RESEARCH ARTICLE OPEN ACCESS

Air Writing Recognition System through Deep Learning

Indrabhan Kurkute[1], Ghorpade Rutuja[2], Gujar Shivani [3], Murdare Shreya[4],Tambe Gayatri[5]
*1 Asst.prof. Department of Computer Technology, Padmashri Dr. Vitthalrao Vikhe Patil Institute of Technology & Engineering(Polytechnic), Loni, Maharashtra, India
*2,3,4,5 Department of Computer Technology, Padmashri Dr. Vitthalrao Vikhe Patil Institute of Technology & Engineering(Polytechnic), Loni, Maharashtra, India

--- ************************ --

Abstract:

A person can engage in air writing by moving their body while they write letters or words on a blank surface. An example of gesture recognition is "air-writing recognition," which involves turning hand gestures into invisible handwriting. You don't need to commit certain gesture patterns to memory in order to practice air-writing, unlike with general motions. It focuses instead on the language and subject matter that the user is interested in. With the advent of smart bands, the need for an extra device with sensor(s) is no longer an issue with conventional air-writing. Consequently, algorithms for air-writing recognition are becoming increasingly flexible on a daily basis. However, signal length variability is one of the key challenges in developing an air-writing recognition model. The uneven signal length is made clear throughout this writing and data-recording process. To make sure the signals were always long, researchers used techniques like padding and truncating, but these approaches led to massive data loss. To ensure minimal data loss for time-series signals, a statistical technique called interpolation can be employed. After extensively investigating different interpolation methods on seven publicly available air-writing datasets, we utilized a 2D-CNN model to identify air-written characters in this study. All datasets showed that our solution outperformed the state-of-the-art solutions in terms of user-dependent and user-independent principles.

Keywords**:** Air-writing recognition, Sensor, Air-writing recognition algorithms, Time-series signals, Statistical technique, 2D-CNN model.

--- ************************ --

I. INTRODUCTION

Our habits of engaging with the digital realm have been steadily expanding over the past decade. Internet access via touchscreens and other electronic gadgets is widespread. The extra effort required to carry and remove physical gadgets like smartphones from one's pocket for use is a major drawback of their widespread adoption. Eliminating the need for intermediary physical devices, such as smartphones, is the main objective of next-generation technologies. It would appear that the future phase of this technology will be dominated by virtual and augmented reality, with the output being projected directly into the user's eyes using specialized glasses. Many people believe that speech recognition is the most natural and intuitive method to communicate with technology and it has been the subject of much research. Nevertheless, there are still some requirements for speech recognition to meet before it can effectively interface with technologies. An increasingly popular means of interaction for future technology, gesture recognition has been the subject of much research and development in recent years. Emerging as new mediums of engagement using gestures instead of conventional keyboards, touchpads, or other pressing and touching instruments are new technologies based on acceleration sensors, photo sensors, electromagnetic and audio impulses, and cameras. Specifically, when it comes to writing methods, conventional touch sensors do not work with VR, AR, and gesture-based technology. In order to fulfill the need of next-generation touchless writing methods, air-writing shows great promise. The term "air-writing" describes the practice of free-form letter or word writing using the fingers or hand motions in an open area. Considering that a gesture is any kind of predetermined movement in the air, it can be seen as a subset of gestures. But it's not easy to recognize air-writing characters. The fine-grained movement of characters sets them apart from general movements, and different people can write them in different ways. Traditional writing methods resemble pen and paper and use many strokes to represent the alphabet and digits. When compared to the traditional writing system, air-writing is distinct. When writing in the air, the motions produced by raising a pen are much less apparent. Without the tactile experience of holding a pen or sheet of paper, users are unable to discern the writing as they would with a traditional pen and paper. Users risk disorientation and a loss of writing orientation due to this. Despite these challenges, there have been encouraging developments in the field recently. A growing number of smartphones now come equipped with built-in sensors. Making a mobile app to gather data from sensors allows users to tailor the data collection procedure to their own needs. Thus, the realm of smartphone-generated sensor datasets is constantly expanding. Typically, for gesture recognition to work, the user must hold or wear motion sensors like an accelerometer and a gyroscope as they make their gestures. However, the unpredictability of signal duration is the greatest obstacle to developing a model for air-writing identification using sensor data. Signals of varying lengths are inherent to data recording and writing processes, yet convolutional neural networks and other deep learning approaches rely on signals of constant length for training and prediction. A common method for padding or discarding values from the beginning and/or end of a signal is fixed-length padding and truncation. Although it's easy to implement, the strategy leads to massive data loss, which means the model misses out on important aspects. Interpolation, in contrast, is a statistical method for estimating future values from existing ones. Without losing any information, it maps the signal data into a specified fixed length, guaranteeing that the signals will remain at a constant length. Hence, there is less data loss. The field of image processing has produced a number of well-studied interpolation methods. We used one-dimensional data instead of two-dimensional data and studied those different interpolation methods on time-series data to get the signals with fixed lengths. Better prediction accuracy and an easier training procedure are two benefits of a well-structured, fine-tuned deep learning model. We achieved state-of-The-art accuracy on seven publically available datasets by designing a 2D-CNN model based on the best approaches proven in the literature.

