**CHAPTER ONE**

**1.0 INTRODUCTION**

The spread of COVID- 19 pandemic disease has created the most crucial global health crisis of the world that has deep impact on humanity and the way the user perceives our world and our everyday lives. The unusual corona virus has resulted in person-to-person transmission but as far as we move, the transmission of the unusual corona causing corona virus disease 2019 (COVID-19) can also be a form a symptomatic carrier with no COVID-19 symptoms. Till now, there is no report about any clinically approved antiviral medicines or vaccine that are effective against COVID-19 virus. It has spread rapidly across the world bringing massive health, economic, environmental and social challenges to the entire human population. The world health organisation has clearly stated that until vaccines are found, the wearing of masks and social distancing are key tools to reduce spread of virus. So, it is important to make people wear masks in public places. In densely populated regions, it is difﬁcult to ﬁnd the persons not wearing the face mask and warn them.

Hence the needs for a real-time system that can determine whether a person wear a facemask or not, in this project we are utilising the power of machine learning, Deep learning and computer vision.

**1.1 BACKGROUND OF STUDY**

Machine learning as defined by (Marco Varone et al., 2020) is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

It is also described in journal of (Machine Learning with MATLAB (2016) as the uses of computational methods to "learn" information directly from data without relying on a predetermined equation as a model.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Deep Learning is as defined by (Zhang et al, 2012) is a process to establish relation between two or more than two parameters, along with attainment of knowledge design between required function to proposed structure. Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called Artificial Neural Networks. The term "Artificial Neural Network" is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes. In this project, the class of ANN we are utilising is Convolutional neural network (CNN).

Convolutional Neural Network is described by The MathWorks, Inc. (2021) as a network architecture for deep learning which learns directly from data, eliminating the need for manual feature extraction.

Computer Vision is described by Technopedia Inc (2018) as a field of computer science that works on enabling computers to see, identify and process images in the same way that human vision does, and then provide appropriate output.

It is like imparting human intelligence and instincts to a computer. In reality though, it is a difficult task to enable computers to recognize images of different objects. Computer vision is closely linked with artificial intelligence, as the computer must interpret what it sees, and then perform appropriate analysis or act accordingly.

**1.2 STATEMENT OF THE PROBLEM**

The spread of COVID- 19 pandemic disease has created the most crucial global health crisis of the world that has deep impact on humanity and the way the user perceives our world and our everyday lives. The unusual corona virus has resulted in person-to-person transmission but as far as we move, the transmission of the unusual corona causing corona virus disease 2019 (COVID-19) can also be a form a symptomatic carrier with no COVID-19 symptoms. Till now, there is no report about any clinically approved antiviral medicines or vaccine that are effective against COVID-19 virus. One of the major ways of preventing its spread is the uses of nose mask.

Design and implementation of facemask detector system using computer vision will solve these problems:

1. To detect if a person is wearing a wear nose mask.
2. To detect if a person is not wearing a nose mask.

**1.3 SIGNIFICANCE OF THE STUDY**

This presents the usefulness, importance and justification of the project. The importance of this project is stated below:

1. To design and implement a face mask detection system.
2. Gaining experience in computer vision, machine learning and image processing techniques.
3. It will also help in gaining experience on how to build and train machine learning model for real-world Application.
4. It will also help in gaining practical experience in deploying a machine learning model in production environment and learn how to evaluate the performance of a model.

**1.4 OBJECTIVE OF THE STUDY**

The aim of the project is to design and implement a facemask detector system using computer vision.

The objectives of the project are as follows:

1. T develop an object detection deep learning model using Convolution Neural Network (CNN) and Transfer Learning.
2. To collect, pre-process and analyse images of faces with and without facemasks.
3. To implement the facemask detection system using a Raspberry Pi 4B microcontroller.
4. To evaluate and test the facemask detection system.
5. To deploy the system in real-life environment

**1.4.1 SCOPE OF THE STUDY**

This project will focus on assembling of the required hardware and components such as raspberry pi display, raspberry pi camera, raspberry pi microcontroller.

It will also cover other aspect such gathering of required dataset, designing and transformation of the design into executable programming code.

**1.4.2 LIMITATION OF THE STUDY**

One of the major limitations of this project is that the model cannot be train from the scratch due to limited dataset and resources such memory, processing power of the hardware to train the model.

The vision device used for taking the sample of the images does not have light so its performance is limited to during day only or only if enough light can be provide during the night.

**1.5 DEFINITION OF TERMS**

MICROCONTROLLER

A microcontroller is a compact microcomputer designed to the govern and co-ordinate the operations of embedded system in motor, vehicles, robots, machines etc. it is a computer built onto a single printed circuit board. The board provides all the necessary circuit for a useful task. Microprocessor, i/o circuits, clock generator, RAM, stored program memory and other support IC’s (integrated circuit).

CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolution Neural Network is type of artificial neural network algorithm that used in extracting features from image. It comprises of three major layers which are convolutional layer, pooling layers fully connected layers.

DEEP LEARNING

Deep Learning is a type of machine learning in which several artificial neural networks are used to extract features (learning weight) from any given data. It is type of computer system that mimic human approaches of solving a problem.

**CHAPTPER TWO**

**LITERATRE REVIEW**

An image is a visual representation of an object either animate or inanimate.

Image is a representation of an object' or 'a reproduction of an object formed using a mirror or lens. Zetie, K. P. (2017)

It is also defined by R. Nick Bryan (2009) as a representation or reproduction of something.

Image is a representation or reproduction of something or an object in its exactness or its similar identity. The object representing can be animate or inanimate, perspective or imaginary.

Long ago, the function of image is for fashion, to attain immortality, it

let’s you see things that you may never notice or seen before. Also, as a form of communication.

In recent times with advent and improvement in technologies, the function of image has increased and improved to the extent that it can be used as tools in decision making with the use a computer vison.

Computer Vision is described by Technopedia Inc (2018) as a field of computer science that works on enabling computers to see, identify and process images in the same way that human vision does, and then provide appropriate output.

This bring about use of computer to detect, analyse and process image in real-time. The output can be used as a control or input for other systems.

There are several approaches the uses of computer vision, that is there are means of processing the image with computer vision such TensorFlow, PyThorch, OpenCV, Theano and Keras.

TensorFlow is an end-to-end open-source deep learning framework developed by Google and released in 2015. It is known for documentation and training support, scalable production and deployment options, multiple abstraction levels, and support for different platforms, such as Android.

TensorFlow is a symbolic math library used for neural networks and is best suited for dataflow programming across a range of tasks. It offers multiple abstraction levels for building and training models.

A promising and fast-growing entry in the world of deep learning, TensorFlow offers a flexible, comprehensive ecosystem of community resources, libraries, and tools that facilitate building and deploying machine learning apps.

Transfer Learning as described by Martin, S. (2019) is the deep learning technique that enables developers to harness a neural network used for one task and apply it to another domain. Transfer learning (TL) is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem (Wikipedia 2019).

In transfer learning, the knowledge of an already trained machine learning model is applied to a different but related problem. The general idea is to use the knowledge a model has learned from a task with a lot of available labelled training data in a new task that doesn't have much data. Instead of starting the learning process from scratch, we start with patterns learned from solving a related task. There is various model outside but we be discussing the major ones.

