Article

A Novel Machine Learning Based Two-Way Communication System for Deaf and Mute

Muhammad Imran Saleem [1,2,*] , Atif Siddiqui [3] , Shaheena Noor [4] , Miguel-Angel Luque-Nieto [1,2] and Pablo Otero [1,2]

1   Telecommunications Engineering School, University of Malaga, 29010 Malaga, Spain
2   Institute of Oceanic Engineering Research, University of Malaga, 29010 Malaga, Spain
3   Airbus Defence and Space, UK
4   Department of Computer Engineering, Faculty of Engineering, Sir Syed University of Engineering and Technology, Karachi 75300, Pakistan
*   Correspondence: imran@uma.es

Abstract: Deaf and mute people are an integral part of society, and it is particularly important to provide them with a platform to be able to communicate without the need for any training or learning. These people rely on sign language, but for effective communication, it is expected that others can understand sign language. Learning sign language is a challenge for those with no impairment. Another challenge is to have a system in which hand gestures of different languages are supported. In this manuscript, a system is presented that provides communication between deaf and mute (DnM) and non-deaf and mute (NDnM). The hand gestures of DnM people are acquired and processed using deep learning, and multiple language support is achieved using supervised machine learning. The NDnM people are provided with an audio interface where the hand gestures are converted into speech and generated through the sound card interface of the computer. Speech from NDnM people is acquired using microphone input and converted into text. The system is easy to use and low cost. The system is modular and can be enhanced by adding data to support more languages in the future. A supervised machine learning dataset is defined and created that provides automated multi-language communication between the DnM and NDnM people. It is expected that this system will support DnM people in communicating effectively with others and restoring a feeling of normalcy in their daily lives. The hand gesture detection accuracy of the system is more than 90% for most, while for certain scenarios, this is between 80% and 90% due to variations in hand gestures between DnM people. The system is validated and evaluated using a series of experiments.

Keywords: deaf-mute person; deep learning; hand gesture recognition; motion controller; speech to text; supervised machine learning

1. Introduction

Technology has become an integral part of our lives. This has a positive impact on society, and people benefit from it. This research is focused on helping people with impairments. Artificial intelligence can support people with impairments to integrate into society effectively. The World Health Organization (WHO) reports that 5% of the world's population is deaf and mute (DnM) [1].

1.1. Sign Language

The method of communication between DnM and others is through sign language, which can only be effective when individuals from both sides understand sign language. This research is a step toward presenting a solution that will improve this communication. One of the main problems highlighted is the lack of understanding of sign language by non-deaf and mute (NDnM) people.

Alternate solutions are considered where one is to use writing for communication. This is an alternative method of communication, but it is thought to be a sluggish and inefficient method [1]. Another solution is to use a sign language interpreter, which is realistic yet expensive, and availability is a bottleneck. This option is also not available to the majority of the DnM community. Due to these and similar issues, DnM people find it hard to communicate with others and are thus unable to integrate into society. This means that society is unable to benefit from its contribution.

Having a smart model that allows communication between DnM and NDnM people will provide a breakthrough in tackling this problem and will have a huge impact on society due to the integration of more people and their contributions. The research work presented in this manuscript takes into consideration the work carried out in [2]. A two-way smart system for communicating between DnM and NDnM people is presented. There are two major contributions in this work, which are (a) Leap Motion [3] for gesture recognition, and (b) an Android application. The Leap Motion device is used to recognize hand gestures, which are in use in Pakistani Sign Language (PSL) courses.

1.2. Artificial Intelligence

The use of artificial intelligence (AI) is increasing, and researchers are solving various problems through its application. This research work also focuses on AI and how it can be used to provide solutions to DnM people.

1.2.1. Machine Learning

In this research, supervised machine learning is applied, and a dataset is created to provide a new feature for DnM people so that they can communicate in more than one language. Some machine learning algorithms are support vector, linear and logistic regression, and support vector.

1.2.2. Deep Learning

The detection of hand gestures and processing of these data are the most important options required for any system that supports DnM. For this research, artificial neural networks (ANN) and convolutional neural networks (CNN) are reviewed.

1.2.3. Artificial Neural Network

ANN is used to tackle a wide variety of computer vision problems, including classification, in today's world. Researchers who are tasked with collecting, analyzing, and interpreting vast amounts of data will find that ANN are very beneficial to their work. In addition, ANN increases processing speed and reduces complicity. The application of AI can be seen in areas such as image processing, natural language processing, intelligent robots, knowledge representation, and automatic reasoning. Some examples of algorithms based on neural networks are multilayer perception (MLP) and Boltzmann neural networks.

1.2.4. Convolutional Neural Network

Other deep learning-based algorithms include CNN, recurrent neural networks (RNN), generative adversarial networks (GAN), and deep belief networks (DBN). One of the important aspects of this research is to process image data, i.e., hand gestures from DnM. For processing image data, CNN is used.

Figure 1 shows the scope of this research within the artificial intelligence (AI) domain. The two solutions are implemented using supervised machine learning and CNN.
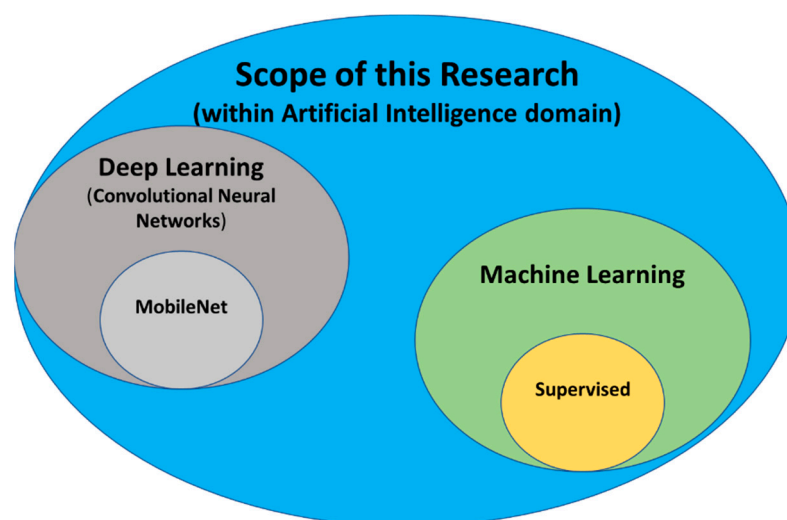
Figure 1. Scope of this research within the artificial intelligence domain.

