IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

AirMouse and AirWrite Intuitive Digital Interaction

M. Krishna Satya Varma[1], M. Chaitanya[2], M. Srihari Bhushan[3], M. J. N.D. Sivasai[4], K. Charan Viswa[5]
Assistant Professor, Department of Information Technology [1]
Students, Department of Information Technology [2,3,4,5]
S.R.K.R. Engineering College, Bhimavaram, Andhra Pradesh, India
krishnasatyavarma@gmail.com, chaitumolleti@gmail.com
sriharimandiga@gmail.com, sivasaimaddula374@gmail.com, charanviswa06@gmail.com

Abstract*: The primary purpose of this project is to redefine the way people will interact with their systems more easily, and this will be achieved by eliminating the touchpads and keyboards to connect to the computer. This cutting-edge system eliminates the need for a regular mouse or a keyboard, allowing for easier device user-interface adaptation. The implementation of the latest gesture recognition technology lets the system not only pick up and decipher the movements of users' hands but also translate them into wider operations that are similar to those of a mouse and keyboard.*

Central to the design of the metaphoric system is the "virtual mouse," which enables the cursor movement and the performance of various mouse buttons without touching anything physically. So, it is a virtual mouse with an on-the-screen cursor that reacts to hand gestures in real-time. The system utilizes certain hand gestures enough to enable the effortless execution of actions commonly associated with a traditional mouse including left-clicking, right-clicking, and dragging across the screen. The virtual mouse feature is enhanced by the air writing technique, commonly acknowledged as a milestone in input development. This capability precisely reads and decodes the movements developed in the space by the fingertips while conveying them to the screen as written words. The air writing system would predict the characters from the user's gestures, creating a fresh, secure, and user-friendly method that favors the option of air typing rather than relying on a physical keyboard.

This project brings innovation in human-computer interaction when implemented using powerful computer vision and machine learning techniques and promotes accessibility and usability of the system where the traditional inputs become inapplicable. On the other hand, its feasible justifications range from medical uses to pioneering in virtual reality, gaming, and education that drives brilliantly the new options in interaction across various fields.

Keywords: Residual Network, air writing recognition, hand tracking, computer vision

I. INTRODUCTION

Nowadays technology is improving day by day, mostly electronic devices. But for some individuals with limited mobility, using technology is very difficult. They cannot interact with computers normally. They cannot use traditional mouse or keyboards normally. Our system provides a touchless interface which is useful for connecting with computers. This technology is perfect for people with impairments since it offers a more accessible, natural, and intuitive method to interact with computers. This touchless interface can completely change how people interact with computers. This system uses computer vision algorithms and image processing techniques, which are used to capture the user's hand gestures. This allows users to interact with a computer through an air mouse which is used to do mouse operations and write the text through air writing. In the initial phase, we have to verify which algorithm is best for this air writing and finalize it. We have to find out the dataset which contains all alphabets and digits which is used to compare with what we have written through air writing. The proposed system solves these problems through the touchless intersect solution that obviously replaces the physical interaction with the devices but also promotes accessibility and ergonomics. Such a system is capable of tracking hand movements in real-time and translating those gestures into mouse actions and text input, hence providing users with outstanding ease and precision to navigate a computer screen and type in texts. Previous base papers that have the concept of air writing and virtual mouse are very useful. We have taken references from them, which are very useful. For example, in [1], The system employs a reverse time-ordered shape context to represent air-writing trajectories in a backward manner, effectively eliminating the need for extraneous starting-lift data. Like this, we have taken references from different base papers.

