss = 1 - Accuracy$$

The test values are determined by Tensorflow through random selection of array data corresponding to the label it is associated with, while the prediction values are randomly selected from the array data which are not chosen as test values. Loss is the difference of accuracy from 1 or 100%. The number of test values are the number of individual values within the array, and the number of correctly predicted values are the total number of individual values in the prediction values which are equal to the individual values in the test values with the same placeholder in their respective array. Both test values and prediction values have the same array size.

A second test will be conducted and this test is a demonstration test which would involve the researchers testing the accuracy of the machine with the action we act out. The accuracy and speed are calculated using the following formulas:

$$Accuracy = \frac{Number\ of\ Tests\ Presented}{Number\ of\ correct\ initial\ Predictions}$$

$$Speed = Time\ Until\ correct\ Prediction - Start\ time\ of\ Action$$

This test could be done in real time where the researchers themselves act out to test the model or the researchers could capture videos of themselves or from others doing the action to test the mode.

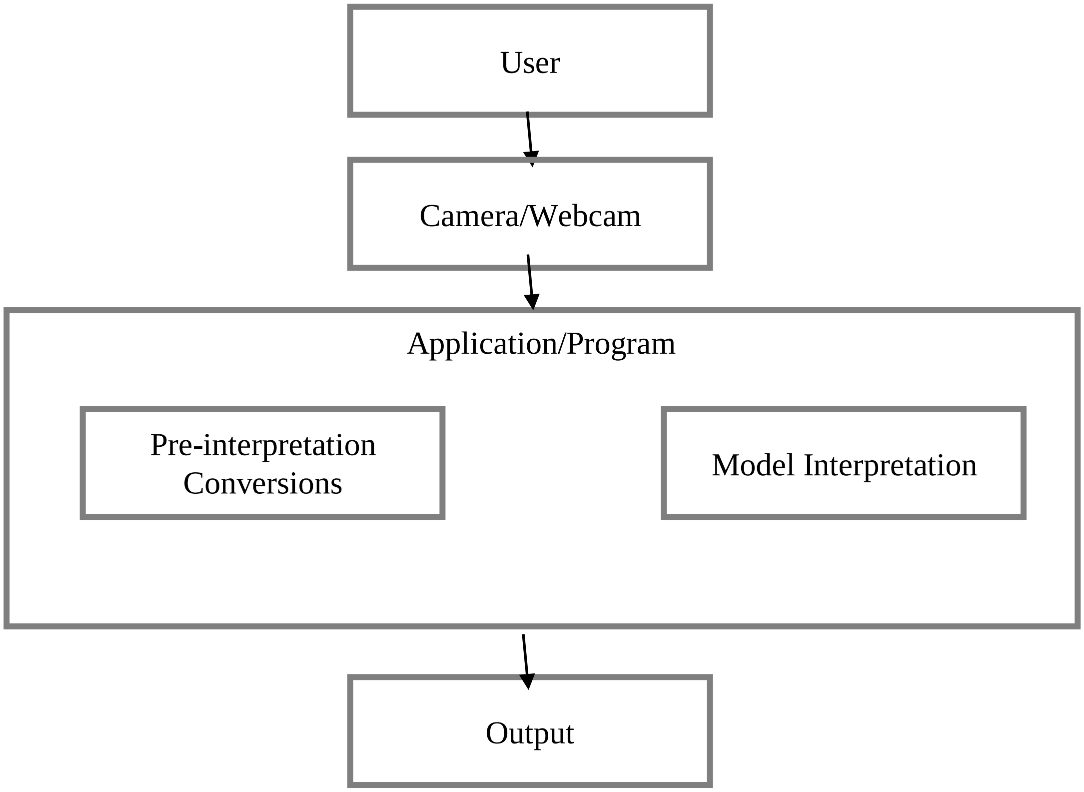
Design Project Flow

The touchless interface as an alternative to public touchscreen project will operate in accordance with the following project designs.

System Architecture

System Architecture is a conceptual model, it shows the structure, behavior, etc. of a system. This simplifies in depicting the system by utilizing diagrams and arrows. Figure 3 shows the system architecture of the application and how data is transferred from each part of the system to another.

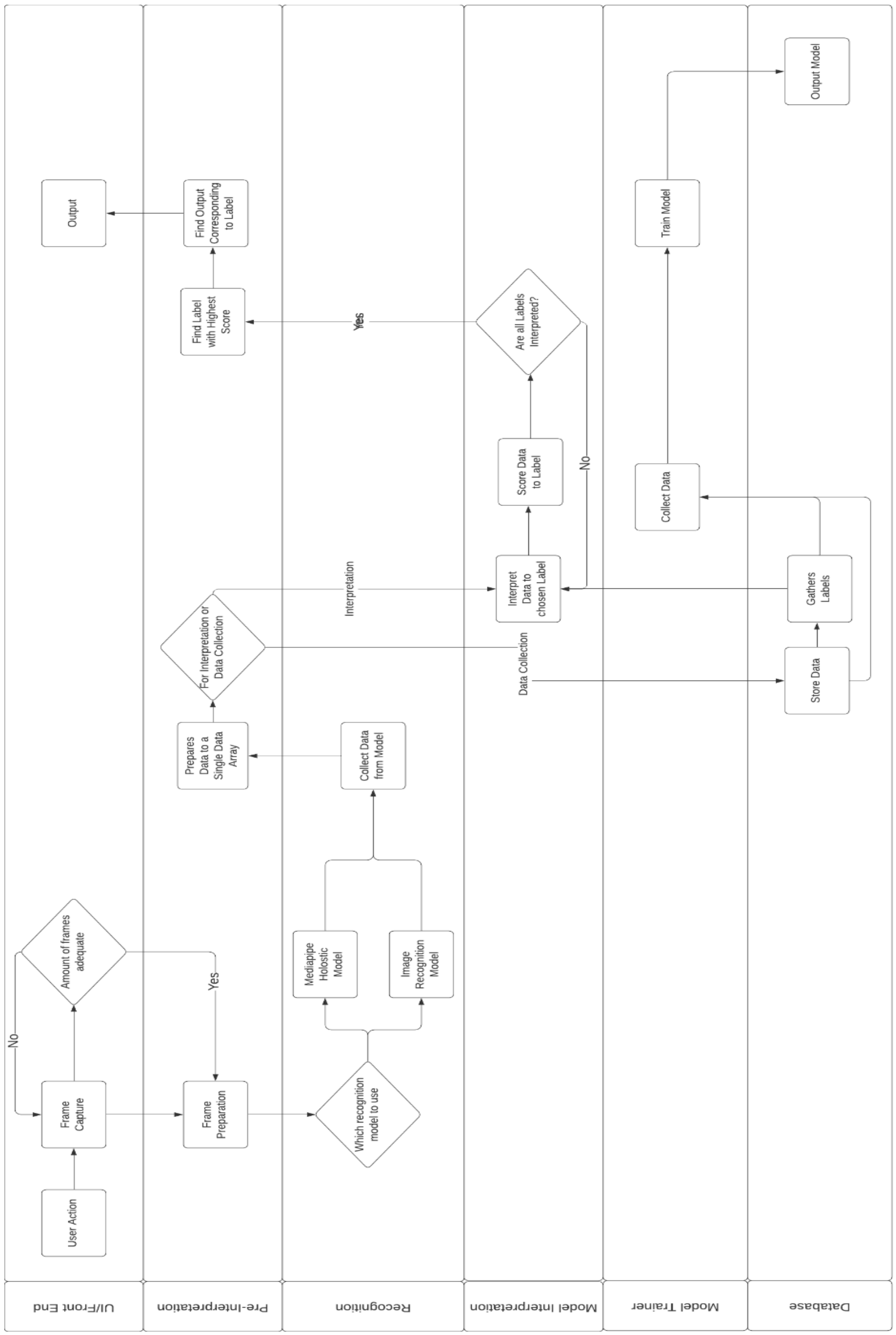
Figure 3. System Architecture



System Flow

System Flow are models that show the activities and decisions that each system executes. System Flow are useful in understanding complex interactions by showing it visually. Figure 4 shows the interactions of different systems in predicting the action of the user, adding data in the database, and creating the AI model for interpretation.