1.3. Manuscript Organization

A summary of the rest of the manuscript is presented in this section. An extensive literature review is carried out for this research work, which is in Section 2. Section 3 presents the research methodology used. The proposed system for DnM people to be able to communicate with NDnM people is presented in Section 4. The experimental setup and results are presented in Section 5. Section 6 is the discussion section. Finally, Section 7 presents the conclusions.

2. Literature Review

In this section, the research work carried out by the authors is reviewed. A fundamental ANN model is presented in [4] and is shown in Figure 1. The structure has one input payer that has an input feature vector of length 'n'. At the output, there is also one layer with output variables of length 'm'. The input features are processed through to the output layer via a hidden layer of neurons of length 'k'. The raw data, i.e., the data to be processed, is entered through the input layer. This layer only provides an interface for the data that is required to be processed. The processing was then carried out using the hidden layer. The hidden layer is the key to this network, where all computational processing of the characteristics acquired from the input layer is done. The processed data are then forwarded to the end user via the output layer. The data sent to the output layer are the solution to the problem interfaced from the input layer. An application of the Mobile Net is presented in [5]. This is a streamlined architecture that comprises lightweight CNNs that use depth-wise separable convolutions.

The first manuscript [6] presents different recognition systems for sign languages. The authors in [7] presented a communication framework for DnM people. They used an automated Sign Language Interpreter (SLI), where they acquired data using motion sensors fitted to a glove. The sensor data are acquired using an Arduino board. The acquired data are then processed using a machine-learning technique. The system achieves an accuracy rate of 93%. In [8], a speech recognition system is presented that identifies a registered speaker's voice. This is a computer-based implementation in which language patterns are stored. The information was processed and acquired using a questionnaire created by the authors in [9]. They collected data from various organizations set up for DnM people. They concluded that the average age when hearing loss is discovered is 2.8 years, while a hearing aid is normally used from 7.6 years of age.

A system is presented in [10], which is referred to as AWAAZ. DnM people can use this system, where their gestures are acquired using image processing. The main components of the gesture recognition system are image acquisition, segmentation, morphological erosion, and feature extraction. In [11], a system that supports two-way communication between

DnM and NDnM people is presented that uses an automatic speech recognition technique and a mobile application based on visualization. The authors developed an Android-based application in [12]. This uses the Principal Component Analysis (PCA) approach and the Video Relay Service (VRS). The VRS works as a manual interpreter, converting speech to hand signals and vice versa. The hand gesture is recognized by the PCA algorithm.

A web cam-based application is developed in [13], where the images are first acquired using a webcam, then PCA is applied to extract features and the characters are recognized using training sets. The authors surveyed [14] and published the issues faced by DnM children. They concluded that it is possible to apply a common solution due to various parameters. Some solutions are presented in this paper. They also concluded that Hand-Gesture-Recognition (HGR) systems have contributed significantly to a shift in the way people interact with computers.

There is a lot of focus on improving sensor technology, which is expected to have a significant impact on how DnM people can communicate. The Leap-Motion-Controller (LMC) [15] is one such device. Despite the success of developing cutting-edge algorithms, there are still limitations because these algorithms have not focused on how to quickly process sequential hand gesture data and are unable to characterize the discriminative representation of different classes of hand gestures. The LMC sensor acquired hand motions and hand motion data where the pattern is sorted using a novel Chronological-Pattern Indexing (CPI) method. This approach interprets the data as a series. Hand gesture recognition systems have recently been demonstrated to offer substantial promise for use in the realm of digital entertainment. These enhancements have been made possible because of recent advancements in machine learning and sensor upgrades. Contactless hand gestures on a Leap Motion device are used to identify dynamic hand motions. This methodology is within the scope of this research study; hence, it has been included and reviewed in detail.

In [16], the authors used recurrent neural networks in combination with Long Short-Term Memory (LSTM) to analyze sequential time series data acquired through the Leap Motion device for gesture recognition. The authors used both the normal unidirectional LSTM and the bidirectional LSTM. A prediction network architecture, identified as the Hybrid Bidirectional Unidirectional LSTM, is created based on the model discussed here, combined with other components. This model significantly improved performance in both forward and reverse directions by considering the spatial and temporal interactions between the Leap Motion data and the network layers.

A portable DnM sign language translator is presented in [17]. This uses a controller to analyze the gestures in a photograph using various image-processing techniques. The use of deep learning has greatly enhanced the performance of this system. After completing the detection process, the hand gesture signs are converted into spoken language. There is a limitation to this application, which is also highlighted. The cost of converting this into a commercial product is high. Researchers are currently placing greater emphasis on the development of Sign Language Recognition systems that are suitable for commercial use [18]. There are several methodologies applied to find a solution to reduce costs. The main building block is data acquisition. The cost of data acquisition devices is generally high. There is a focus on finding a solution to reduce this cost, which will reduce the overall cost of a commercial product. Another aspect of the research is to review and evaluate how Sign Language Recognition (SLR) can be improved. Different Sign Language Recognition (SLR) systems being developed have their pros and cons. In [19], the authors developed a glove using flex sensors. The hand gesture data from the glove are acquired using an Arduino board.

A leap motion controller is an interactive tool with many applications, such as manufacturing, 3D modeling, and other fields. This is also true in detecting hand gestures. It follows the movement of the hand with the help of cameras and infrared LEDs [20]. The device is used in combination with the SDK development toolkit, which collects hand gesture data. The results are favorable. The focus of the research in [21] is acoustic communication.

The authors highlighted the difficulties that DnM people face in communicating with others. A real-time device that aids DnM people, where they use a specifically designed glove fitted with five flex sensors and one accelerometer. The sensor output will change based on hand gestures. The acquired data are then processed using an Android application. This application can interpret gestures in Arabic Sign Language (ArSL). The output is in the form of both text and voice. The prototype is a reliable and accurate device developed at a low cost.