[1] Akash Kumar M. B. et al. introduce a video-based calculator that lets users type English letter sets and numbers using the author's camera while in the air. To begin with, it follows the finger in the video frames, and then it uses convolutional neural networks (CNNs) over plotted images to understand the written letters. be that as it may, this work provides a distinctive human framework link that eliminates the necessity of a console, mouse, etc., for character production. All you need is a webcam and a flexible camera to rearrange your fingers. The author has constructed this project by combining OpenCV with the Python language. [2] in It is predicted by Fuad Al Abir et al. that air-writing recognition will play a crucial role in the globe following the fourth industrial revolution. Researchers in this work used time-series data to create an algorithm that can identify individual letters and numbers in the English alphabet. Data preparation for deep learning methods using sensors with little loss of data is a tough challenge. Several interpolation methods, commonly employed for pictures but seldom considered for time-series signals, were thoroughly investigated by the author. Using the Bicubic interpolation algorithm to interpolate the raw data yields the greatest results in our use case, according to the author's testing. The author used this extrapolated data to train a 2D convolutional neural network (CNN) model that could distinguish between different letter types; this model significantly outperformed the benchmark algorithms. Before the air-writing recognition system can be employed in the real world, it needs to be fine-tuned. Additionally, the author has the option to combine user-dependent and user-

independent methods, provide the user with feedback or advice, and incorporate an auto-correction system.

[3] The sequence gestures need to be identify for Air writing can be inspired by the work of Shubham gade et al, is the result of an innovative and comprehensive integration of the Hungarian job allocation system with the H-net neural network. At this point in the procedure, author will make use of the sequence list in order to enhance the results of the slot allocation. It is possible to make real-time adjustments to the scheduling strategy that is utilized in order to discover the nearest charging station for electric vehicles with the assistance of this crucial data. Using the decision-tree method, the sorted list that was obtained in this stage of the choice tree process is used to make the decision to check the available slots for the electric vehicle charging station that is located closest to the user. Short-term and long-term applications of these technologies include the scheduling of air writing gestures in the good speed.

II. LITERATURE SURVEY

[4] The work that expanded a traditional machine learning introductory project—handwriting analysis—to new interfaces so that it may be utilized in multiple ways is discussed in by M. Saranya et al. The author considered that their method offered a more user-friendly interface, therefore they took on the challenge of applying a new approach to the misperception problem. While there is room for improvement, the author hopes that their modest contribution highlights the subtleties of various approaches to computer vision and machine learning problems and prompts other researchers to consider users' needs when developing new technologies.

[5] The authors Ayush Tripathi et al. provide a system for air writing identification that uses deep learning-based models to decipher the alphabets written on time-series data collected from a wrist-mounted Inertial Measurement Unit (IMU). Two sets of three-channel images are formed from the signals captured by the three-axis accelerometer and gyroscope in the IMU. These images are encoded using several approaches, including MTF, Self Similarity Matrix (SSM), and Gramian Angular Field (GAF). Letter predictions are then derived by averaging the class conditional probabilities derived from the two independent classification models. Various ResNet, DenseNet, VGGNet, AlexNet, and GoogleNet variations, among others, are used as standard model architectures for picture categorization. The effectiveness of the suggested approach is proven by experiments conducted on two datasets that are publicly accessible. You will be able to find our implementation code at this address: https://github.com/ayushayt/ImAiR.