We have trained different algorithms to get accurate results. At first, we trained the CNN algorithm with reference from [2]. Later other algorithms like MRF CNN, VGG16, and RESNET CNN are verified. We trained these algorithms against the EMNST dataset which contains alphabets and digits. The accurate results are obtained with the help of RESNET CNN. The RESNET's residual block consists of shortcut connections which will skip some layers. We have also made reference to implementing an air mouse system. For example, in [7], the system uses the HSV color detection technique for capturing the hand gestures of the user. It also uses Python and OpenCV libraries for virtual mouse interaction. We also use other libraries and frameworks like TensorFlow, Media pipe, pyautogui, pyttsx3, and many more. We can do mouse operations like left-click, right-click, and dragging of mouse curser. This air mouse and air writing system has many features. In the air mouse system, we can drag the curser with the help of our hand gesture. With specific hand gestures we can do left click, right click functions. We can also increase or decrease the volume of the computer with the help of hand gestures. In the air writing system, what we have written in the air will be saved in a text file temporarily. Later, based on the user's requirements it can be turned into speech or it will be saved permanently in either .txt or .docx format. Turning text into speech means what we have written will be spoken by the system. We can write both digits and alphabets. An air mouse and air writing system could increase accessibility for people with impairments and improve accessibility to computers for all users. There are some disadvantages and problems with traditional input devices. They can spread illness from one user to another user through the touching of a physical mouse or keyboard. This system will decrease the spreading of illness. Because of this virtual mouse, the germs do not spread that easily as there is no physical contact. In the coming steps of this project, we will try to increase the accuracy and increase the number of gestures and their operations. The study is implemented in a way that involves conducting a series of experiments to evaluate the accuracy of the system and its usability. In this regard, the demonstration of the system's effectiveness in various tasks such as navigating interfaces, clicking on objects, and writing text in the air will take place. Therefore, it also touches on what a wide range of areas can be covered by the system, including health, education, gaming, and accessibility.

II. EXISTING SYSTEMS

Modern studies examine the use of hand movements for controlling computer devices and their interaction with the users' environment. I have explored this research due to the fact that I need a more natural and intuitive scheme of human-machine interaction.

The given paper called "Deep Learning-Based Real-Time AI Virtual Mouse System Using Computer Vision for COVID-19 Epidemic Prevention"[9], appears with the notion that the Media Pipe package is used on for tracking the hands as well for tracking the tip of the hands. Historically, this approach to disease prevention has been found to be accurate and effective. OpenCV library (Open-Source Computer Vision) is a computer vision library. The system takes a snap of humans hand gestures in real time, and the technology is able to grasp many gestures in an accurate way.

The paper "Air-Writing Recognition Based on Deep Convolutional Neural Networks [8], about the system that captures hand gestures, through a web camera," examined. They claimed that CNNA type of AI deep learning with high accuracy level of 98% for fingerprint detection and 99% for gestures and hand position of the sportswomen are possible.

In the "Virtual mouse using object tracking" [7] system, is based on the HSV color detection technique for sweeping an excess of the user's hand waver. The second step is about how to use the keyboard as an alternative, and the last one is the system used to move the cursor and perform other mouse operations such as left click and right click.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

Limitations

Sensitive to Multiple hands: Multiple features in the background opposition of current pixel detector system do recognize the single hand. It brings an issue of accurate motion tracking and of prediction of the hand movement. Apart from this, there are some other algorithms needed to separate different hands, which could be a tricky part as they may somehow create confusion and inaccuracies in hand detection.

Sensitivity to hand gestures: Virtual keyboards and mice utilizing hand motions can be unforgiving in terms of gesture orientation, causing accidental actions or tracking inaccuracies.

Individual differences and calibration requirements: The same device could aid different users, but they all may have different writing styles or gestures, so the performance can be recalibrated or even customized, in order to do its best work.

Sensitivity to environmental conditions: Effects of virtual mouse systems and Air-writing recognition systems on surrounding factors such as ambient noise and lighting conditions are present, thus giving the low accuracy percentage.

Need of external equipment: The system needed the user to put either colored hats or lines on their fingers.

III. PROBLEM STATEMENT

This project aims to design a system that employs both virtual mouse and air writing technology to allow users to work with computers. Therefore, it develops a computer input system, which uses hand movements to move the mouse cursor. Besides that, it has the capability of finger air signing onto the computer screen. The system must be able to recognize the gestures of the user accurately and respond to them. Each gesture must be designated for an individual purpose. This system ought to be equipped with the capability, for example, to control the mouse cursor and finger air writing for text input. This system will generate an interface in which the user is able to interact with the computer without any physical input devices such as a mouse and keyboard. This makes it possible for a disabled person who has limited mobility to use a computer through the simple movement of his or her hands in the air.