DETECT FACEMASK

RASPBERRY PI

CAMERA

DISPLAY

CLASSIFY

IMAGE

CAPTURING

LOAD

MODEL

POWER

SUPPLY

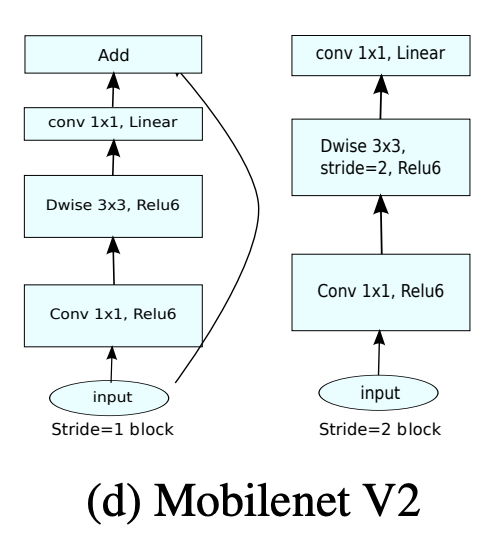
**Block Diagram of the system**

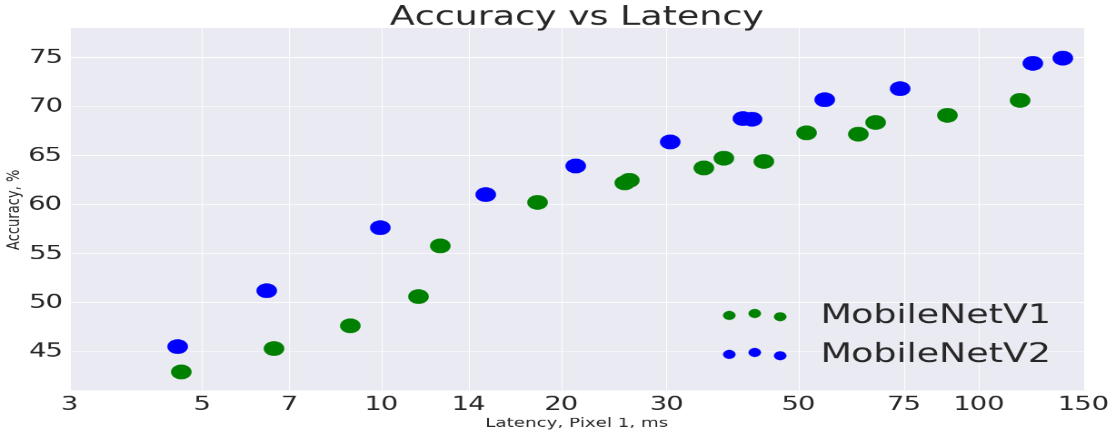
**2.1 MobileNetV2**

MobileNetV2 is a classification model (distinct from MobileNetSSDv2) developed by Google. It provides real-time classification capabilities under computing constraints in devices like smartphones. This implementation leverages transfer learning from ImageNet to your dataset.

**2.1.1 MobileNetV2 Architecture**

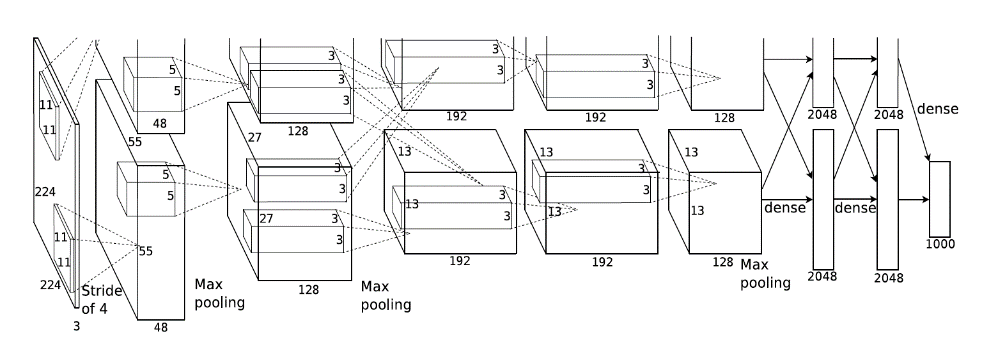
The MobileNetV2 architecture utilizes an inverted residual structure where the input and output of the residual blocks are thin bottleneck layers. MobileNetV2 also uses lightweight convolutions to filter features in the expansion layer. Finally, it removes non-linearities in the narrow layers.





**2.2 AlexNet**

AlexNet was designed by Hinton, winner of the 2012 ImageNet competition, and his student Alex Krizhevsky. It was also after that year that more and deeper neural networks were proposed, such as the excellent vgg, GoogleLeNet. Its official data model has an accuracy rate of 57.1% and top 1-5 reaches 80.2%. This is already quite outstanding for traditional machine learning classification algorithms.

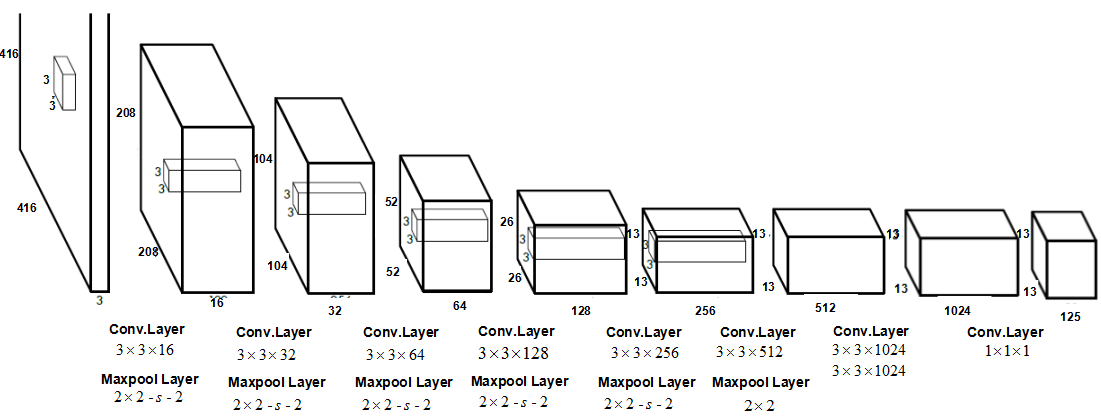


**2.3 YOLO**

YOLO is an abbreviation for the term ‘You Only Look Once’. This is an algorithm that detects and recognizes various objects in a picture (in real-time). Object detection in YOLO is done as a regression problem and provides the class probabilities of the detected images.

YOLO algorithm employs convolutional neural networks (CNN) to detect objects in real-time. As the name suggests, the algorithm requires only a single forward propagation through a neural network to detect objects.