[6] Have you ever wished you could draw anything by just waving your finger? That's what Vijay Kamble et al. describe. Dissemination of ideas and information by verbal, written, or visual means is what we mean when we talk about communication. An innovative method for controlling computers just by waving your fingertips has been proposed by the author. There are a plethora of modern technologies that can accomplish this. Using the media pipe module, the author suggests detecting fingertips in the air in front of the camera in real-time. Author recommends RNN model with Bezier curve for handwriting pattern recognition. Here are the outcomes that can be expected from the proposed method: The system will begin by recognizing the hand and finger in the video frame. Then, it will analyze the hand's pattern and translate it into text. Also, this approach removes the need for any type of character input device, allowing for true Human-System connection.

[7] Using the FMCW radar sensor, Shahzad Ahmed et al. presented a novel approach to in-air digit recognition. It was suggested to use a multi-stream convolutional neural network (CNN) model that could extract data from range-time, Doppler-time, and angle-time patterns. Overall, the MS-CNN model outperforms the traditional CNN methods because it concatenates features at a later stage after combining them from several input streams. Twelve individuals in twelve different locations had their data collected in an effort to increase diversity and decrease prejudice. The first experiments demonstrated a very high classification accuracy of 94.20% when it came to identifying all ten base

digits. Conventional CNNs that use range-time and Doppler patterns are much less accurate than MS-CNNs. Aside from in-air digit writing, the technology can be applied to various gesture detection challenges thanks to MS-CNN's excellent accuracy. A human hand's pliability causes the micro-Doppler effect, which is a slew of extra vibrations introduced by the hand's fingers, palm, and other soft tissues. The CNN's latency varied between 400 and 500 ms, which imposed a strain on computing, but the classification accuracy was good enough to be useful even when testing in a variety of physical settings. This paper proposes an in-air writing system that can classify a single digit simultaneously. The suggested technique appears to handle each digit gesture independently. Nobody has looked into the possibility of recognizing continuous digit writing just yet.

[8] According to Siddhi Bhalerao and colleagues, that is. One natural method to control a computer is with radar, which can pick up on hand movements even when no contact is made. "Air-writing" refers to the practice of making hand motions while writing. The authors have developed a millimeter wave radar-based air-writing system. There are two stages to the author's method. The first is to locate your hand and follow its movements. The authors then put this data to use in two ways: first, by training a specific kind of neural network to decipher the hand's trajectory and identify the characters it draws; second, by transforming the trajectory into an image and training yet another neural network to decipher the letters it draws. In character recognition, the first method is just as effective as the second, with an accuracy rate of 98.33%. Using actual data collected at 60 GHz by three radar sensors, the author verified this.

[9] The WiTA task requires continuous air writing recognition, which XUHANG TAN et al. successfully handles using an oval end-to-end air-writing recognition algorithm based on the transformer model. Author builds end-to-end Air-Writing model using transformer model, which effectively converts video frame sequences from Air-Writing into character sequences. The author's model learns the mapping relationship between authoren frame sequences and characters by the use of an attention mechanism within an encoder-decoder framework, enabling automatic handwriting character recognition. The character error rate (CER) is used to evaluate the model's performance, whereas D-fps is used to ensure particular real-time performance. At 29.86%, this study's CER value is the highest of any data yet reported. In addition, the algorithm's real-time performance is confirmed by its D-fps result of 186.75 fps. Experiments comparing Author's model to other end-to-end methods show that it improves overall accuracy by producing better CER results while keeping a certain level of real-time performance.

[10] A CNN-BiLSTM–based hybrid deep learning model for character recognition was proposed by TaikiWatanabe et al. Graffiti characters, which are written with a single stroke, were employed by the author. In order to identify characters, the author used both alphabetic and numeric characters independently. Before beginning character recognition, the author created two datasets: one containing images and another containing padded sequential data. One model was trained using CNN data, and the other used BiLSTM for padding sequential data. Also, to boost recognition accuracy, the author mixed the CNN and Bi-LSTM models. A hybrid deep learning model was the name given to this integrated model in this study. Our suggested solution outperformed previous research with experimental findings showing a recognition accuracy of 99.3% on the alphabetic letter dataset and 99.5% on the digit character dataset.