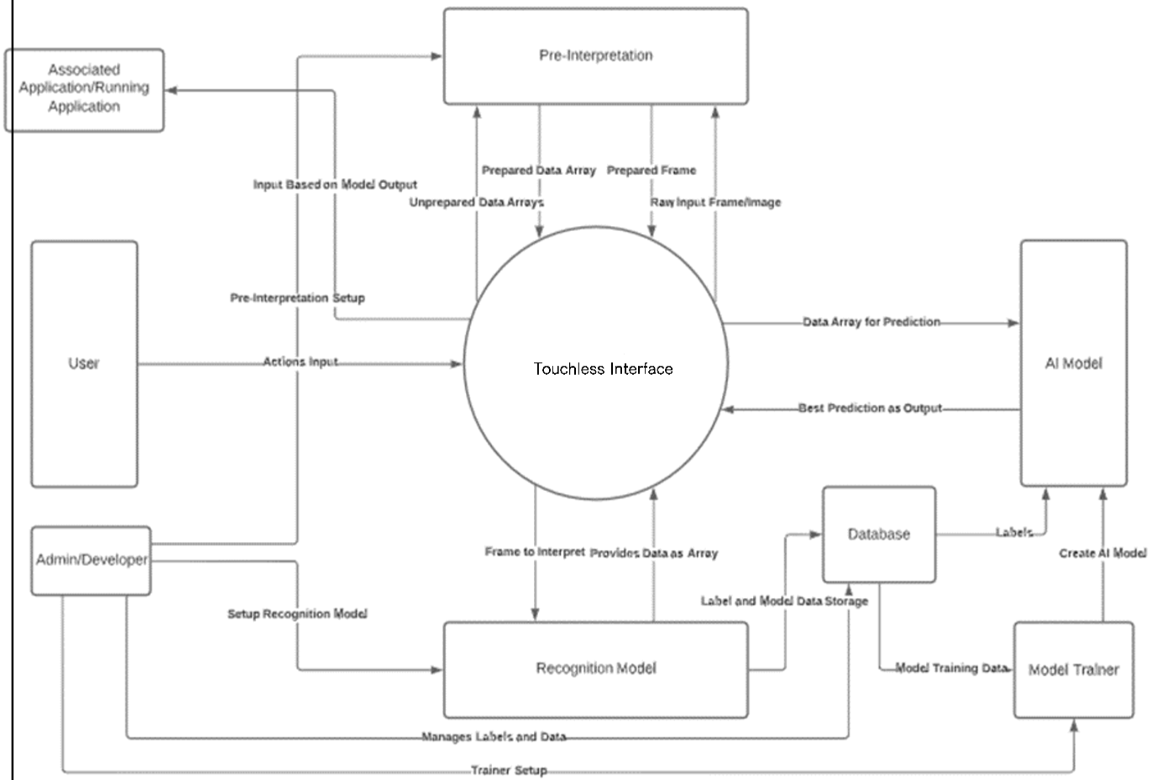
Figure 4. System Flow



Context Diagram with Data Flow Diagram

Context Diagram is a basic overview of the whole system, it is designed to be an at-a-glance view, which shows the system as a single process and its relationships with other entities. Figure 5 shows the process, relationships, and functions of the system and its entities that form the whole touchless interface system.

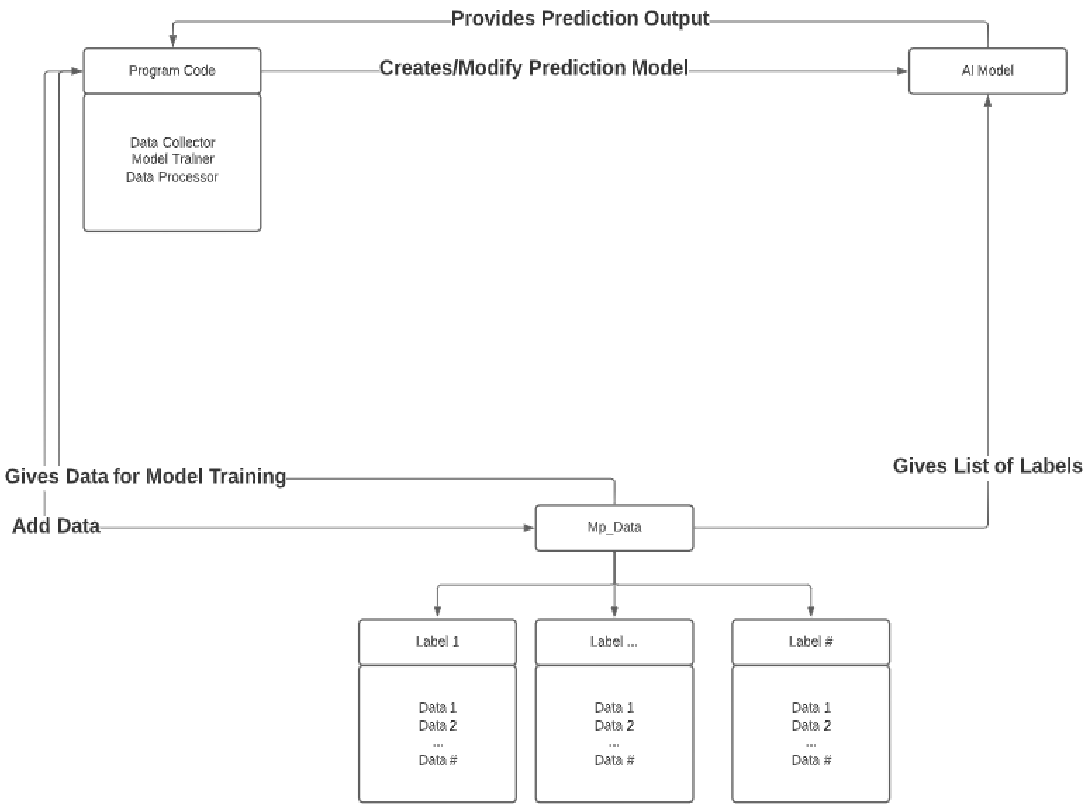
Figure 5. Context Diagram with Data Flow Diagram



Entity Relationship Diagram

ER-Diagram or Entity-Relationship Diagram displays the relationship of entities in the database. Figure 6 below shows the connections and actions of each entity in the database for the application.

Figure 6. Entity Relationship Diagram



Chapter 4

RESULTS AND DISCUSSION

This chapter includes the presentation, analysis and interpretation of the data gathered. The sequence is based on the specific research problems as stated in Chapter 1. Necessary figures and tables are included to present the topics discussions in detail and answer the problems stated.

1. What are the stages in development of Touchless Interface as an alternative to touchscreens using developmental and prototyping methods?

Prototyping method was utilized in relevance to the development of this system. To increase the viability of the study, the method begins by gathering all of the prerequisites, including timely publications and journals. Making a viable design is the next step in the prototyping process. The prototype was quickly designed, paving the way for the following stage, which is building the system. The following documents that were utilized to create the first prototype that supports the desired functions were gathered. Once faults or malfunctions have been identified, the prototype is improved further to eliminate them in the refinement process. The process continues until the user is satisfy with the results and approves the prototype for the community to utilize it.

The steps of the prototyping method are shown in Figure 7.