IV. PROPOSED SYSTEM

With the help of our invention, users can do away entirely with mouse devices and make hand gestures in the air as the mouse will function that way. The clarity and easy navigation are made possible by this natural and intuitive manner of interaction where the screen elements change which is the feature of a seamless user experience.In our system, to increase the accuracy, we apply RESNET (the Residual Networks). In short, RESNET's residual learning approach made significant breakthroughs in the development of deep learning strategies by resolving the problem of training very much deeper neural networks. The resoluteness of RESNET tackling vanishing gradients will also facilitate deep learning, and likely to have a higher precision of air writing recognition. Our system depends on EMNIST and gives the symbol of what was written in the air. We will be using the EMNIST dataset which has alphabets and digits to be able to do alphabets and digits as well.Our proposed system includes an air mouse, which is an advantage for various functions like moving of curser, left click, right click, volume increase and decrease, etc. The OpenCV library for computer vision was our choice for this project. This library is used to process images. It outlines what is depicted by a picture, which is transmitted by a webcam. Additionally, we use the media pipe framework to facilitate hand tracking and image recognition. The Media pipe framework connects pieces of real-time image recognition as a streaming pipeline. This system relies on various other libraries such as Keras, pyautogui, and pyttsx3. pyttsx3 is a library here, which is used to convert text to speech. Our library plays a vital role in the speech features of auto-writing systems.Our envisioned system will be established through libraries and existing algorithms. These systems can be designed in a way that they enable direct communication with the computer system for its users without the need for a physical mouse and keyboard hardware. The system that is suggested in the proposed method is user-friendly.

Advantages of Proposed System

The proposed system has considerable merits because of the fact that is better than the existing one. The inadequacies in the old system which have been set out above are overcome. Our system is made usable and functional with a full set of options that embrace both standard mouse control and something more.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

- No multiple hand detection: We will also ensure that hand-over areas such as doors, hallways and stairs will not be detected by the system. The hand tracking module helps in hand accuracy that prevents multiple hand tracking detections.
- RESNET CNN: While varying users all have their own unique style, RESNET CNN architecture is still actually able to predict from the hand gestures of the different users without the same handwriting style. RESNET, the famous residual network with fully connected layers, is used for correctly classifying hand gestures under the visual data input.
- No need for external equipment: For the sake of the wearer, the device doesn't need any coloured caps or fingertips with typical tapes for hand recognition, which is the last step.

Additional Features:

- The main feature in this air mouse is that it is being able to do more than the standard air mouse. Such includes increasing and decreasing the volume of PC with hand gestures.
- Air writing provides a possibility to keep the text from written in the air by saving files in some popular formats such as .docx or .txt as well.
- This system has a programming called text-to-speech conversion for the visuals for those users that would rather hear the text instead of the read it.
- Furthermore, we have also added colour change buttons to make it easier for you to modify the color of the handwritten virtual letter brush, and also added clear canvas option.

Basically, the proposed system provides the user with features such as straightforward interaction, amusing interface, a boost in the user's productivity level, accessibility, diversity, and so much potential to develop other system that leaves it as a noteworthy and creative approach to computing interfaces.

V. ARCHITECTURE

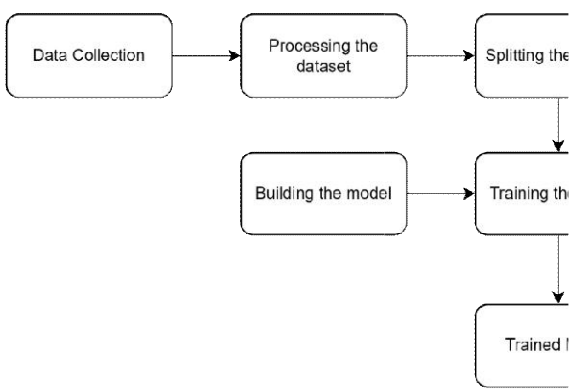

Fig 1: Model Building Architecture

Data Collection: Involving data collection is a necessary step for the development of a machine learning model. We have to determine the data access feature of our system. The goals of this ideation session include involving all sources of data to acquire our needed data. The data sources can range from open data such as the Internet repositories, and scrap data of webpages to private data. EMNIST dataset is the dataset that we have used for collecting the data we required from the online tool. Every data set merges Alice, bice, counts, and digits. These are the basic requirements of our system.