The importance of sign language is highlighted in [22]. The focus is also on NDnM people, as they also need to learn sign language so that they can communicate with DnM people. The available material is less effective in terms of training sign language for NDnM people. The authors developed an interactive Chinese sign language teaching system for smartphones that makes it easier to learn sign language. The application teaches sign language in an interactive way, where a mobile phone camera is used to record the activity. The researchers also added a vocabulary of 100 words. The application also carries out analysis and assessment, which helps the user.

In [23], a mobile phone application 'BridgeApp' is developed for DnM people. They used speech-to-text, speech-to-visual, and sign language features in their implementation. This application works in offline mode. They have implemented American Sign Language (ASL) and Filipino Sign Language (FSL). A system based on vibrotactile, and visual feedback is developed in [24]. DnM people can communicate on their own without the use of sign language using this system. An Android application is also developed by the authors.

The authors in [25] used ARM LPC-2148 to connect different sensors and actuators to a Braille keypad, which is a user-friendly application for blind people. They have also installed sensors in walking sticks to aid people who are visually impaired. A Hand Gesture Recognition System for DnM people is designed using a Fuzzy-Neural Network [26]. The data was acquired from a hand gesture translating glove. The accuracy of gesture recognition using fuzzy logic is discussed. In [27], an application is developed for DnM, which detects facial emotion. The application performs eye-tracking and event-related potential analysis. The authors carried out extensive testing using 630 images of hand gestures.

A deep learning-based hand gesture translation system is presented in [28]. The software automatically recognized hand gestures at 94% accuracy. A process to convert to and from sign language is discussed in [29]. The gestures are detected and then converted into text. To develop an effective model, a variety of ML and AI techniques are used, in addition to natural language processing (NLP) and convolutional neural networks CNN). The authors also used attention-based long short-term memory (LSTM) to detect rapid and continuous lip movements.

An intelligent glove for sign language communication is discussed in [30]. The researchers used a flex sensor and a GY-521 module interfaced with Arduino. The flex sensor measures the movement of the fingers and generates a voltage output. The acquired hand gesture data are then converted into text. In [31], the authors developed a module that translated sign language. The user can wear it around the neck. The device processes hand gesture images using image processing techniques and deep learning models to recognize the sign, which is then converted into a voice using a text-to-speech convertor.

A CNN-based system is presented in [32], which also provides a training option. This system also supports multiple languages. In [33], a low-cost Arduino-based system connected to flex sensors is presented. This system is used for gesture detection. The system also uses a gyroscope and an accelerometer. A picture-based communication application is presented in [34]. It is a portable, easy-to-use communication tool that helps people communicate with each other. A Chinese language translation and processing tool is presented in [35]. The DnM people can watch videos with Chinese subtitles using this tool.

In [36], a video chat application is developed. This is based on the Indian Sign language (ISL). The process is initiated when a user starts making hand gestures, which are then picked up by a camera. The algorithm, which is based on CNN, then converts this into

phrases and numbers. A medical consultation system is presented in [37], where the DnM person can attend a hospital and communicate with the doctor. The system provides two-way communication. In [38,39], the authors applied machine learning to solve a new problem for setting up electronic product test sites. They created a dataset for the machine learning algorithm.

The creation of a sign language dataset is outside the scope of this work. The following references provide the details of some available sign language datasets.

In [40], the authors presented the Argentinian sign language dataset. They included 3200 videos in the dataset. They highlighted the importance of having the entire dataset for training purposes. The authors in [41] proposed a large-scale ASL dataset that includes 25,000 videos. A Chinese sign language dataset is collected and processed by the authors in [42]. The dataset includes 500 categories and was evaluated by the authors, including another dataset. The authors in [43] created a new Russian sign language dataset. A new Turkish sign language dataset is presented, which includes 226 signs that the authors processed using CNN [44]. In [45], the authors presented a word-level ASL dataset. The authors in [46] worked on word-level ASL and collected references related to dataset creation. In [21], the authors focused on Arabic sign language and created a glove for detecting this dataset.

Table 1 is a summary of the literature review and details of the hardware platform and software or algorithm used by the researchers. The last column highlights some features of the work carried out.

Table 1. Summary of literature review.

References	Hardware	Software/Algorithm	Features/Limitations
[7]	Gloves, Sensors, Arduino	KNN	Custom made, require validation; DnM to NDnM only
[8]	PC based	None	NDnM to DnM to only
[9]	None	None	Survey only
[10]	PC based	Image processing	DnM to NDnM only
[11]	Mobile app	Speech recognition	DnM to NDnM only
[12]	Mobile app	ANN	DnM to NDnM only
[13]	PC based	PCA	DnM to NDnM only
[14]	None	None	Survey only
[15]	Leap Motion	CPI	DnM to NDnM only
[16]	Leap Motion	LSTM	DnM to NDnM only
[17]	Raspberry Pi	Deep learning	DnM to NDnM only
[18]	PC based	Various	Survey only
[19]	Gloves, Sensors, Arduino	None	DnM to NDnM only
[20]	Leap Motion	None	DnM to NDnM only
[21]	Gloves, Sensors, Mobile app	None	DnM to NDnM only
[22]	Mobile app	None	DnM to NDnM only
[23]	Mobile app	None	DnM to NDnM only
[24]	Mobile app	None	DnM to NDnM to DnM
[25]	Sensors, Standalone	None	DnM to NDnM only
[26]	PC, Microcontroller	FNN	DnM to NDnM only
[27]	Gloves, Microcontroller	None	DnM to NDnM only
[28]	PC based	Deep learning	DnM to NDnM only
[29]	PC based	CNN	DnM to NDnM only
[30]	Gloves, Sensors, Arduino	KNN	DnM to NDnM only
[31]	PC based	NN	DnM to NDnM only
[32]	PC based	CNN	DnM to NDnM only
[37]	Mobile app	Mobile Net	DnM to NDnM to DnM

Research Questions

This section provides the conclusion of the literature review and highlights some limitations. In the research work reviewed in Section 2, the implementation is mostly meant for NDnM to understand sign language using either a PC-based or a standalone system. Some researchers created their devices using sensors and gloves and demonstrated their prototypes. It is important to understand that converting a prototype into a product is itself a huge task and requires a lot of work and research. These products will then have to go through a tedious validation and verification process. This means that it will take a lot of time for these products to be ready for use.