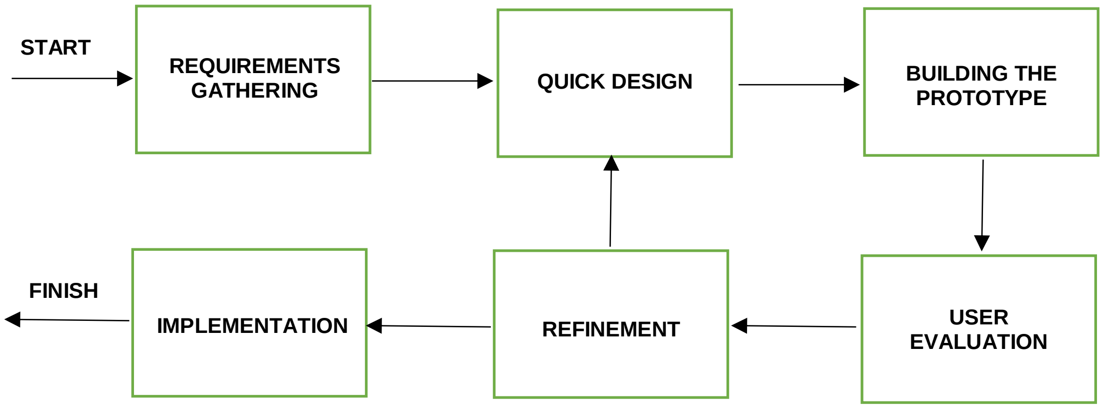


Figure 7. Stages in the Development of the System Using Prototyping Method

Gathering and analyzing requirements Gathering and assessing the necessary components is the first step in the prototyping process. The information was obtained and analyzed through papers and interviews. The specifications, functionality, and availability are the basis of evaluating the plan in building the system.

Quick Design. The initial design will be shown prior to the project's actual construction. The design of the touchless interface integration for the ticket vending machines will be built, finalized, and the appropriate materials will be chosen to fully utilize its effectiveness. This involved the creation of a sample simulation in a computer for testing. And the materials for main testing and transferring of the software to the ticket vending machines were chosen with the assistance of specification documents and professional guidance.

Being able to maneuver the cursor and click the desired button without touching is needed by the commuters in the LRT and MRT station in order to avoid contact to the surface if the said ticket vending machine are infected with unwanted virus that may cause harm to the body and endanger one's life.

The camera utilized in the system are just normal web camera that is used in a typical desktop computer and not particularly unique nor expensive. The main backbone of the system is the machine learning algorithm written in python. The web cam is used to capture every frame of the user's hand movements while hovering through the interface.

The first testing of the system would be in a computer simulation before applying the actual software in the ticket vending machines. The computer that would be used in this project has an AMD Ryzen 5 3400G CPU with Radeon Graphics (8 CPUs) that has a base clock speed of 3.7GHz. It also has a 16GB of RAM and runs in Windows 10 Pro 64-bit operating system. Visual Studio is the Integrated Development Environment (IDE) that would be used to code the python algorithm. The researchers also build a simple ticket vending machine software application that processes the purchases done by the commuter.

To simulate the system effectively, the researchers imitated the User Interface (UI) of the LRT/MRT kiosk by building a prototype that emulate the main functions of the said ticket vending machines. The program is built in Unity which is a cross-platform

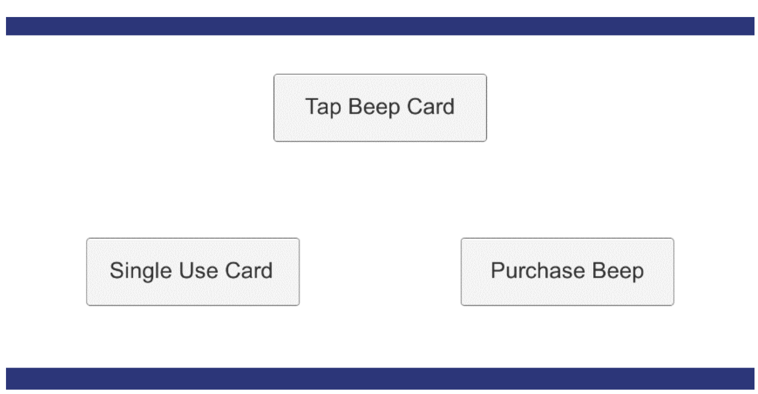


Figure 8. Main Interface Made in Unity

game engine with a built-in IDE and used to develop video games for desktop platforms, consoles and mobile devices.

The main interface of the kiosk prototype consists of three buttons redirecting to each of the features that can be seen in an actual kiosk in LRT/MRT stations.

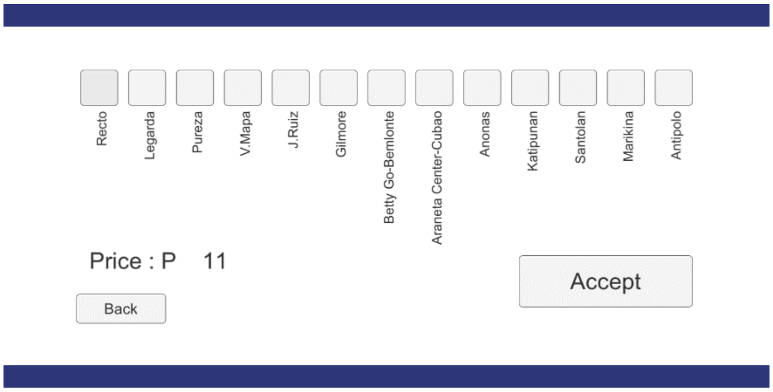


Figure 9. Single Used Card Feature

The single use card functions as a regular ticket vending machine that dispenses on way ticket to a specific station after purchasing it. The interface features the LRT 2 different stations that the commuter can go to starting from Recto to Antipolo. The prize of each station can also be seen in the interface by tapping the square icon beside it. After the person made a decision, he/she can push the accept button to

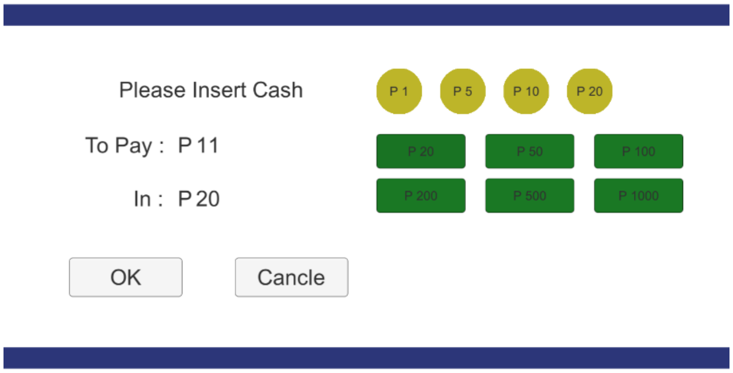


Figure 10. Proceed Purchase

process the purchase or press the back button if he/she wants to go back to the main interface.

The beep card button is used when loading a pre-owned beep card. The user can load any amount he/she desired by selecting different combination of amounts that can be seen within the interface. After deciding, the user can tap the ok button to process the payment or press the cancel button to cancel the purchase.

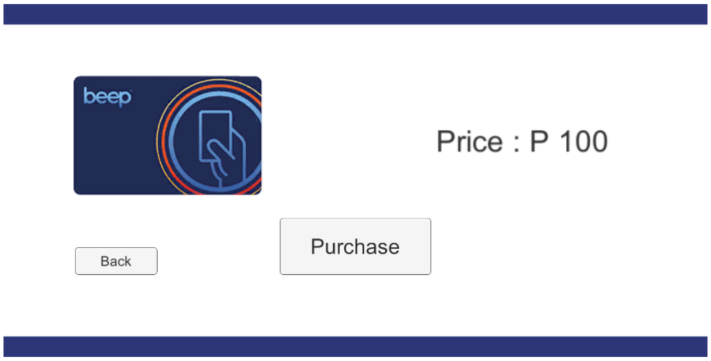


Figure 11. Tap Beep Card Feature

The purchase beep button enables the commuter to buy a personal beep card that can be loaded and be used for a long term. After deciding, the user can tap the purchase button to process the payment or press the back button to cancel the purchase.