Processing the dataset: The dataset we will be working on will need to be to come in a particular format to train and evaluate our model. Processing datasets is, as a matter of fact, the cleaning of the dataset. Purging of the dataset consists of the removal of duplicated copies, and data with noisy and abnormal values.

Splitting the dataset: To learn the model, we need some parts of the dataset. Therefore, we divided the dataset into two groups as train dataset and test dataset respectively. We applied splitting in different ways like random splitting, stratified splitting, or time-based splitting etc. We did random splitting, training data is part-set of the whole data, and the testing dataset is the remaining part of it.

Building the Model: It is essential for us to make the decision concerning what algorithm is better for our system on this stage. We need to look at various algorithms, research them, and test them. Ensure that you choose an algorithm that provides the best-fitting results.

Training the Model: We must train the model which we selected in the above step. The above split train dataset is used to train the model. This training helps the model to learn the relationships and patterns present in the training data. Now the trained model will be able to do the predictions on new data.

VI. IMPLEMENTATION

Algorithm Used

RESNET CNN:

The resnet (Residual Network) is an innovative neural network design, which enables the training of much deeper networks than was possible before. Unlike the conventional architectures, the resnet retains many of the stacked layers without going through them. He et al. brought to manifest the Resnet which over makes up for the disappearing and exploding gradient problem that hindered deep neural networks training by employing residual blocks with skip connections. Innovation enabled effective training of deep networks with depths of more than a hundred layers. These are significant steps toward accomplishment of tasks in the field of computer vision.

Resent, which underlies the learner on the principle of residual learning, is in place. Blocks in Resent structure require fewer convolutional layers but still have a few batch normalization layers and ReLU activation functions. Unlike regular blocks, in RESNET, the shortcut connection allows skipping the non-linear transformations. In this case, the input term is accumulated at the output stage, which eventually helps to maintain the signal gradient signal across many layers of the network. This facilitates the training of deeper networks that show stable and competent performance throughout.

Generally, the convolutional layers are the first in the architecture, having a function of processing the input image, followed by a max-pooling layer. The principal of the topology is the stack of sequential residual blocks, where each next stage from the top can double the filter's number while decreasing the spatial dimensionality of the feature map with the first layer as a stride-2 convolution. It is with this structure that RESNET appears capable of learning progressively complex and abstract features at different scales.

An important part of the RESNET consists in using global average pooling without changing if any layer at the end of the network where each feature map is not a matrix but a number. This shows a graceful degradation of the network, as the number of its parameters is reduced dramatically, encoding a dense layer in opposite to completely connected ones. This final layer, a fully connected layer which performs a mapping of the pooled features to class scores for the task at hand might be image classification in this case, forms the network. After this conversion, a SoftMax layer takes the scores as input and converts them into a probability distribution over the classes.

RESNET was used on lots of sets and tasks and those include the classification topic too where the ImageNet dataset is taken as a base. It is open-source, and it exists in many versions of its product family: ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152, which are distinguished by the number of layers and hence, by computational complexity and performance. The selection of a RESNET specific variant is reliant on the linking task, as well as the available computational resources.