The proposed system focuses on the limitations of existing systems in terms of providing an end-to-end system where both DnM and NDnM people can communicate in real time. The system uses a commercial-off-the-shelf (COTS) device to acquire hand gesture signals, which means there will be no need to design and then validate custom-made hardware. This will allow the product to be available for use quickly. The other feature is that the system is not application dependent and can be used in any environment where communication is required between DnM and NDnM people. The research is also focused on a low-cost system that is affordable and can be used with minimum training.

3. Research Methodology

This section presents the research methodology of DnM-related research work. The research is carried out in two steps, which are shown in Figure 2.
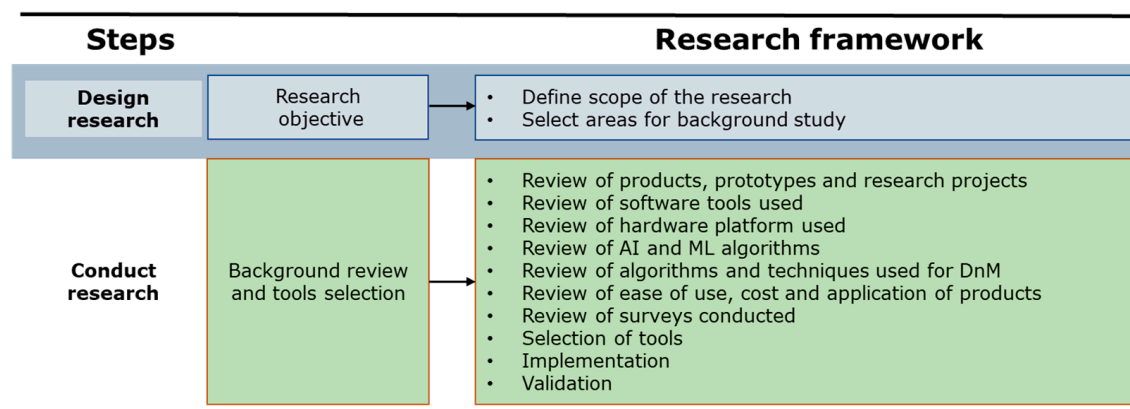

Figure 2. Research Methodology.

3.1. Design Research

The first task is to define the scope of this research and identify the activities to be carried out within the scope. The sequence of the activities is also defined. The first step is listing the areas where a review is required. Literature related to the area of this research is selected for review. The sub-category to review includes the software and hardware tools used, ease of use, how performance is evaluated, the algorithm, and techniques used.

3.2. Conduct Research

The second task is to conduct research and complete the tasks defined in the previous task. The first step is to collect existing research work within the scope of this research. The next step is to review some AI and ML algorithms and techniques. An extensive review is then carried out on the research work selected, which includes a review of the hardware and software platform, quality of the product or prototype, ease of use, and accuracy. After completing the review, tools are selected for progressing with this research. The last steps are implementation and validation.

4. Proposed System Block Diagram and Novel Features

In this section, the novel features of the proposed systems are presented, followed by a detailed description of the system.

4.1. Novel Features of the Proposed System

Figure 3 presents the six key features of this PC-based system. The proposed system is low cost and provides two-way communication between DnM and NDnM people. The system proposes a dataset for supervised machine learning to detect languages. The system uses a leap device t and a sound card for two-way communication. The proposed system provides a complete solution for communication between DnM and NDnM people. A unique feature of this system is its proposed dataset for detecting multiple languages. The PC software application is developed in LabVIEW version 2019 [47].
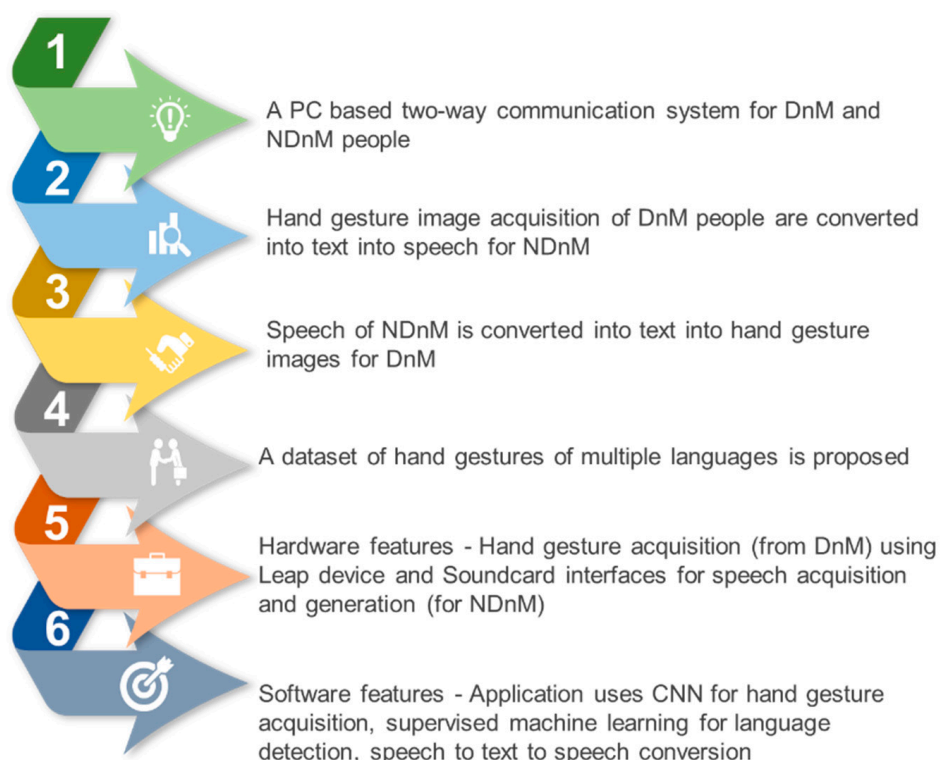

Figure 3. Novel features of the system.