This means that prediction in the entire image is done in a single algorithm run. The CNN is used to predict various class probabilities and bounding boxes simultaneously.



**COMPARATIVE ANALYSIS**

Different metrics have been proposed to measure object localization accuracy. The Intersection over Union (IoU) which is also called Jaccard Index, is commonly used to evaluate the accuracy of detections. It can be calculated as the area of overlap between a predicted detection and its corresponding ground-truth divided by the area of the union between the predicted detection and the ground truth. The mean IoU for an image is computed by taking the IoU of each

class and averaging them, for the binary or multi-class detection problems. This can be applied to all the images of the test dataset to have an average IoU value. Another related detection metric is the F1-score (also called Dice Coefficient), which is calculated as two times by the area of overlap divided by the total number of pixels contained in the detected and the ground truth regions. This measure can be represented in terms of Precision and Recall metrics. It also can be applied to all the target objects present in an image and we can compute the average F1 score for all images of the test dataset. The IoU and F1-score metrics are related and positively correlated for given fixed ground truth. This means that, while comparing two models using IoU if the first model is better than the second one using this metric, it will also be better using the F1 score. When taking the average score over a set of detections in images, the IoU metric has a tendency to penalize quantitatively single inaccurate detections more than the F1-score even when both of them can predict a given object instance is badly detected.

The standard metrics normally used for analysing object detection accuracy and speed include recall, precision, F1 score (F1), mean average precision (MAP), and frames per second (FPS). In the target detection process, precision is the ratio of correctly detected targets to the number of all detected targets and recall is the ratio of the number of accurately detected targets to all targets in the sample set. F1 represents the weighted harmonic average of precision and recall. Average precision (AP) is the precision across all elements of a category of pills, as defined in the formula given below:

Numerically, MAP is the average value of the AP sum across all categories, and this value is used to evaluate the overall performance of the model.

FPS is an indicator that is commonly used for evaluating the speed of model detection. The number of images that can be processed per second is referred to as FPS.

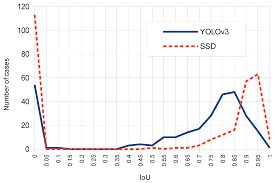
Both detectors can produce acceptable results for different object sizes, illumination conditions, image perspective, partial occlusion, complex background and multiple objects in scenes. One of the major strengths of SSD model is the almost elimination of FP cases which is preferable in applications related to the analysis. On the other side, YOLOv3 produces better average results. YOLO struggles to localize objects properly, but SSD is quicker than the previous progressive for single-shot detectors.

For real-time purposes, speed and accuracy are determining factors for smooth functioning. YOLO variants (especially up to YOLOv3) provide excellent accuracy but require computation-intensive hardware. For such devices, this model

would suffice the speed requirement. MobileNet-SSD V2 also provides a somewhat similar speed to that of YOLOv5s, but it just lacks in the accuracy. SSD could be a higher choice when we have a tendency to square measurable to run it on a video and therefore the truth trade-off is extremely modest. YOLO is a better option when exactness is considered than you want to go super quick. So, either of the models can be chosen depending on the requirement of various applications.

**CONCLUSIONS**

Real-time object detection and tracking on video streams is a crucial topic of surveillance systems in many field applications. The objective of our paper is to make a comparative study on two object recognition systems using CNN to identify the objects in the images. We studied and analysed the YOLO object detection model and MobileNet SSD model for performance evaluation in different scenarios. Each of the compared models has its own unique properties and is successful in its respective applications. YOLO provides better accuracy compared to MobileNet SSD, which provides more detection speed.



**2.3 CAMERA**

A camera is a device used to capture photographs or videos. It consists of a lens, a mechanism for capturing light, and a recording medium (such as film or a digital image sensor). When you press the button to take a photo or start recording a video, the camera captures light through the lens and records it onto the medium. The resulting image or video is a representation of the scene or subject that the camera was pointed at when the photograph or video was taken. The main classification of cameras are analogue camera and digital camera.



TYPICAL IMAGE OF A CAMERA

**2.3.1 ANALOGUE CAMERA**

An analogue camera is a type of camera that captures and stores photographs on film, rather than in digital format. Film is a strip of light-sensitive material that is exposed to light when a photograph is taken. The image is captured on the film through the camera's lens and is then developed using a chemical process. Analogue cameras are also known as film cameras. They were the first type of camera to be invented, and for many years were the only type of camera available. The output of analogue camera is called picture.



ANALOGUE CAMERA

**2.3.2 DIGITAL CAMERA**

A digital camera is a camera that captures and stores photographs in digital format rather than on film. Digital cameras use an image sensor to capture light and convert it into digital data, which is then stored on a memory card or other storage device. One of the main advantages of digital cameras is that they allow you to review and edit your photos on a computer or other device, rather than having to wait for film to be developed. Digital cameras also offer a variety of features and settings that allow you to control the quality and style of your photos, such as white balance, exposure, and resolution. There are many different types of digital cameras, including DSLR cameras, point-and-shoot cameras, and smartphone cameras. The output of digital camera is called an image.