Building the Prototype. The actual prototype will be built and developed using the analyzed data that were acquired during the design process. To establish the system's efficiency and viability, certain parameters will be examined.

Figure 12. Four Frames of Different Hand Shapes Captured by the Camera

Figure 12 shows four different images highlighting the changing of hand signs that are defected and captured by the web camera. From the figure, red dots and white lines can be seen in front of the hand. This indicates that the combination of opencv and mediapipe library used to code the system are able to detect the front and back of the hand at every angle and shape. The red dots represent the joints located in the hand and the white lines connects every joint to form the skeleton like figure of the hand. Counting from left to right while prioritizing top to bottom, the first image reveals the palm of the hand. The second image detects the hand imitating the point and push of a button gesture. The third image highlights the back of the hand counting to three while the fourth and last one shows a rock n roll hand sign.

User Evaluation. The technology will be tested when it has been built. The device's functionality is the most important feature that must be presented in order for the evaluators to thoroughly examine the prototype. The assessors will evaluate the application to see whether it can be improved further.

Refinement. All of the evaluators' comments and suggestions were documented after the prototype presentation. Changes to the system and features were implemented based on the evaluators' recommendations. Applying adjustments to the prototype's schematic and upgrading the system's additional features were completed in the previous phase.

Implementation. The prototype model is now in its final phase. During this phase, the product was reviewed to see if it met the municipality's needs. This is a functioning application that has previously been implemented, but it is being maintained and progressed to avoid failures.

The system's software would be install in the ticket vending machines located at every station of LRT and MRT as shown in figure 13.

Figure 13. Photos of the Ticket Vending Machines in LRT 2

2. What will the application use as the user's input in terms of:

1.1 Hand Controlled Pointer

The combination of Opencv and MediaPipe Holistic are able to help the system recognize the different inputs of the user in terms of hand controlled pointer and hand signs/hand gestures. Opencv was used to provide the standard framework for computer vision of this application and accelerate the perception when used in machine learning. The MediaPipe library offers accurate and fast real-time perception of simultaneous hand tracking and human pose estimation and enable the different inputs to be detected by locking each keypoint to its corresponding joint located in the hand, arms, torso and the lower part of the body. These keypoints can be used to code different instructions and execute based on the instruction given to it. The researchers utilized MediaPipe Hands and MediaPipe pose to get the accurate input of the user.

MediaPipe pose is model with high-fidelity body pose tracking. It is one of the machine learning solution used that has 33 3D landmarks but the researchers only used the 22 keypoints the can be seen in the arms and chest leaving the 11 keypoints starting from the left and right hip to the index of the left and right foot since these body parts are the only thing that moves when the user maneuver the cursor in the interface of the system. The researchers also took advantage of the background segmentation mask on the whole body in order to precisely render specific joint of the body in different backgrounds.

MediaPipe hands is a vital component to improve the user experience across the touchless interface system. This machine learning solution has the ability to recognize the shape and movement of hands by applying keypoints to each joint located in it. This keypoints which are like the skeletal system of the hand allows the AI to track every motion that are seen within the perspective of the camera. Since there are a total of 21 joint in one hand, this high-fidelity hand and finger tracking solution consist of 21 3D landmark which are the keypoints captured from a single frame. This hand perception functionality is used as the main element for the hand controlled pointer of the touchless interface system.

The following lists of figures contain the different hand gestures that the system can recognize to execute a certain instruction in the user interface.

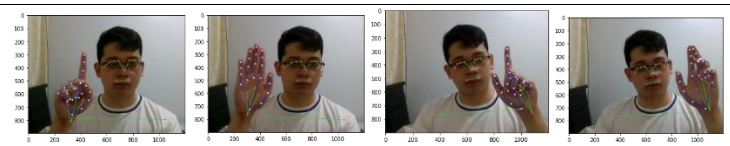
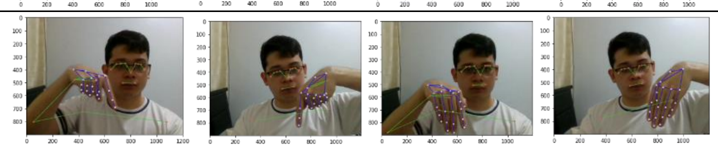
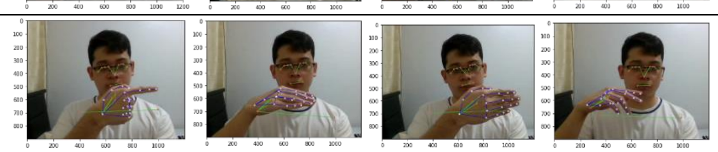
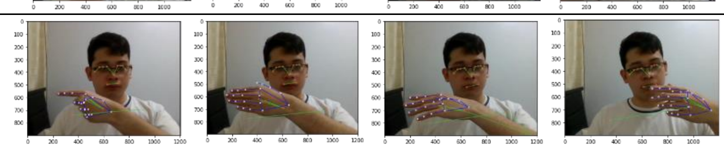
Hand Controlled Pointer	Different Frames of Hand Gestures Captured
Point Up	
Point Down	
Point Left	
Point Right	

Table 1. Different Hand Shapes with Its Corresponding Function

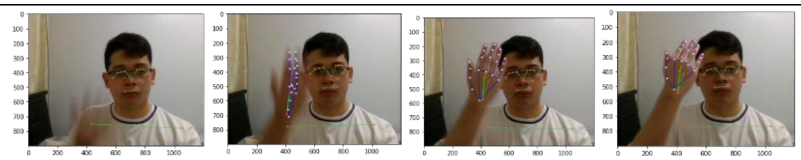
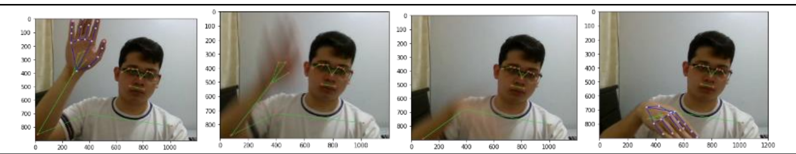
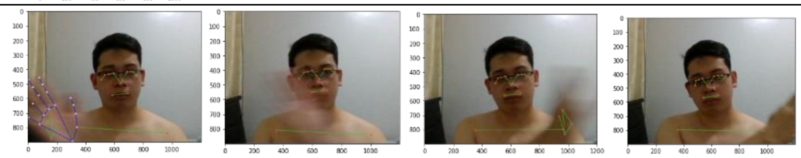
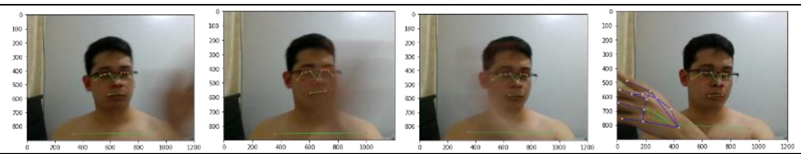
Hand Signs/Hand Gestures	**Hand Movement Patterns**
Swipe Up	
Swipe Down	
Swipe Left	
Swipe Right	

Figure 2. Different Hand Movement Pattern with Its Corresponding Function

3. How effective will the application be in reading the user's input in terms of:

1.1 Accuracy

There are four metrics that are initially used in order to measure the accuracy of the system. The confusion matrix is an object detection metric that is used to visually observe how well the machine learning model predicts on the testing data. There are 250 data that were tested within the model and the researchers used sklearn library to calculate different parameters that are crucial in determining the accuracy of the system. The other three metrics which are the recall, precision, and f1-score are calculated after getting the result from the confusion matrix. The matplotlib is a handy tool or library in python to visually plot the table of the confusion matrix.