[11] According to Logeswari.N et al., Air Writing Recognition is a cutting-edge method of interacting between computers and humans. In order to analyze hand movements and transform them into digital text images, it is helpful to be able to recognize gesture based writing in Air. This allows you to avoid using a touchpad or screen. An enhanced Air Writing Recognition system that utilizes a wrist-worn smart band is suggested by the methodology. The Bluetooth module allows for the wireless transfer of data acquired by the smart band. With the help of the collected data, three ML algorithms were trained: RF, KNN, and GBM. Several metrics were used to compare the machine learning model's performance, including recall, sensitivity, specificity, and accuracy. Compared to the other two algorithms, the KNN model outperforms them in terms of numerical accuracy. Based on the results

of the simulations, the KNN model outperforms the RF model by 49% and the GBM model by 27.57% in terms of accuracy.

[12] In their description of the Input via text is highly desirable in many XR use scenarios, but is especially important for knowledge and office tasks. This article examines the use of physically aligned and mid-air surfaces to facilitate the entry of handwritten text in Virtual Reality (VR) and Video See-Through Augmented Reality (VST AR), specifically comparing the two for the purpose of writing both basic and complex sentences. A total of 72 participants were involved in the 2x2x2 experimental design. Over the course of two 10-minute sessions, they were asked to write ten basic sentences and ten complex sentences, which were designed to mimic text input found in real-world circumstances. Our digital ink-based handwriting recognition program works with a variety of XR displays and surface alignments. The author took into account factors including task load, handwriting style, text input performance, usability, and user experience. Excellent usability and a smooth transition from the real world to the digital one are the outcomes of the author's research. There was no correlation between the speed and accuracy of text input and XR displays or surface alignments. Simple sentences (17.85 WPM, 0.51% MSD ER) resulted in faster input speeds and fewer errors than complicated sentences (15.07 WPM, 1.74% MSD ER), but sentence complexity did have an effect.

[13] To address the many problems that hostel administrators encounter when trying to keep track of student approvals, S. Gobi et al. developed the Gate Pass Management System. Using quick response (QR) codes and an intuitive web app, the method improves security, makes things easier to understand, and encourages responsibility and openness. The system guarantees effective and efficient handling of hostel gate pass requests by integrating numerous features adapted to the demands of different stakeholders, such as students, administrators, wardens, and security professionals. In the future, this system's implementation could completely transform how hostels are managed, leading to more efficiency and an improved student experience.

[14] A multi-stream convolutional neural network (CNN) model that efficiently utilizes information from range-time, Doppler-time, and angle-time patterns is presented by Amith K. R. et al. as an application of in-air digit identification employing FMCW radar sensor technology. In comparison to more conventional CNN methods, the MS-CNN model outperforms them by integrating characteristics from numerous input streams. Classification accuracy for identifying the 10 basic digits is 94.20%, according to experimental data performed with various participants in different physical settings.

III. METHODOLOGY

The proposed model for the identification of Air Writing is explained in detail with below mentioned steps.

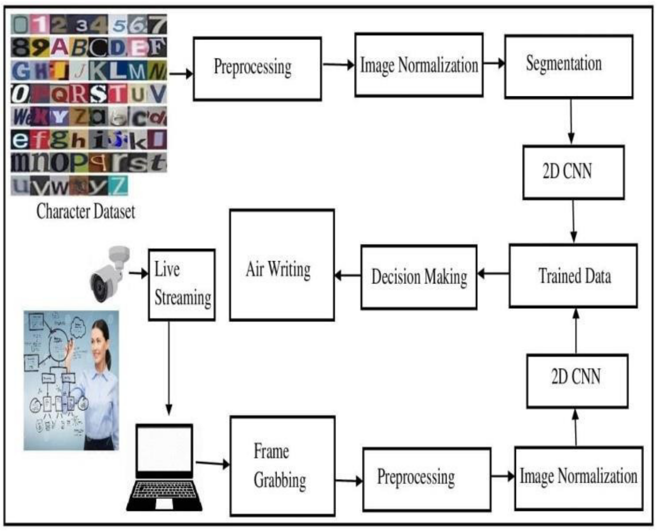

Figure 1 System Overview

Step 1: Dataset preparation:

This is the initial step of the proposed system where "mnist" character dataset is being downloaded to train the model using 2D CNN. This "mnist" dataset is being stored in the object form to label them the same. This dataset is consists of the alphabets from a to z, A to Z and digits from 0 to 9 of English language.

Step 2: pre-processing -

Here in this step every images is being considered to convert it into a grayscale image to reshape it in a dimension of 28 X 28. A rescaling factor with 1: 255 is set to train the whole character dataset as mentioned in the next step of the proposed system.