4.2. Proposed System Block Diagram

The proposed system provides two-way communication between DnM and NDnM people. The hardware block diagram is shown in Figure 4. Hand gestures from DnM people are acquired using a Leap device, and the data are then sent to the PC application. The leap device captures the hand gesture data as an image, which is then processed to obtain the recognizable gesture. The data then goes through several activities implemented in the software and will be discussed later. The data are finally converted into speech, which is generated through sound card speaker output. NDnM people can listen to the speech and then respond through the speech, which is then acquired using PC microphone input. The acquired audio is then displayed as text on the screen for the DnM person to read.
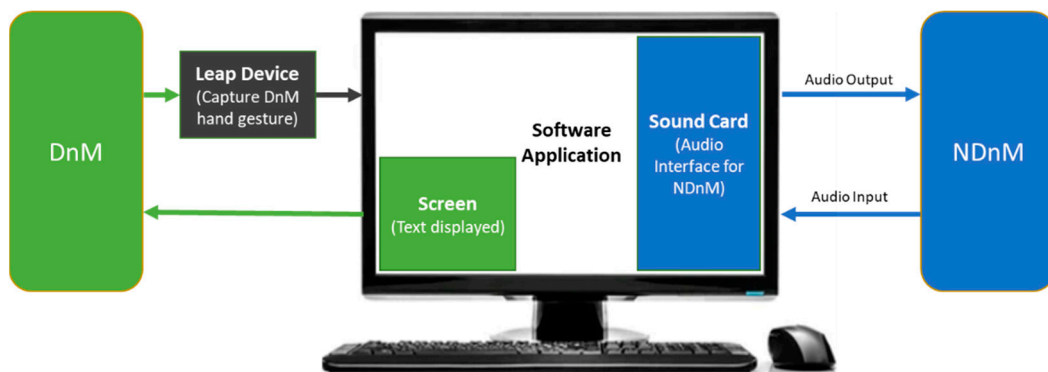
Figure 4. Hardware Interface Block Diagram—Communication between DnM and NDnM.

Figure 5 shows the various software building blocks. The gesture data are acquired and processed by the Leap device. The processing is based on MobileNet to detect gesture data. The data are then fed to the PC software application, which is developed in LabVIEW. The data are first processed using a supervised machine-learning algorithm that detects the language. This is done using a dataset created as part of this research work. The LabVIEW application then converts this into text that the NDnM person can read. The last step is to convert the text into speech, which is generated through a PC sound card. This completed the communication from DnM to NDnM. The NDnM person records the speech using a PC sound card microphone input. Speech is acquired, and then language is detected using supervised machine learning. The detected language is then converted into text, which is finally displayed for the DnM person to read.
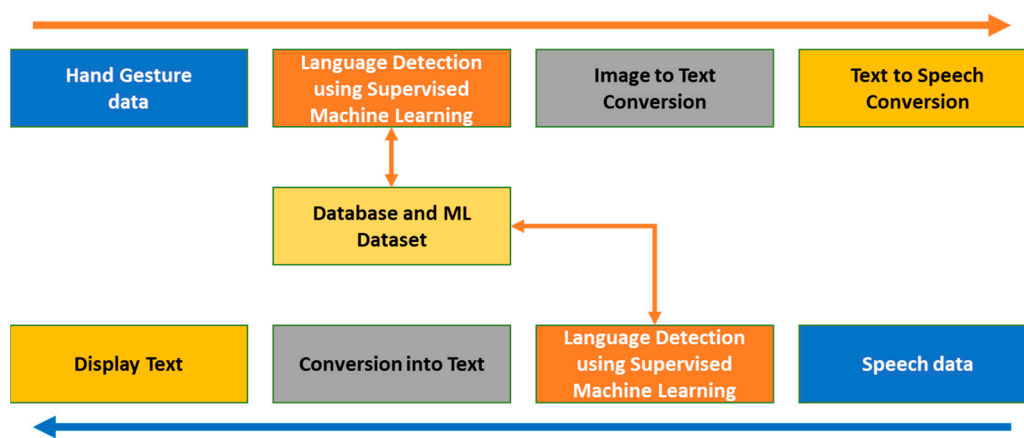

Figure 5. Software Block Diagram—Communication between DnM and NDnM.

Figure 6 shows the machine learning-based implementation through 8 steps. The system acquires hand gesture images using a Leap motion device, which is then processed, and the results are generated. The proposed system collects data that are used to improve system performance. The system also collects user information, which is very useful for improving the machine learning algorithm and implementation. Some parameters used for processing the image are listed in steps 2 and 3.

In Figure 6, the process starts by providing relevant information to the machine learning algorithm through the input layer as shown by '1'. The processing is completed within the hidden layers and activities are marked by '2' till '7'. Step '2' is the initialization stage where the input image data goes through different tasks as listed. Step '3' is the processing stage where the image data is processed to detect certain features. Sign language dataset is detected in step '4'. The next step is to detect language. Currently there is only one language i.e., English as listed in step '5'. Data is stored in step '6' while the last step