DIGITAL CAMERA

**2. RASPBERRY PI**

Raspberry Pi is a small, single-board computer that was developed in the United Kingdom by the Raspberry Pi Foundation to promote the teaching of basic computer science in schools and in developing countries. The Raspberry Pi is a low-cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. It is a fully functional computer, capable of running a variety of operating systems and software applications, including word processors, games, and media players. Despite its small size and low price, the Raspberry Pi is a powerful and versatile device that can be used for a wide range of projects, from building a home media centre to learning to program or creating a home automation system.



**RASPBERRY PI**

**CHAPTER THREE**

**3.0 METHODOLOGY**

It is significant to make sure that that the project run smoothly as predicted earlier. Thus, the method used in the overall implementation of the project are stated below.

1. Uses of camera to capture the image and send it to the raspberry microprocessor
2. The raspberry will load the trained data from the storage disk.
3. Process the image.
4. Application of face mask detector to determine “mask” or “no mask”
5. The result will be display on the screen.

LOAD DATA

TRAIN FACE MASK CLASSIFIER

APPLY FACEMASK CLASSIFIFIER TO EACH FACE TO DETERMINE MASKED OR NO MASKED

FACE FROM IMAGE/VIDEO

SHOW RESULT

MASKED OR NO MASKED

EXTRACT EACH FACE

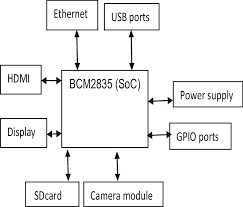
**FLOW CHART OF THE PROJECT**

**3.1 DESIGN ANALYSIS**

To achieve the objectives of design, development and construction of nose mask detecting system, the process was divided into two stage. First stage involves the setting up of hardware component which include the connection of raspberry pi microcontroller with other hardware components such camera and the display. Second stage involve the development of the software part with the uses of python and its associated libraries.

**3.1.1 RASPBERRY PI MICROCONTROLLER**

This is the brain of the project. It is a computer on a small board on which the necessary software required are installed. It also has memory on which the trained that are stored. It controls and coordinate all other component such as camera and the display device. The type used in the project is raspberry pi 4B with 8gigabyte of RAM

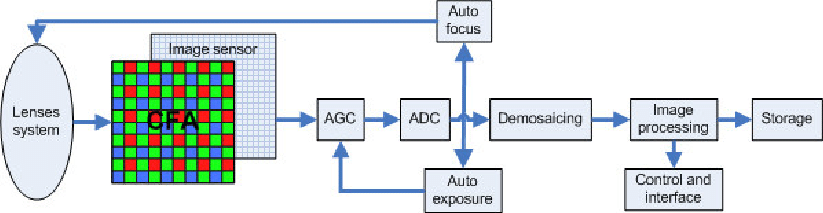
****

**BLOCK DIAGRAM OF RASPBERRY PI**

**3.1.2 CAMERA**

This is the device for taking the sample of pictures.

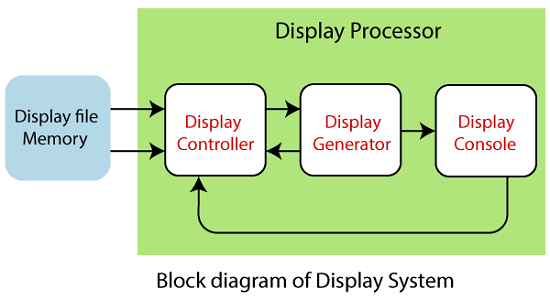
It is an input device for the raspberry pi, the type used here is digital camera because it provides direct manipulation of the images.



**BLOCK DIAGRAM OF CAMERA**

**3.1.3 DISPLAY**

This is the output device on which the result will be presented. The type used here is raspberry pi display device which capacitive touchscreen with screen size of 7 inches.



**3.2 SELECTION OF MATERIALS**

This work considers some factors during selection of material such processing time of the raspberry pi because this project must produce its output in a real-time that is its output is a function of time, it must give us result in a finite amount of time that’s why we choose latest version of raspberry pi with high configuration.

We also carefully selected the camera used because we need direct manipulation of images taken by the camera that’s why we choose digital camera because analogue camera lacks that feature.

also, we put weight of the project into consideration by choosing digital camera because of its smaller size as compared to analogue camera.

The display device used is 7inches capacitive touchscreen which gives clear view of the output and eliminates the needs for external keyboard inputting of commands.