Step 3: Training with 2D Convolution neural network (2D CNN)-

To train the obtained images a 2D Convolution neural network is being employed using the ketas and tensor flow libraries in python.

For the purpose of deployment a sequential neural network model is selected for the 2 layers of neurons. In the first layer a 32 kernels are set to the size 3×3 along with the activation function Relu for the set dimension and the color channel of 3. Followed by the first layer second layer is also set the same dimension as of first layer with 64 kernels each of size 3 X3 with activation function Relu. A max pooling layer is added after the second layer with the dimension size of 3 X3 followed by a dropout layer of 50%.

LAYER	ACTIVATION
conv 2D 32 (3X3)	Relu
conv 2D 64 (3X3)	Relu
MaxPooling 2D 3X3	
DropOut 0.5	
Flatten	
Dense 250	Sigmoid
Dense 10	Softmax
Optimizer	Adadelta
batch_size	500

Figure 2: 2D CNN Architecture

The training of the mist dataset accuracy and loss can be seen in the below figure 3.

After the 2 layers a flatten layer is stopping the process of training with a dense layer of size 250 and with an activation function called Sigmoid to be followed with a Dense layer of size of 10 with an activation function of Softmax is being employed. Then finally the data is gathered using another dense layer of " categorical " size along with the activation function softmax. An adadelta optimizer is used to optimize the precision of neuron values, this finally yields a trained data to store in a file with an extension of. H5. The Whole learning process through 2D CNN is shown in the below mentioned architectural table.

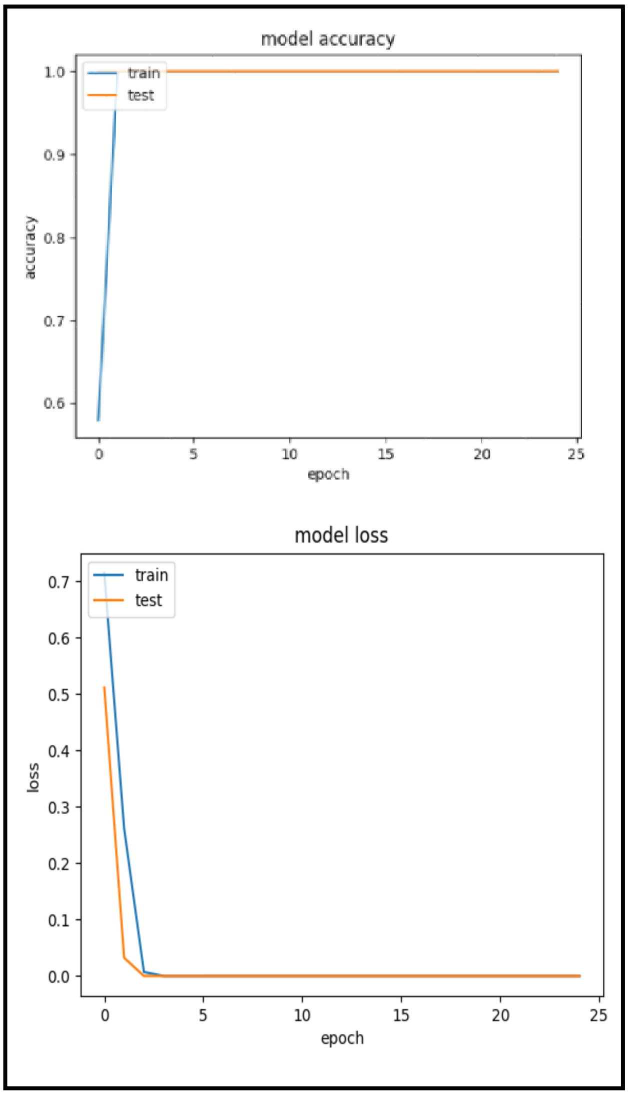

Figure 3: Accuracy and loss for CNN

IV. RESULTS AND DISCUSSION

The proposed method for Air Writing Recognition System through Deep Learning. was developed using the Anaconda framework, Python, and the Spyder IDE. The development computer has 1 terabyte of secondary memory and 8 gigabytes of main RAM. A number of factors have been considered in order to determine how feasible the proposed plan is. In this part, we detail the results of the experimental study.

The obtained results for confusion matrix are depicted below in the following figures.