i.e., '7' in the hidden layer is for user data collection. The results are generated through the output layer as marked by step '8'.
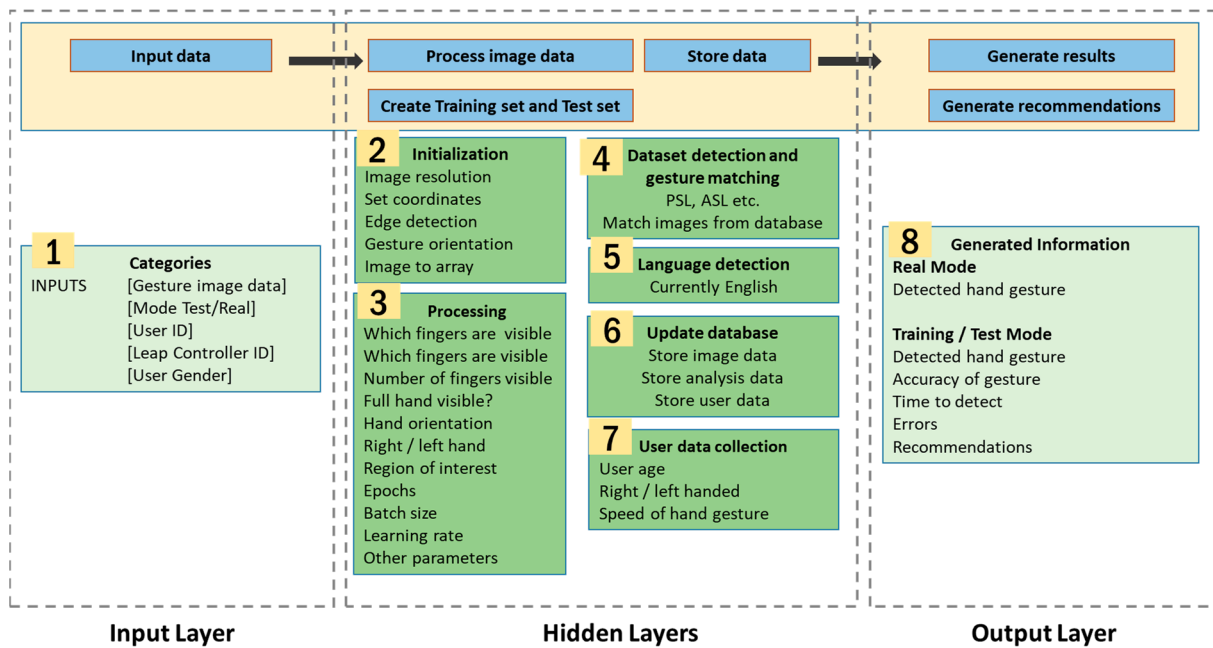

Figure 6. Machine Learning-based Actual and Training System Implementation.

Figure 7 shows a screenshot of the main software application. In this instance, the DnM person generated 5 gestures, i.e., 'H', 'E', 'L', 'L', and 'O', which are detected at a time. The software application created the word using these. Finally, the text is converted into audio and generated for NDnM people to listen to. In response, the NDnM person replied with the word 'GOOD', which is acquired through microphone input and converted from speech to text for the DnM person to read.
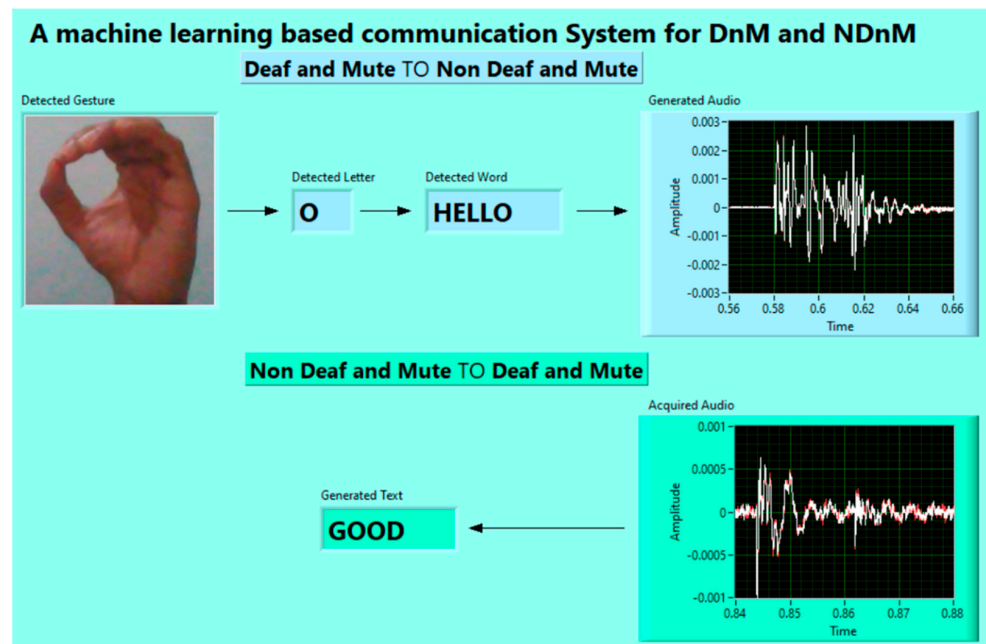

Figure 7. Software Application Screenshot.

5. Experimental Setup and Results

The validation of the proposed system and subsequent results are presented in this section. Section 5.1. lists the criteria and data used to validate the proposed system.

5.1. System Validation Dataset and Criteria

To validate the proposed system, ASL is selected mainly due to the availability of a full dataset, which is needed to create an effective training set. This dataset is used globally; thus, the number of end-users will increase. Other parameters considered are that this dataset can be modified and redistributed, open source, use its own unique grammar, and how many people are using this dataset. An in-depth review of the sign language dataset with the aim of creating a new dataset is outside the scope of what is presented in this manuscript.

For any new proposed system, the first step is always to select known input test data and determine the expected output. Through this approach, the system can be validated, limitations can be understood, and accuracy can be defined. Once this is done, the second step is to expand the input test dataset and process it through the system. This process will help enhance the capabilities of this new system.

Using a known dataset is the key to validating the proposed system. A fully working and validated system will then open the door for further research by processing new datasets through the system.

Here, a validation strategy is defined in which some parameters are fixed to evaluate the proposed system. The first is to select a dataset for hand gestures. For this, the images are used from Kaggle [48]. The hand gesture images used for validation of the proposed system are of American Sign Language (ASL). This process is used to review the proposed system for accuracy, speed, ease of communication, and effectiveness of the machine learning algorithm. The images used are in color and RGB form, although other forms can also be used.