**3.3 CONSTRUCTION PROCEDURES**

There are 2 major steps in the designing and the construction of this project. The steps include:

STEP 1(HARDWARE PART)

Connecting the input and output components with the raspberry pi microcontroller. The input device is the camera that takes the picture or the live video, process it and send to the raspberry pi. The raspberry pi will extract the require features from the image, analyse it and compare it with given model then it will decide the class which the picture belongs to either “with mask” or “No mask”. The result will be display on the 7 inches capacitive touchscreen connected to the raspberry pi that serves as the output device.

STEP 2 (SOFTWARE PART)

\* The first step involved installation of operating system on the raspberry pi using raspberry Operating System imager. An imager is an IDE for downloading and installing raspberry pi operating system on a new raspberry pi.

\* Installation of python programming language and other required dependencies and its associated libraries (OpenCV, TensorFlow, NumPy, sciplot, keras etc.)

\* Training of the face mask detector to generate model

\* uses of pre-trained model (MobileNetV2)

\* Comparation between trained model and pre-trained model.

**3.4 BILL OF ENGINEERING MATERIALS AND EVALUATIONS**

|  |  |  |  |
| --- | --- | --- | --- |
| S/N | COMPONENTS | QUANTITY | PRICE |
| 1 | Raspberry pi 4 Model B | 1 | #58, 000 |
| 2 | Raspberry pi camera module | 1 | #19, 000 |
| 3 | Raspberry pi Touch Display | 1 | #45,000 |
| 4 | Casing stand for the system | 1 | #19,000 |
| 5 | Raspberry pi case | 1 | #4, 000 |
| 6 | Storage device (SD 32G) | 1 | #5, 000 |
| 7 | Total | 6 | #150, 000 |

**CHAPTER FOUR**

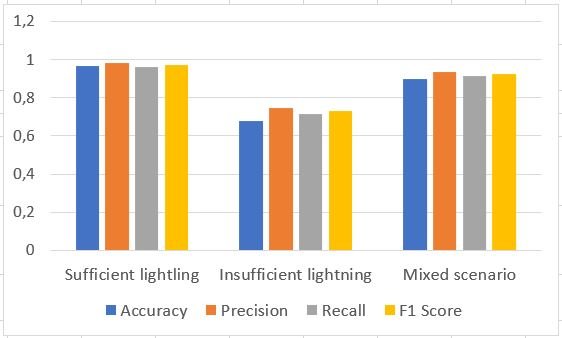
**4.0 RESULT AND DISCUSSION**

**4.1 PERFOMNACE TEST**

The testing of the entire project was carried out in stages.

As mentioned in the previous chapter in the dataset section, of the 2,400 images collected, 80% of the images will be used to train the cascade and the remaining 20% will be used as a testing dataset for the trained model.