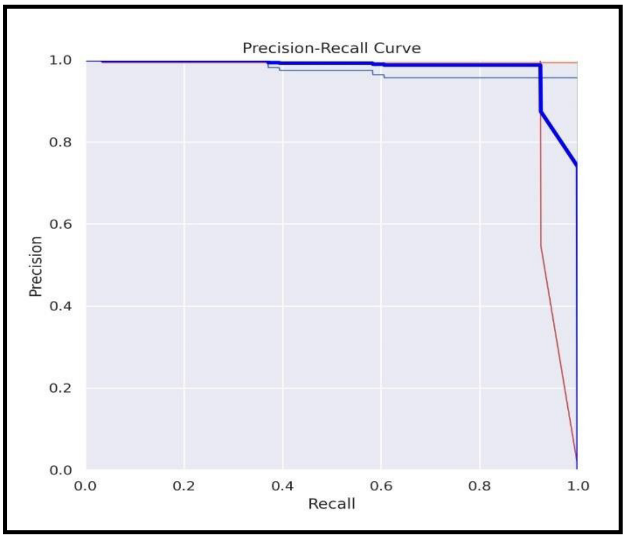

Figure 2: Precision-Recall Curve

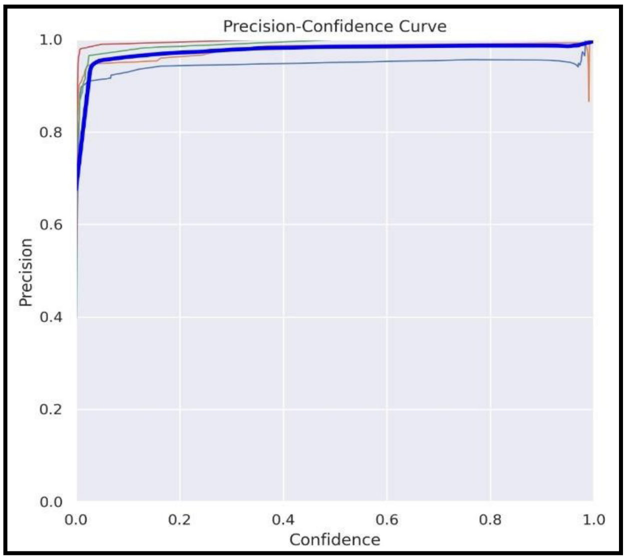

Figure 3:Precision-Confidance Curve

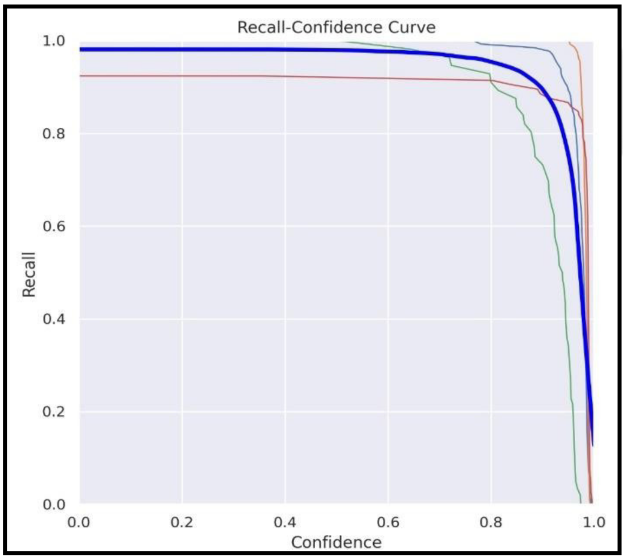

Figure 4: Recall-Confidence Curve

V. CONCLUSION AND FUTURE SCOPE

After the fourth industrial revolution, the ability to read air writing will be essential. Here, we construct a system that uses time-series data to differentiate between the English alphabet's letters and numbers. Minimizing data loss while preparing sensor data for deep learning techniques is a challenging topic. We have investigated in depth a large number of interpolation methods that are commonly used for pictures but often ignored when dealing with time-series images. In our use case, we found that the Bacubic correlation approach of interpolating the raw data produced the best results. Our suggested 2D-CNN model was trained using this interpolated data to classify the letters with a substantial improvement over the state-of-the-art methods. The air-writing recognition system needs to be fine-tuned before it can be used worldwide. In addition, we can mix user-dependent and user-independent methods, give the user instructions or feedback, and incorporate an auto-correction system. We can also explore the subjects' attributes in further detail. We intended to make the procedure universally applicable, thus we overlooked the fact that the datasets indicated some of the participants' features. Last but not least, compared to word recognition in isolation letter or digit situations, word recognition in similar circumstances necessitates further research. All of these issues must be resolved before an air-writing recognition system can be built. Along that path, we provide our work as a springboard.