Alphabets A to Z and numbers 0 to 9 are used for validation. Within the dataset, there are 1820 images of alphabets A to Z and 700 images of numbers 0 to 9. These images show different variants of these alphabets and numbers. The number of images is appropriate for the evaluation of the algorithm. Images of some alphabets and numbers selected randomly are shown in Figure 8. These images provide a variety of gestures that are detected and evaluated.
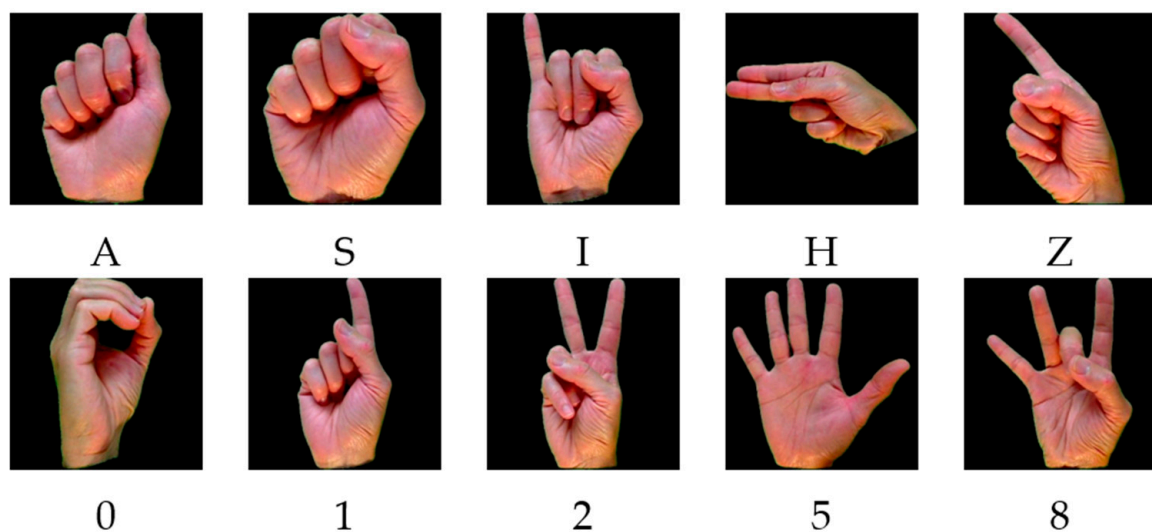

Figure 8. A subset of alphabets and numbers from the Kaggle dataset.

The dataset is split into two groups. The first includes 70% of the data, which is used for training, while the remaining 30% is used for testing. This is a standard split for evaluating machine learning-based algorithms. In other combinations, researchers use 60% of the data for training, while 20% each is used for validation and testing, respectively.

For alphabet classification, 1274 images, i.e., 70% of 1820, are used for training, while the remaining 30%, i.e., 546 images, are used for testing. Similarly, for the evaluation of numbers, 70% of images, i.e., 490 out of 700, are allocated for training, while the remaining 210 images are used for testing.

5.2. Experiment 1—Processing Quality of Acquired Gesture

In this experiment, different variants of one image of 'A' are selected, and hand gesture detection is processed. The gestures in all the images are identical, and only the quality is varied. Figure 9 shows the software screenshot showing the different qualities of the same image of gesture 'A'. The image quality is changed to create five more images of gesture 'A', and then these images are processed using the image detection algorithm.
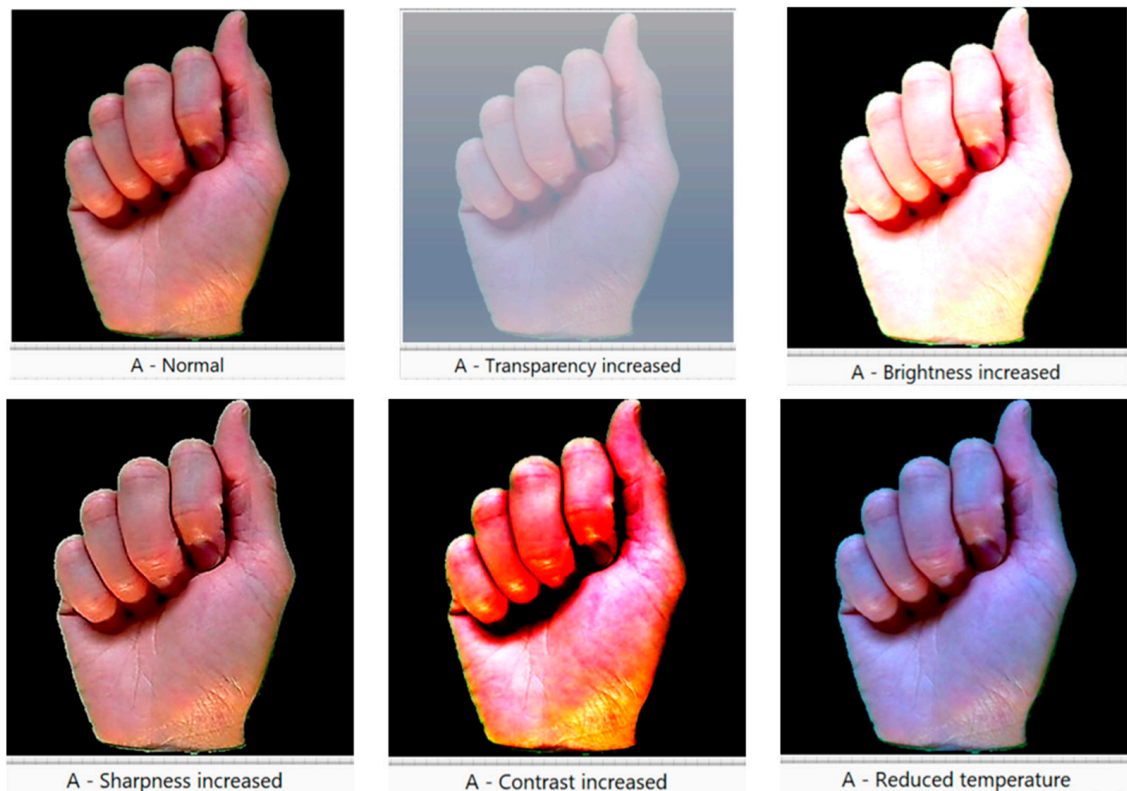

Figure 9. Hand gesture dataset for experiment 1.