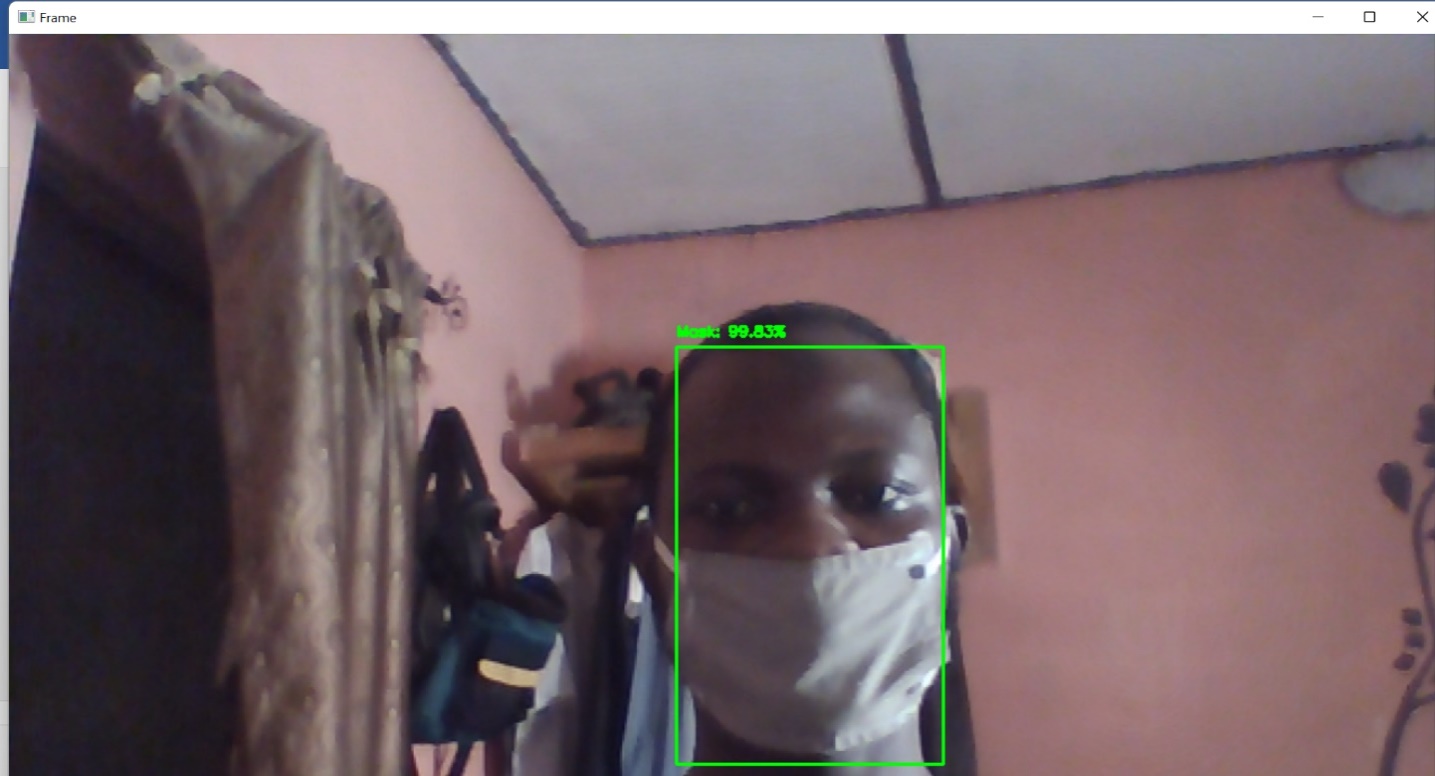
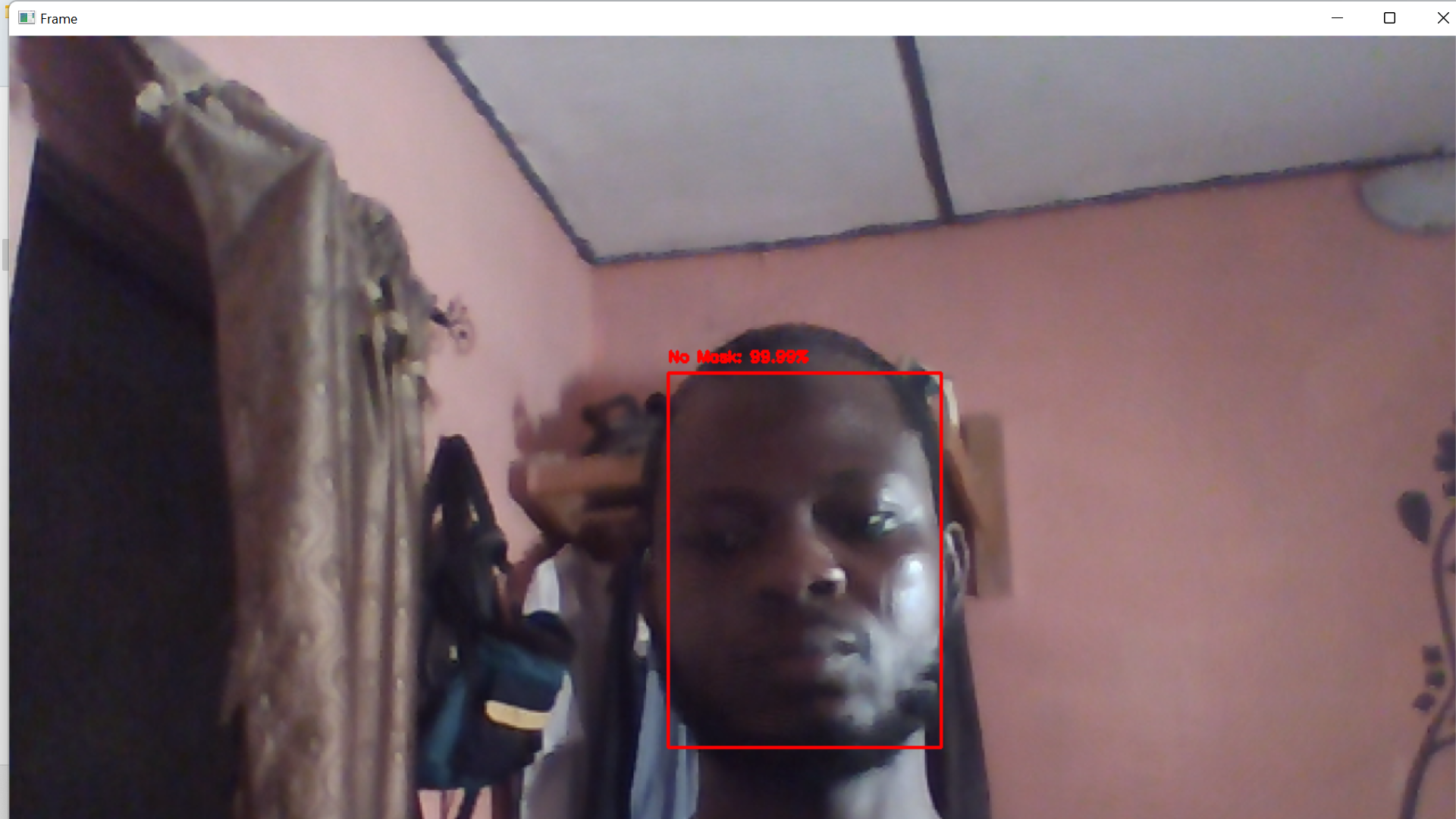
In the following test, the lighting condition of the facemask detection is taken into consideration. The data set is divided into 3 categories, namely sufﬁcient lighting, insufﬁcient lighting, and mixed scenarios. The ﬁrst scenario with a dataset containing images of faces with and without masks that have sufﬁcient lighting has the highest accuracy, namely 96.6%. The image dataset that has insufﬁcient lighting has the lowest accuracy, namely 67.7%. In the mixed image dataset between sufﬁcient lighting and insufﬁcient lighting, the accuracy is 89.7%. In addition, as in the previous test, in each scenario the precision value is consistently higher than the recall value.