Possible Future Developments: ¬ Implementation of the system in mobile apps ¬ Support for real-time operations in offices and other locations.

VI. REFERENCES

[1] AkashKumar M B, Ashwini R, Dhanaraj S, Shauthortha B K, Dr.V Keerthika, "Computer Vision Based Air Writing Recognition," Vol-7 Issue-4 2021, IJARIIE-ISSN(O)-2395-4396.

[2] Abir, F.A.; Siam, M.A.; Sayeed, A.; Hasan, M.A.M.; Shin, J. Deep Learning Based Air-Writing Recognition with the Choice of Proper Interpolation Technique. Sensors 2021, 21, 8407. https:// doi.org/10.3390/s21248407.

[3] Shubham Gade, Amita Singh, Shubham Sarote, " Efficient H-net model-based slot assignment solution to accelerate the EV charging station searching process ",ISSN: 2349-6002,Volume:11 ,Issue: 6,PageNo: 2590-2597

[4] M. Saranya, Kariketi Tharun Reddy, Madhumitha Raju, Manoj Kutala, "Air-Writing Alphanumeric Character Recognition," 2022 IJCRT, volume 10, Issue 7 July 2022, ISSN: 2320-2882.

[5] Ayush Tripathi, Arnab Kumar Mondal, Lalan Kumar, Prathosh A.P., "ImAiR: Airwriting Recognition framework using Image Representation of IMU Signals," DOI 10.1109,LSENS.2022.3206307.

[6] Vijay Kamble, Tanvi Ghule, Anuja Kotkar, Vinayak Kulkarni, "Air Writing and Recognition System," Volume 9, Issue 11, www.jetir.org (ISSN-2349-5162).

[7] Shahzad Ahmed, Wancheol Kim, Junbyung Park, Sung Ho Cho, "Radar-Based Air-Writing Gesture Recognition Using a Novel Multistream CNN Approach," IEEE INTERNET OF THINGS JOURNAL, vol. 9, no. 23, 1 December 2022

[8] Siddhi Bhalerao, Dr. Meenakshi Thalor, Aissms Ioit, "Character Recognition In Air-Writing Based On Network Of Radars For Human-Machine Interface," Vol. 1, No. 1, 2023 : 64-72, DOI: https://doi.org/10.59890/ijaamr.v1i2.345, https://journal.multitechpublisher.com/index.php/ijaamr

[9] X. Tan, J. Tong, T. Matsumaru, V. Dutta and X. He, "An End-to-End Air Writing Recognition Method Based on Transformer," in IEEE Access, vol. 11, pp. 109885-109898,2023,doi: 10.1109/ACCESS.2023.3321807.

[10] Watanabe, T.; Maniruzzaman, M.; Hasan, M.A.M.; Lee, H.-S.; Jang, S.-W.; Shin, J. 2D Camera-Based Air-Writing Recognition Using Hand Pose Estimation and Hybrid Deep Learning Model. Electronics 2023, 12, 995. https://doi.org/10.3390/ electronics12040995.

[11] Logeswari.N, Amutha.R, "Air Writing Recognition using Machine Learning Algorithms," Journal of University of Shanghai for Science and Technology, Volume 26, Issue 4, ISSN: 1007-6735.

[12] Florian Kern, Jonathan Tschanter , MarcErich Latoschik, "Handwriting for Text Input and the Impact of XR Displays, Surface Alignments, and Sentence Complexities," IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, Vol. 30, No. 5, May 2024

[13] S. Gobi, V. Pasumpon Arivarasan , S. Sivaraman, T. Aadhavan, Mrs. A. Dhivya, M.E, " Air-Writing Recognition in Tamil Characters Using Inception V3," International Journal of Research Publication and Reviews, Vol 5, no 5, pp 12255-12259 May 2024, www.ijrpr.com ISSN 2582-7421

[14] Amith K R, Nikhil Holla R, Prashanth J, "Air Writing Detection and Recognition," Volume 4, Issue 1, International Journal of Advanced Research in Science, Communication and Technology, ISSN (Online) 2581-9429, DOI: 10.48175/IJARSCT-15381, ISSN (Online) 2581-9429.