Data set Used

We used EMNIST dataset in this system. EMNIST is an extended version of MNIST dataset. MNIST (Modified National Institute of Standards and Technology) dataset contains only digits from 0 to 9. It does not contain any alphabets. EMNIST is an extension of MNIST which contains both alphabets and digits. Based on the survey done on EMNIST and MNIST [5], we came to know that EMNIST is a widely used dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

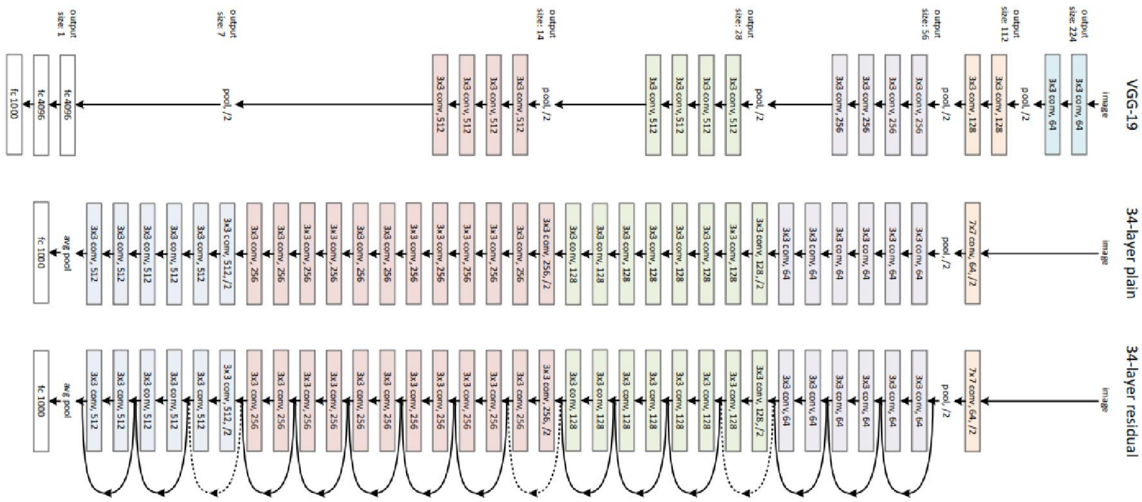

Fig 2 : RESNET CNN Architecture

The handwritten digits from the NIST Special Database 19 are collected in the EMNIST dataset, which is arranged in a format compatible with the MNIST dataset and can be accessed at https://www.kaggle.com/datasets/crawford/emnist. The handwritten letters in the EMNIST Letters dataset are organized into 26 classes, one for each letter of the alphabet. The 103,600 images in the dataset are divided into 14,800 test images and 88,800 training images.The 240,000 training and 40,000 test images in the EMNIST Digits dataset contain 10 classes of handwritten digits.

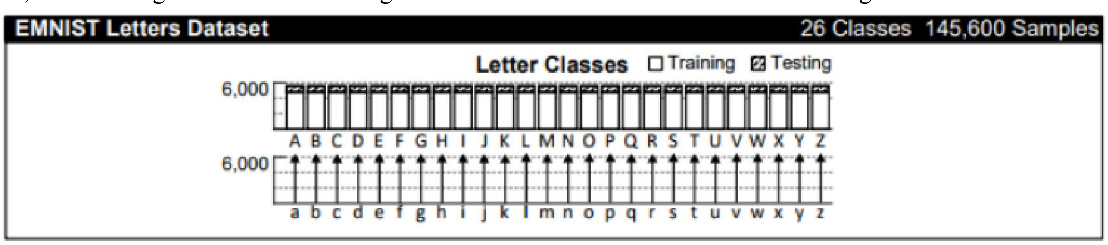

Fig 3: Letters Dataset

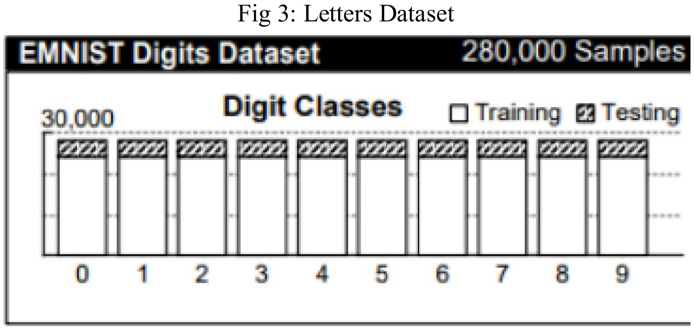

Fig 4: Digits Dataset

Case Study

Through this study case, we will have a long look at posture from the hand images. This analysis will contribute to computer vision areas such as image recognition and hand gesture recognition. These preprocesses will be beneficial for ensuring that all data is well prepared and that all necessary features are extracted for the next step of our research. Combining the one image example-prepared data will be drawn for followers' explanations. A stepwise procedure has been illustrated. With the approval of an external source, a hand posture image was secured. It is important to take the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