Hand Pointers

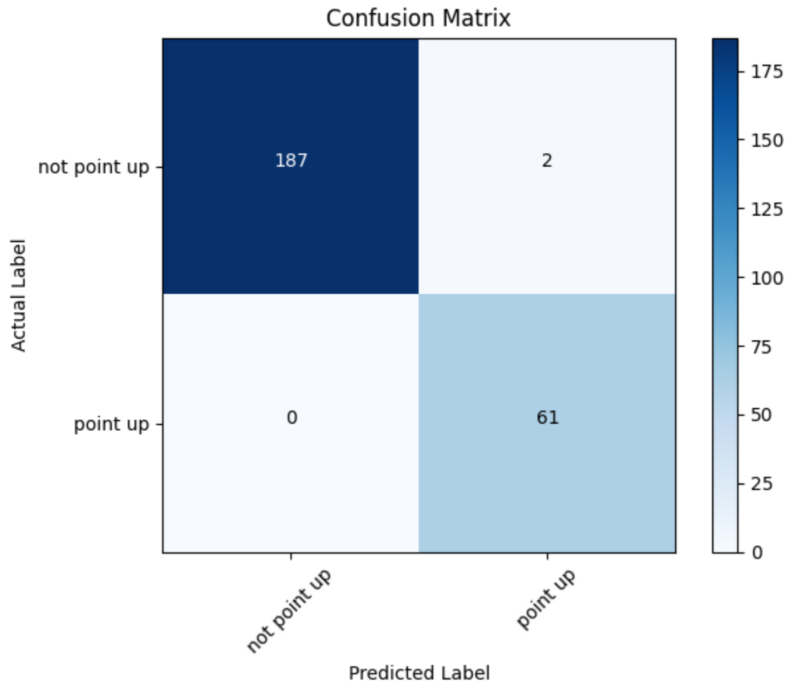


Figure 14. Point Up

The confusion matrix of point up pointer are made up of four boxes containing a specific value. The box in the upper left is the true negative (TN) which the 188 value is the number of times when the actual label is not point up and the model predicted it successfully. In the lower right corner is the true positive (TP) which means that the model predicted point up 61 times where the actual label is also point up. In the upper right where the box contains the value of two is called the false negative (FN) and this tells us that the actual label is not point up but the model thought it was point up which is wrong. In the lower left, there is zero value since the model did not have any mistake predicting the point up for its opposite label and this is also called the false positive (FP). The whole point of the confusion matrix tells us that the model is well in predicting the point up hand pointer since it only got two mistakes for predicting point up where the actual label is not point up. The model were able to distinguish all the point up gesture out of the 250 sample data given.

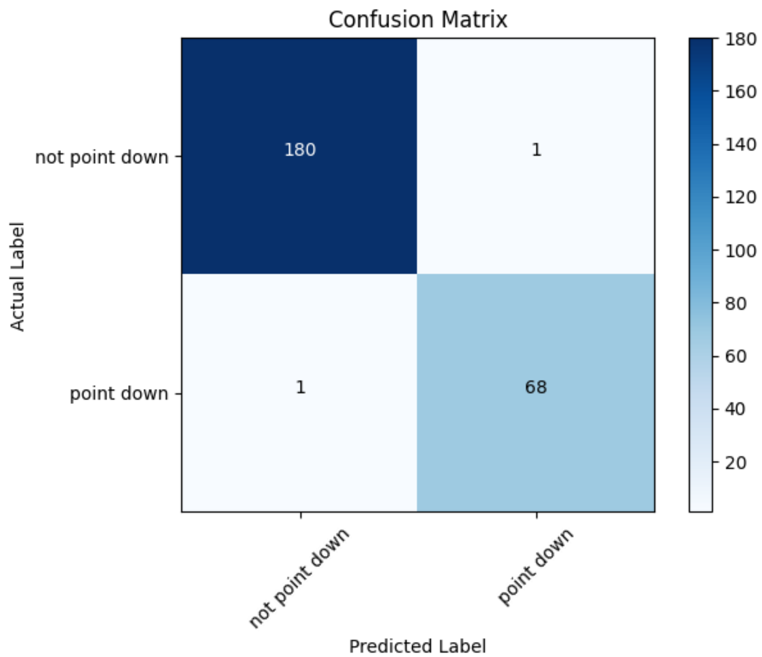


Figure 15. Point Down

When predicting which gesture is point down, the model did well since it only got two mistakes by predicting not point down where it was actually point down in the false positive side and predicting point down where it was actually not a point down gesture in the false negative side. The model got 68 when predicting point down correctly and predicted 180 times in the true negative section.

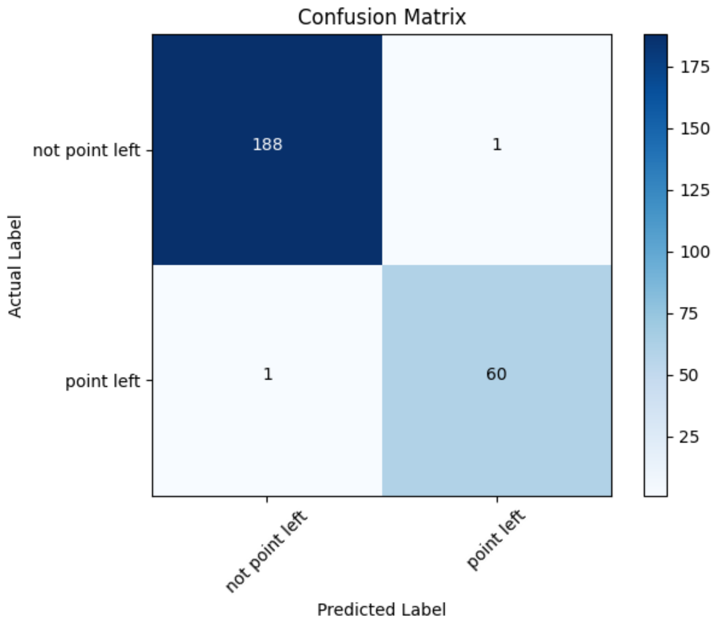


Figure 16. Point Left

The model got 188 score in the true negative side where it predicted not point left successfully and 60 in the true positive where it predicted point left correctly. It also have one score for each in the false negative and false positive.

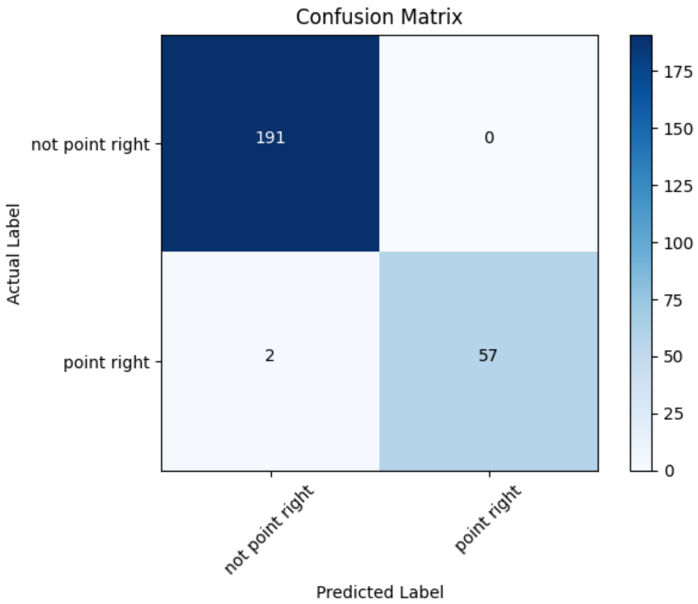


Figure 17. Point Right

The model got 191 score in the true negative side where it predicted not point left successfully and 57 in the true positive where it predicted point left correctly. It also have two score for in the false negative and none in the false positive.