COMPARESON BETWEEN THE THREE CONDITIONS

**4.2 PRESENTATION OF RESULT**

The face mask detecting system utilized a raspberry camera for snapping the image, the output of the camera is fed into the raspberry pi microcontroller which load the data from its storage disk, process the image then output the result as either masked or no mask. The sample of the two possible outcomes from the project is shown in the fig below.

Face with mask Face without mask

**CHAPTER FIVE**

**5.0 CONCLUSION AND RECOMMENDATION**

**5.1 CONCLUSSION**

This project was designed considering some factors such as response time, process time, economy durability, availability of research materials and the availability of the components of the design specification. The general operation of this project depends on the training of the model and its accuracy depends on the amount of dataset used in training the model. Therefore, there is sufficient training and enough dataset to aid the performance and accuracy of the system.

The construction of the whole system is modular that is any part of the system can remove and replace easily which makes its maintenance and repair easier in the event of any breakdown. The work was quite tedious but eventually was s success. Its successful completion shows that there are many ways to manipulate image, computer vision with advance in technology can be combine to use as great tools in decision making, and finally computer vision can be deployed in the detection and identification of any objects.

**5.2 RECOMMENDATION**

There are many improvements that can be made in the future model of this project. Since this project aims at identifying weather someone is wearing a nose mask or not, there are many improvements that can be made which includes

1. An advanced camera which has light and can take more vivid and clear image can be used for taking of images.
2. The storage device on the system could be increase to accommodate more image and dataset.
3. Also, the microprocessor (RAPSBERRY PI) used can be replaced with a better microprocessor manufactured specially for this project so as to improve the processing power and reduce the cost of production.
4. Instead of using display only for the output, the output could be sent to database for future use, also to get more dataset to train the model which will in turn improve the accuracy of the system.

**REFERENCES**

* Abbass, H. (2021, April 1). Editorial: What is Artificial Intelligence? IEEE Transactions on Artificial Intelligence. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/TAI.2021.3096243>
* Varone, M., Mayer, D., & Melegari, A. (2020). What is Machine Learning? A definition. Retrieved from https://expertsystem.com/machine-learning-definition/#:~:text=A definition, -Expert System Team text=Machine learning is an application, use it learn for themselves.
* MathWorks. (2018). Machine Learning with MATLAB. The MathWorks, Inc, 1. Retrieved from https://www.mathworks.com/content/dam/mathworks/ebook/gated/machine-learning-ebook-all-chapters.pdf%0Ahttps://www.mathworks.com/campaigns/products/offer/machine-learning-with-matlab.html
* Zhang, X., Wang, L., & Su, Y. (2021). Visual place recognition: A survey from deep learning perspective. Pattern Recognition, 113. https://doi.org/10.1016/j.patcog.2020.107760

Guo, W., Huang, L., & Liang, L. (2020). A weld seam dataset and automatic

**.** R. Nick Bryan bIntroduction to the Science of Medical Imaging (2009)

. Zetie, K. P. (2017). What is an image? Physics Education, 52(2). <https://doi.org/10.1088/1361-6552/aa544e>. Martin, S. (2019). What Is Transfer Learning? | NVIDIA Blog. Retrieved from https://blogs.nvidia.com/blog/2019/02/07/what-is-transfer-learning