image to the needed size. To bring down image resolution, we changed the image width and height into a 6x6 pixels grid.

This resizing step was utilized because it had two main purposes. Firstly, it allowed for easier processing, and it also helped to decrease the computational complexity. To allow the image data to get processed and manipulated in the AI-powered application, the resized image was first converted to a NumPy array. To perform the subsequent analysis, the array.numPy was structured to provide for a number of operations and activities. Now the colorful resized image gets resized in grayscale. The acquainting with this grayscale presentation, we make up the intensity information. This grayscale image is also known as black and white, and it carries no color at all.

By doing so, these illustrations eliminate the level of complexity. It will make the number of resources used for computing substantially less. Now, proceed to the drawing of the features from the obtained representation. The process of feature extraction is performed by means of the Prewitt filter. The most famous feature of this filter was the filtering line that avoided blurring of the image and examination of the controlled gradient. The output of this process will be composed of those distinctive existing factors which will be used for further investigation.

Fig 5: Sample hand gesture

```
import cv2
import numpy as np
original_image = cv2.imread('/content/Three_finger_salute.jpg')
resized_image = cv2.resize(original_image, (6,6))
image_array = np.array(resized_image)
print("Resized Image as NumPy array:")
print(image_array)
print("grayscale")
grayscale_image = cv2.cvtColor(resized_image, cv2.COLOR_BGR2GRAY)
print(grayscale_image)
```

Fig 6: Code for Grayscale

```
grayscale
[[119 115 107 114 115 114]
 [122 115 100 102 113 116]
 [119 109 184 196 101 110]
 [113  97 176  83  93  98]
 [ 95  92  78 104  83  95]
 [ 88  85 166  79  80  93]]
```

Fig 7: Grayscale

Stage -1:

119	115	107	114	115	114
122	115	100	102	113	116
119	109	184	196	101	110
113	97	176	83	93	98
95	92	78	104	83	95
88	85	166	79	80	93

Input 6X6

×

+1	+1	+1
0	0	0
-1	-1	-1

Kernel 3X3

→

-71	-153	-145	-64
-49	-39	-37	57
147	215	216	125
47	26	27	155

Output 4X4

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

Stage -2:

-71	-153	-145	-64
-49	-39	-37	57
147	215	216	125
47	26	27	155

Input 4X4

×

+1	+1	+1
0	0	0
-1	-1	-1

Kernel 3X3

→

-947	-918
-225	-94

Output 2X2

Stage -3:

-947	-917
-225	-94

Input 2X2

MAXPOOL 1X2 →

-918
-94

→ Dense layer

Workflow

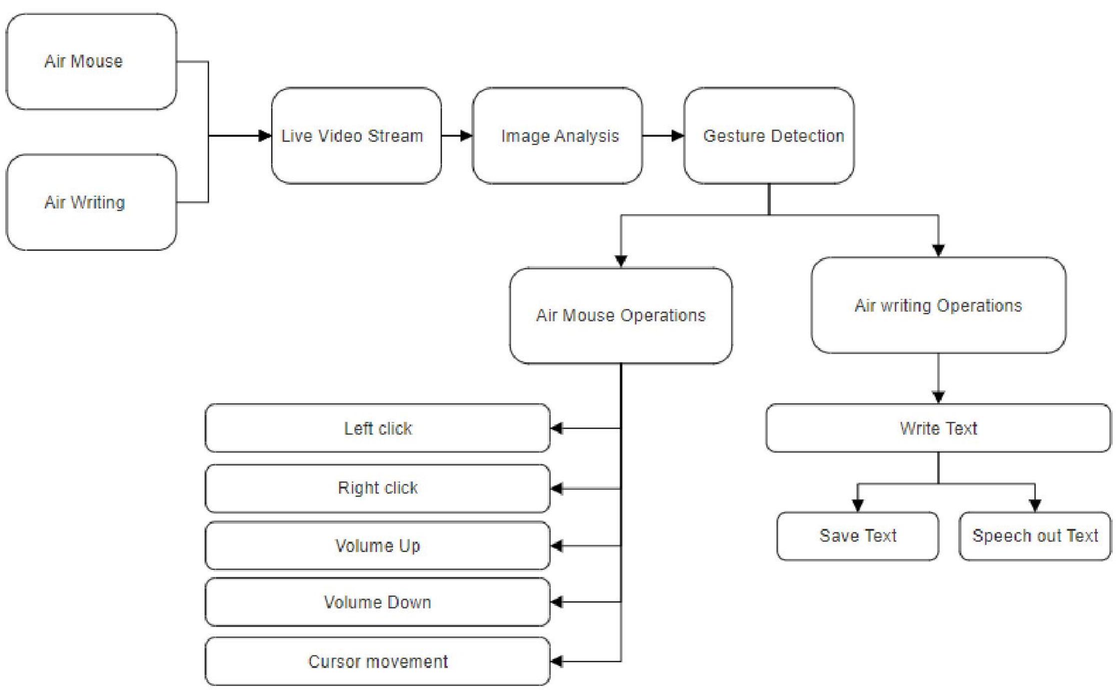

Fig 8: System Workflow

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

Implementation of the system is illustrated in the above figure. The given steps are depicted above and the structure is as mentioned:

1. Live video stream
2. Image analysis
3. Gesture detection
4. Air mouse operations
5. Air writing operations

Additionally, the next action should be set to either air writing or air mouse option. Later, the live video capture of the scene is done by using OpenCV library. There is a library that uses a web camera to record these video narrations. The user will be doing them himself/herself depending on their specific needs. Now the think that gesture will be examined from all sides. The signal which is provided by user shall be detected in gesture sensation step. Now, users can involve in air mouse operations as well as air writing. The air mouse process covers the cursor movements, left click, right click and volume increase and decrease. There is a gesture for every operation in humanize. This is how the air writing system works: we write, and it displays what we have written in the editor. The script can save the text in an acceptable format, so either .docx or .txt. This system employs speech feature that is utilized for transformation of the writing into verbal. By switching colour of virtual pens, we can spread our messages. This is the preserve of an air writing system.