Hand Signs/Hand Gestures

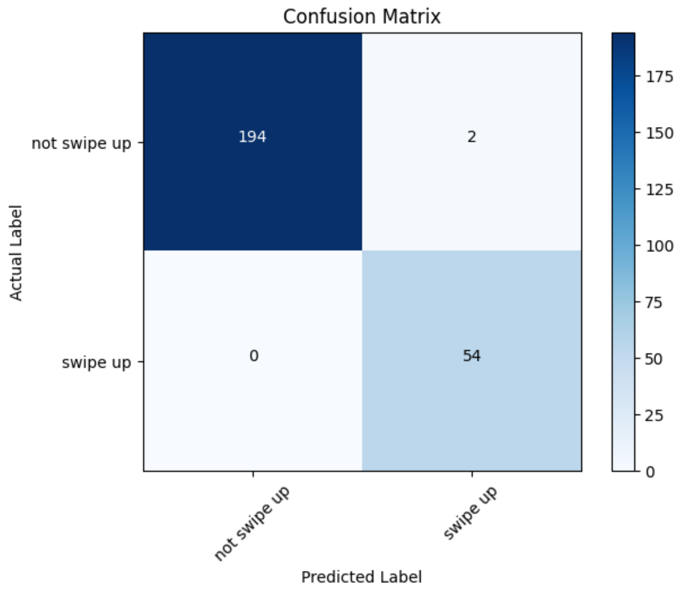


Figure 18. Swipe Up

The model got 194 score in the true negative side where it predicted not swipe up to its corresponding label successfully and 54 in the true positive where it predicted swipe up correctly. It also got two score for in the false negative and none in the false positive.

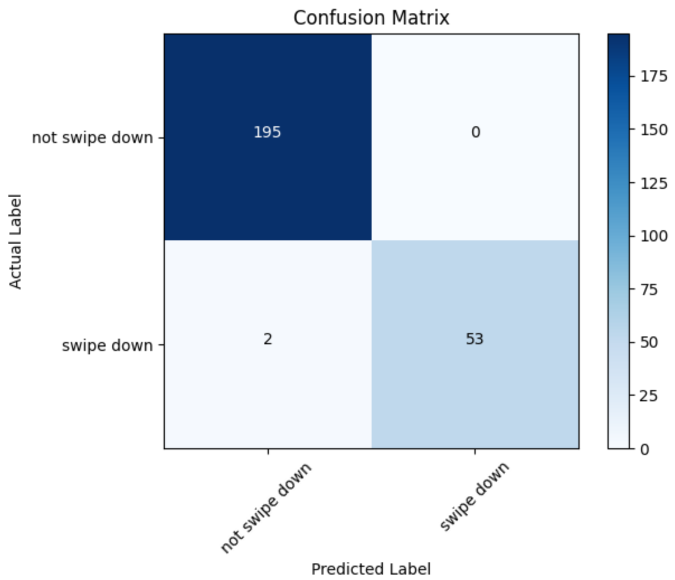


Figure 19. Swipe Down

The model got 195 score in the true negative side where it predicted not swipe down to its corresponding label successfully and 53 in the true positive where it predicted swipe down correctly. It also got two score for in the false positive and none in the false negative.

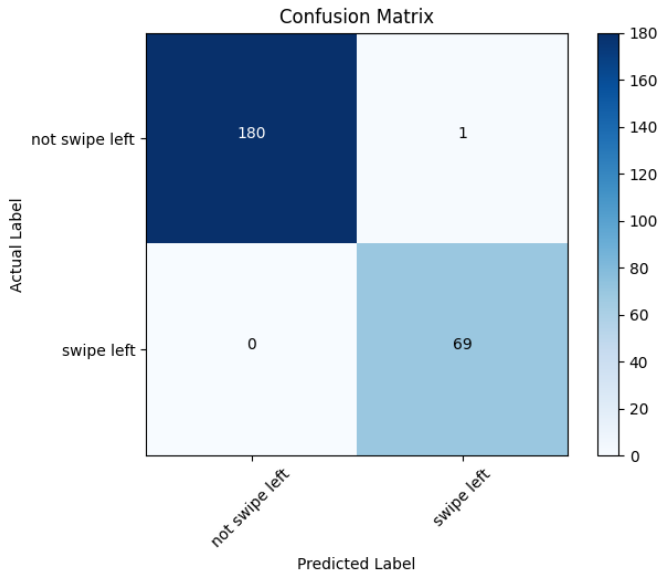


Figure 20. Swipe Left

The model got 180 score in the true negative side where it predicted not swipe left to its corresponding label successfully and 69 in the true positive where it predicted swipe left correctly. It also got one score for in the false negative and none in the false positive.

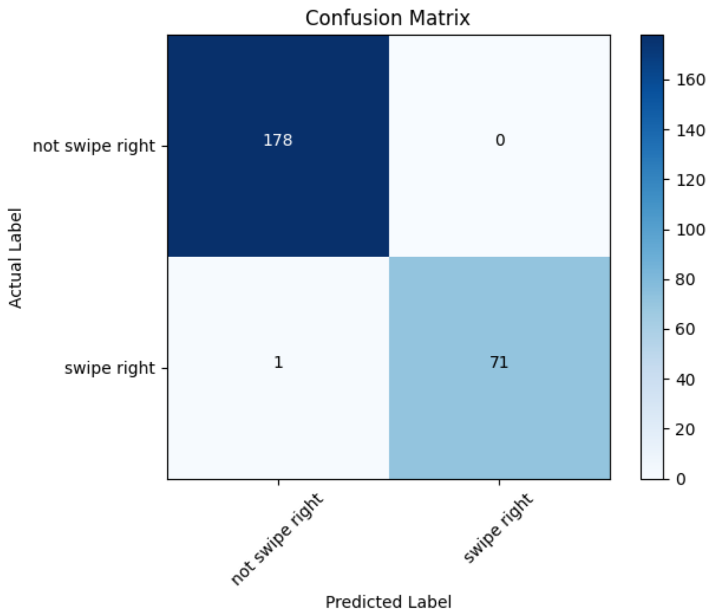


Figure 21. Swipe Right

The model got 178 score in the true negative side where it predicted not swipe right to its corresponding label successfully and 71 in the true positive where it predicted swipe left correctly. It also got one score for in the false positive and none in the false negative.

After observing the values within the confusion matrix of the hand point and hand swipe, the precision and recall can now be calculated. The precision is equal to the value of true positive (TP) divided by the total number true positive (TP) and false negative (FN). Precision answers the question when it predicts the right label, how often it is correct? Meanwhile, the recall can be measured by the value of true positive (TP) divided by the total number of the actual right label (true positive (TP) + false positive (FP)). Recall answers the question when it actually is the right label, how often does it predict it? The result of the precision and recall are now used to get the f1-score of each label in the prediction. The accuracy of the machine learning model is equivalent to the average f1-score of the prediction labels. The table below shows the summary of the result of the precision and recall of each hand pointer and hand swipe and its corresponding f1-score. The percentage accuracy is also seen below the f1-score of the model and the support column where it contains the total number of the specific label available out of the 250 testing data.

	Precision	**Recall**	**F1-score**	**Support**
Point Up	97%	100%	98%	61
Point Down	99%	99%	99%	69
Point Left	98%	98%	98%	69
Point Right	100%	97%	98%	59
Accuracy			**98%**	

Table 3. Different Metrics Used in Hand Pointer

	Precision	**Recall**	**F1-score**	**Support**
Swipe Up	99%	100%	99%	69
Swipe Down	100%	99%	99%	72
Swipe Left	96%	100%	98%	54
Swipe Right	100%	96%	98%	55
Accuracy			**99%**	

Table 4. Different Metrics Used in Hand Swipe

1.2 Speed

Test	Hand Mesh AI Model (sec)	Hand Mesh Pointer (sec)
1	2.60	1.91
2	2.97	1.90
3	2.90	1.91
4	3.14	1.92
5	3.00	1.92
6	3.38	1.9
7	2.97	1.91
8	3.00	1.90
9	2.95	1.93
10	3.18	1.93
Average	**3.009**	**1.913**

Table 5. Delay of the System in Terms of Seconds

The beginning of the time starts right before the first frame is read and ends after 30 frames had passed. This is made because the AI model will be trained to input short sequences of videos which consist of 30 frames. 30 frames is ideal as a good starting point than below 30 frames because it could be too fast to collect enough frames for the action and make misinterpretations. More than 30 is too long also because it could make the system make delay even more. The table shows the speed at which the system decodes the input and let the AI model predict. This values provides an idea to show the delay or lag from performing the action and the one displayed on the screen.

4. **Which recognition system will take less time to execute while being successful in interpreting:**

Test	Real-Time Image Processing	Hand Mesh AI Model	Hand Mesh Pointer
1	24.00%	21.80%	20.60%
2	22.30%	25.00%	22.30%
3	24.90%	24.00%	19.40%
Average	**23.73%**	**23.60%**	**20.76%**

Table 6. Usage of CPU in Percentage

The prototype is tested in a 4.3 GHz 4 Cores Processor (CPU). Table 4 shows the percentage of how much the CPU is being used when running the prototype. After taking three test, real time image processing gives an average of 23.73% while hand mesh AI model got an average of 23.60% and hand mesh pointer obtain 20.76% of average.

Test	Real Time Image Processing	Hand Mesh AI Model	Hand Mesh Pointer
1	301.5	439.5	367.5
2	298.3	432.9	367.4
3	310.6	433.4	367.5
Average	**303.47**	**435.27**	**367.47**

Table 7. Memory Usage in Megabytes

The memory usage of the prototype is tested in 8GB of RAM (Random Access Memory). Table 5 shows how many megabytes of main memory are being used when the system is running. Real-time image processing has a 303.47 average, hand mesh AI model obtain 435.27 megabytes and the average of hand mesh pointer is 367.47 megabytes. The data within table 4 and table 5 provides an idea of the hardware limits that the computer system can take.

Test	Real-Time Image Recognition (sec)	Hand Mesh AI Model (sec)	Hand Mesh Pointer (sec)
1	2.24	2.60	1.91
2	1.99	2.97	1.90
3	2.02	2.90	1.91
4	2.18	3.14	1.92
5	1.96	3.00	1.92
6	2.00	3.38	1.9
7	2.04	2.97	1.91
8	2.12	3.00	1.90
9	1.98	2.95	1.93
10	2.24	3.18	1.93
Average	**2.077**	**3.009**	**1.913**

Table 8. Speed of each Model in Terms of Execution

Table 6 represent the time of execution it takes the AI model to recognize the action done by the user. The data shows that real time image recognition had an average of 2.77 seconds while the hand mesh AI model took longer than expected since it got an average of 3.009 seconds. Meanwhile, hand mesh pointer model had the fastest execution time since it obtain 1.913 as an average.

Test	Real-Time Image Recognition	Hand Mesh AI Model
1	64.00%	99.54%
2	68.00%	99.08%
3	60.00%	99.70%
4	76.00%	99.09%
5	72.00%	99.04%
6	64.00%	99.20%
7	68.00%	99.69%
8	72.00%	98.30%
9	80.00%	99.07%
10	72.00%	99.53%
Average	**69.60%**	**99.22%**

Table 9. Accuracy Based on the AI Model

The used in the testing the accuracy are collected by creating a program that takes a random sample in the training data as testing data. The testing data are represented to the two AI model for it to predict which had the highest accuracy. Different questions are used because the real time image recognition can only use a single as an input while hand mesh approach can use up to 30 frames of data. The hand mesh AI model got an outstanding average of 99.22% while the model that uses single frame image processing did poorly since in only have 69.60% of accuracy.

Chapter 5

SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the summary of the findings made using the methodologies used, the conclusions made from these findings, and the recommendations on how to improve this study in terms of data gathering, processing and system designing. These improvements were part of the observations during the conduct of this study.

Summary of Findings

1. In the development of Touchless Interface as an Alternative to Public Touchscreen Technologies, prototyping method which is relevant in producing this system was effectively utilized. Every stages using this method were meticulously followed in order to pave the way for great results. Each of the processes in the prototyping method were fully satisfied until the implementation stage and lead to the completion of the system.
2. The effectiveness of Touchless Interface as an Alternative to Public Touchscreen Technologies was proven after garnering high percentage of accuracy. The accuracy of the system is measured by calculating the f1-score given by the results of precision and recall. The results of the f1-score is equal to the accuracy and effectiveness of the system. After testing the two AI models, its precision is 98% and 98% of recall value so the f1-score would be 98%. This means the responsiveness of clicking a button or maneuvering the cursor would be highly accurate.

3. The respondents found that the developed system is highly acceptable in terms of its functionality and usability. They highly accepts the system to be used and does its functions. They also found that the system is acceptable in terms of reliability. This means that the system accepted to be reliable in serving its purpose. This is a functioning application that has previously been implemented, but it is being maintained and progressed to avoid failures.
4. The researchers found out that the hand mesh approach is more suitable and accurate than the real-time image recognition in recognizing the different hand gestures used in the system.

Recommendations

1. Make the system user friendly to all people, especially for people with disabilities like to those who don’t have hands or limbs to use the interface.
2. Reduce the main memory usage when using the system.
3. Improve the responsiveness in recognizing hand gestures.
4. Reduce the CPU usage when using the system.
5. Fix some bugs where the application just give out randomly and output even when the user did not initiate an input.

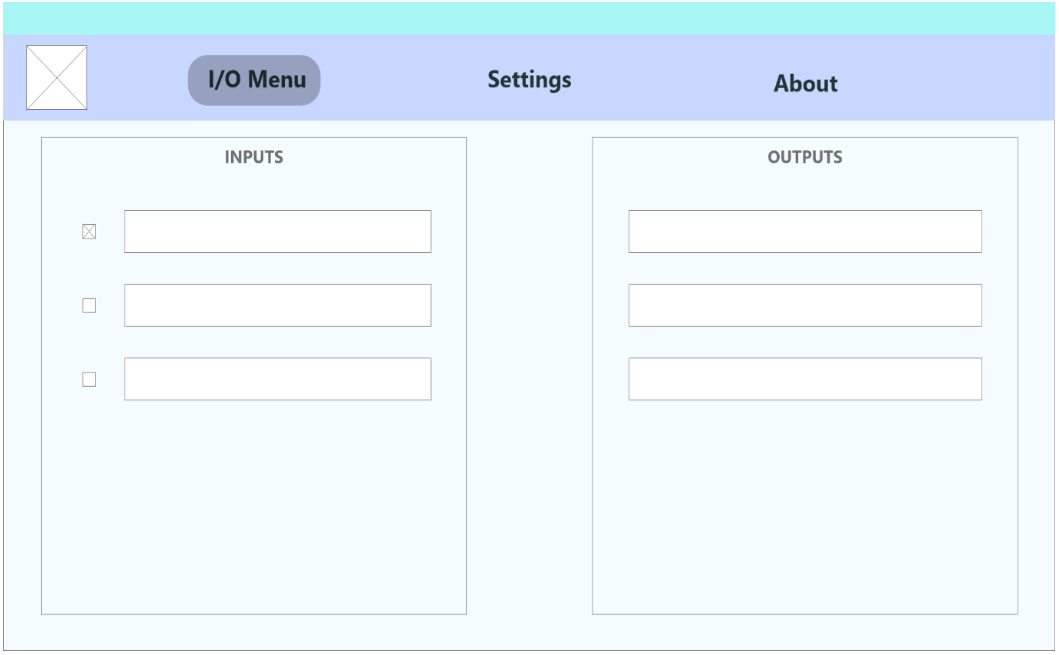
Appendices

Appendix A. Wire Frame

Wireframe is a visual guide of the skeletal framework of an application or website. Wireframe acts as the visual representation of a yet to be built website or application. Appendix A shows the wireframe of the research project, depicting the UI in which the users/admin will see.