VII. RESULTS

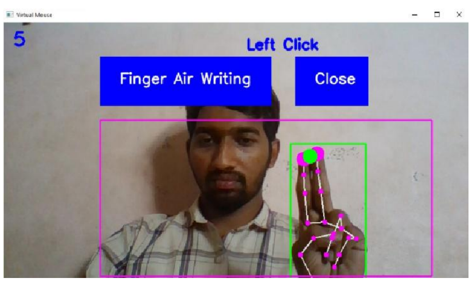

Fig. 1. Left click

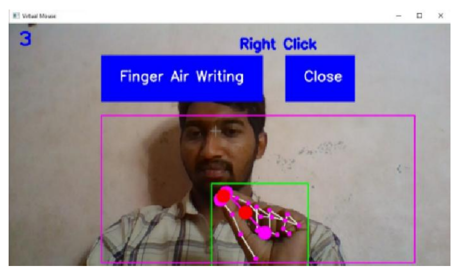

Fig. 2. Right click

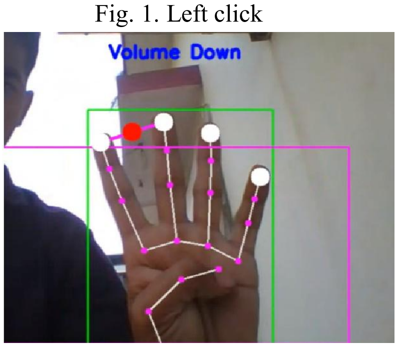

Fig. 3. Volume down

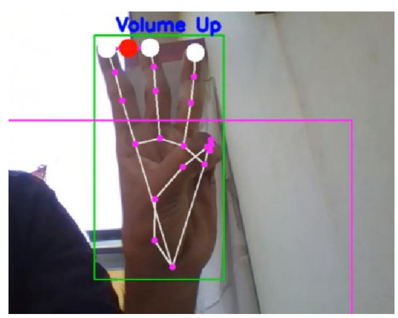

Fig. 4. Volume up

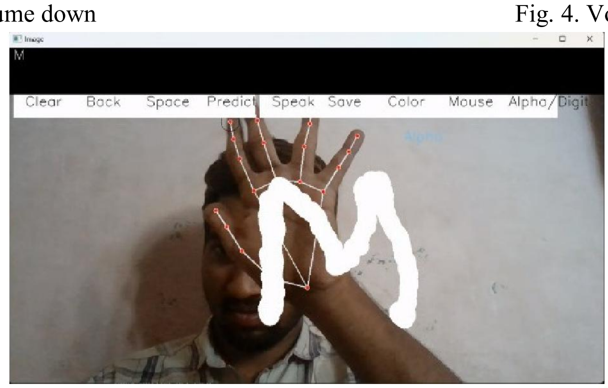

Fig. 5. Air writing

VIII. CONCLUSION

In this project, we used the RESNET CNN algorithm to get accurate results. We used the EMNIST dataset which contains both alphabet and digits images that will be compared with our air writing characters. This project implements techniques with the help of RESNET CNN algorithm and libraries like OpenCV, and Media Pipe to improve the accuracy and performance of the virtual mouse and air writing system. This project will improve the user experience.

REFERENCES

[1]. "Air-writing Recognition using reverse Time Ordered Stroke Context.", IEEE, 2017, Tsung HsienTsai and JunWeiHsieh. IEEE Journal of Information and Computational Science.

[2]. "Recognition of Handwritten Digit using CNN", IEEE, 2020, Peiyu Ma.IEEE Trans on systems, Man, and cybernetics.

[3]. "Air writing recognition Modelling and Recognition of Characters, words and Connecting Motions", IEEE, 2017, P Anjaneyulu, Y Jampaiah, R Karthik, T Vijetha. IEEE Trans. Pattern Analysis and Machine Intelligence.

[4]. "Hand-Mouse Interface Using Virtual Monitor Concept for Natural Interaction.", IEEE, 2017, CHANGHYUN JEON, OH-JIN KWON, DONGIL SHIN. IEEE Journal. Internet Compute. Service

[5]. "A Survey of Handwritten Character Recognition with MNIST and EMNIST" Alejandro Baldominos, Yago Saez and Pedro Isasi Published on MDPI: 4 August 2019

[6]. "Pattern Recognition - Recognition of Handwritten Document Using Convolutional Neural Networks.", IEEE, 2019, M.Rajalakshmi, P.Saranya,P. Shanmugavadivu. International Journal of Engineering and Computer Science.

[7]. "Virtual Mouse Using Object Tracking" Monali Shetty, Christina A Daniel, Manthan K Bhatkar, Ofrin P Lopes Journal Name: IEEE Xplore, Year of Publication: 2020

[8]. "Air-Writing Recognition Based on Deep Convolutional Neural Networks.", IEEE, 2021, chaur-heh hsieh, you-shen lo, jen-yang chen 3, and sheng-kai tang. IEEE Trans.Human-Mach.

[9]. "Deep Learning-Based Real-Time AI Virtual Mouse System Using Computer Vision to Avoid COVID-19 Spread." S. Shriram et al. [2021] Hindawi, Journal of Healthcare Engineering.

[10]. "Virtual Canvas for Interactive Learning using OpenCV" Palak Rai, Reeya Gupta, Vinicia Dsouza, Dipti Jadhav IEEE 3rd Global Conference for Advancement in Technology (GCAT) - 2022