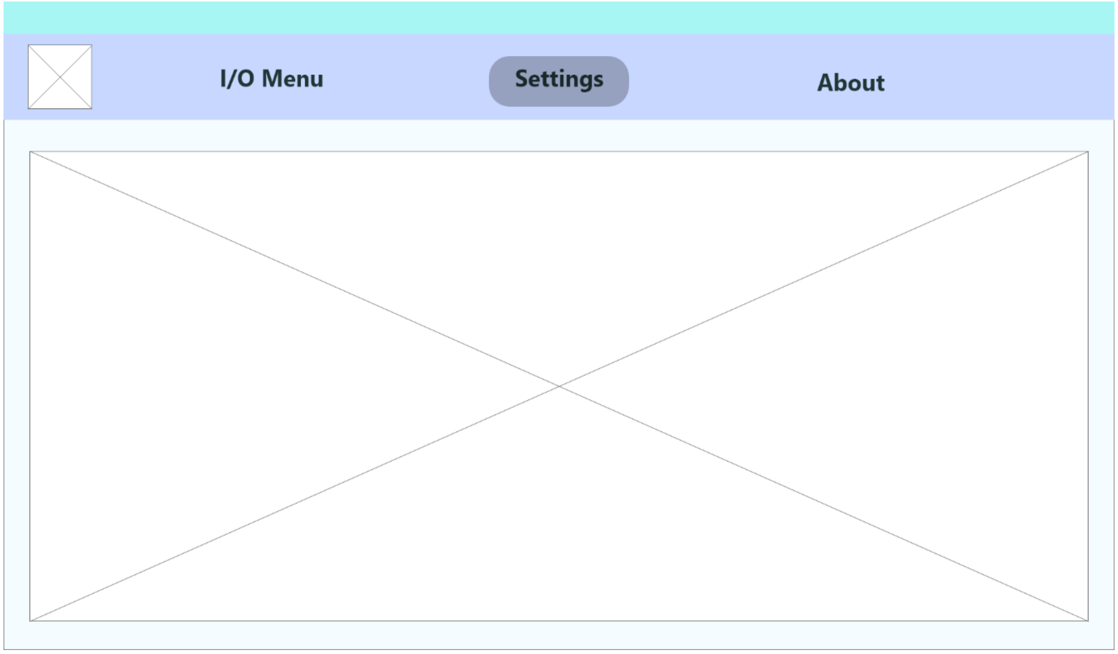
a. I/O Menu

Figure 22. I/O Menu Wireframe



b. Settings

Figure 23. Settings Wireframe



c. About

Figure 24. About Wireframe

References

Albawi, S., Mohammed, T., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. International Conference on Engineering and Technology (ICET), 1-6.

Alhmidi, H., Cadnum, J. L., Piedrahita BS, C. T., John MD, A. R., & Donskey MD, C. J. (2018). Evaluation of an automated ultraviolet-C light disinfection device and patient hand hygiene for reduction of pathogen transfer from interactive touchscreen computer kiosks. American Journal of Infection Control, 464-467.

Bezdan, T., & Bacanin, N. (2019, 1 1). Convolutional Neural Network Layers and Architectures. Retrieved from Researchgate.net: https://www.researchgate.net/publication/333242381_Convolutional_Neural_Network_Layers_and_Architectures

Bi, Q., Goodman, C. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. American Journal of Epidemiology, 2222–2239.

Butler, K. T., Davies, W. D., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. Nature, 547–555.

Cockburn, A., Gutwin, C., Palanque, P., Deleris, Y., Trask, C., Coveney, A., . . . Maclean, K. (2017). Turbulent Touch: Touchscreen Input for Cockpit Flight. International Conference for Human-Computer Interaction (CHI 2017), 6742-6753.

Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2020, January 1). Fundamental Concepts of Convolutional Neural Network. In A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, & D. De, Recent Trends and Advances in Artificial Intelligence and Internet of Things (pp. pp.519-567). Retrieved from Researchgate.net: https://www.researchgate.net/publication/337401161_Fundamental_Concepts_of_Convolutional_Neural_Network

Ireri, D., Belal, E., Okinda, C., Makange, N., & Ji, C. (2019). A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artificial Intelligence in Agriculture, 28-37.

Jekins, A., Lindsay, S., Eslambolchilar, P., Thornton, I. M., & Tales, A. (2016). Administering Cognitive Tests Through Touch Screen Tablet Devices: Potential Issues. Journal of Alzheimer's Disease, 1169-1182.

Kubát, M. (2017). An Introduction to Machine Learning. Springer International Publishing.

Kumar, V. B., Kumar, S. S., & Saboo, V. (2016). Dermatological disease detection using image processing and machine learning. Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), 1-6.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. Sensors, 2674.

Lucidchart. (n.d.). *What is Data Flow Diagram*. Retrieved from lucidchart.com: https://www.lucidchart.com/pages/data-flow-diagram

Lucidchart. (n.d.). *What is Flowchart*. Retrieved from lucidchart.com: https://www.lucidchart.com/pages/what-is-a-flowchart-tutorial

Mediapipe.dev. (n.d.). Retrieved from Mediapipe.com: https://opensource.google/projects/mediapipe

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundation of Machine Learning (Second Edition). Cambridge, Massachusettes: The MIT Press.

Nixon, M., & Aguado, A. (2019). Feature Extraction and Image Processing for Computer Vision Fourth Edition. London: Academic Press.

Orphanides, A. K., & Nam, C. S. (2017). Touchscreen interfaces in context: A systematic review of research into touchscreens across settings, populations, and implementations. Applied Ergonomics, 116-143.

Price, M., & Sup, F. (2016). A robotic touchscreen totem for two-dimensional haptic force display. IEEE Haptics Symposium (HAPTICS), 72-77.

Python (programming language). (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Python_(programming_language)

Seilevel. (n.d.). *System Flow* . Retrieved from https://seilevel.com/: https://seilevel.com/business-analyst-resources/business-requirements-models-templates/system-flow/

Soliman, N., & ALEnezi, A. (2019). A Method Of Skin Disease Detection Using Image Processing And Machine Learning. Procedia Computer Science, 85-92.

Tensoflow. (n.d.). Retrieved from Tensorflow: https://www.tensorflow.org/

Tensorflow. (2021, June 25). tf.keras.metrics.Accuracy. Retrieved from tensorflow.org: https://www.tensorflow.org/api_docs/python/tf/keras/metrics/Accuracy?version=nightly

Wang, Z., Di, H., Shatif, M. A., Aluadah, Y., & AlRegib, G. (2018). Successful leveraging of image processing and machine learning in seismic structural interpretation: A review. The Leading Edge, 402–480.

Wikipedia. (n.d.). *Entity-Relationship Model*. Retrieved from wikipedia.org: https://en.wikipedia.org/wiki/Entity%E2%80%93relationship_model

Wikipedia. (n.d.). *System Architecture* . Retrieved from wikipedia.org: https://en.wikipedia.org/wiki/Systems_architecture

Wikipedia. (n.d.). *Website Wireframe*. Retrieved from wikidepia.org: https://en.wikipedia.org/wiki/Website_wireframe

World HealthOraganization. (2020, March 27). who.int. Retrieved from who.int: https://www.who.int/docs/default-source/coronaviruse/risk-comms-updates/update-20-epi-win-covid-19.pdf?sfvrsn=5e0b2d74_2

Wu, J. (2017, May 1). Introduction to Convolutional Neural Networks. Retrieved from National Key Lab for Novel Software Technology: https://cs.nju.edu.cn/wujx/paper/CNN.pdf

Xia, K., Huang, J., & Wang, H. (2020, March 20). LSTM-CNN Architecture for Human Activity Recognition. Retrieved from Researchgate.net: https://www.researchgate.net/publication/340073805_LSTM-CNN_Architecture_for_Human_Activity_Recognition

Zhang, Y. (2021). Image Processing. Berlin: Walter de Gruyter GmbH & Co KG.

Zhu, R., & Li, Z. (2016). An Ergonomic Study on Influence of Touch-screen Phone Size on Single-hand Operation Performance. MATEC Web of Conferences .