The results are shown graphically in Figure 10. Each image is processed 100 times, and the correct detection percentage is shown. It is concluded from this experiment that gesture detection is more than 93% for different image qualities. The average detection percentage of the six images is 96%.
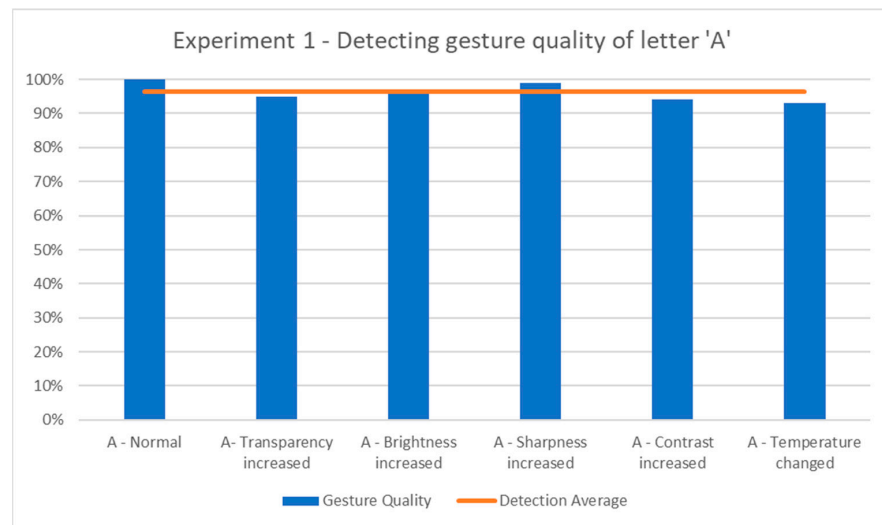
Figure 10. Quality detection accuracy of the algorithm.

5.3. Experiment 2—Processing Variation in Gestures

In this experiment, two variants of three gestures, i.e., 'H', 'I', and 'Y', are used. The two gestures of each letter are different. Here, the algorithm is evaluated for detecting a gesture with a variation. Figure 11 shows images of the three gestures used for this experiment. The images in each row are of the same gesture but slightly different. The system performance is evaluated for this scenario by checking the accuracy of detection for these hand gestures.

Figure 11. Hand gesture dataset for experiment 2.

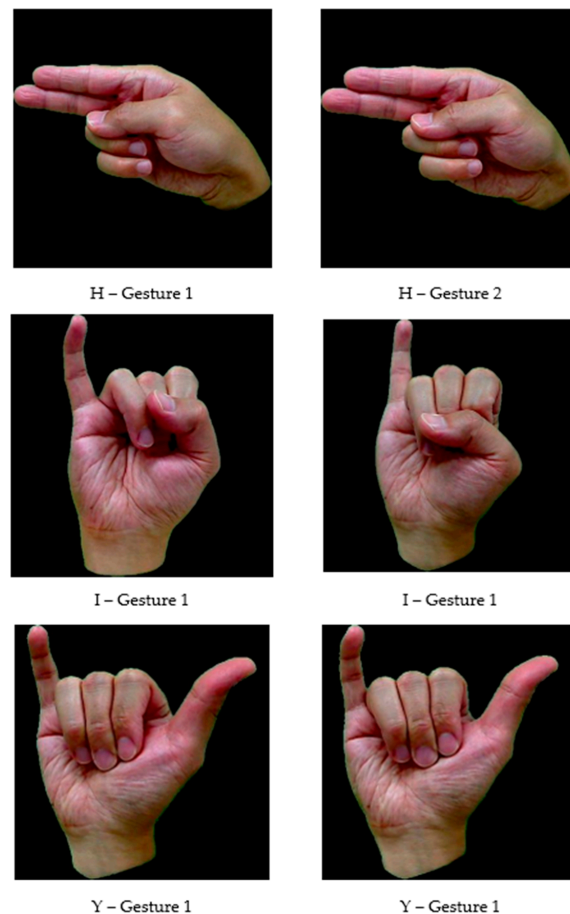The results of the experiment are shown graphically in Figure 12. For this experiment, three different letters are selected with two slightly different gestures each. The detection accuracy is between 81% and 100%. As mentioned, in experiment 1, each image is processed 100 times. It is concluded from this experiment that the accuracy of some gesture variations is less than 90%. The overall average of correct detection in percentage is approximately 93%.
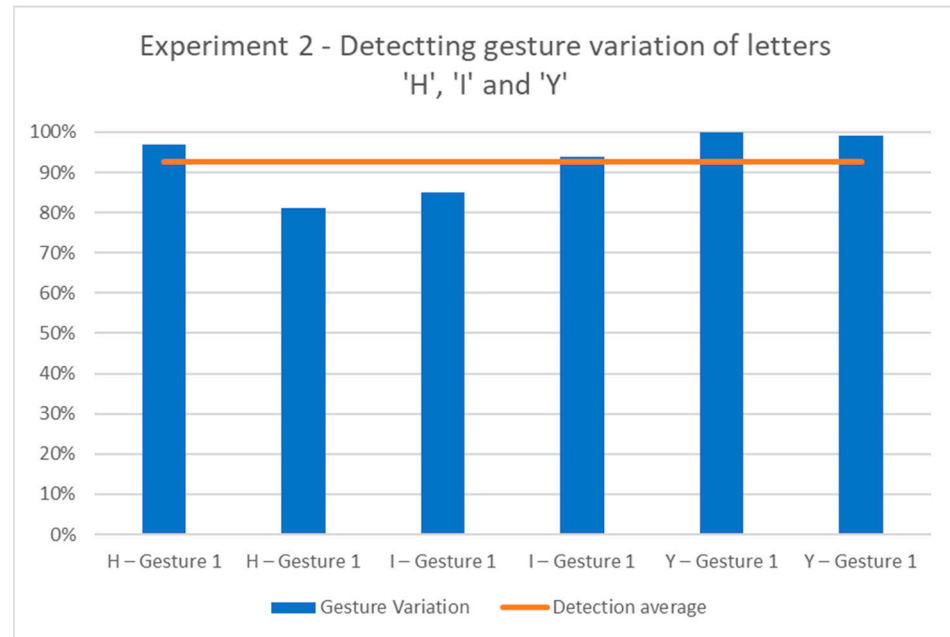

Figure 12. Variation detection accuracy of the algorithm.

5.4. Experiment 3—Machine Learning Algorithm Repeatability

The accuracy of a model is a metric for determining which model is the best at recognizing correlations and patterns between variables in a dataset based on the input or training data. Accuracy is defined as the percentage of correct predictions, which is simply the ratio of correct guess or detection to the total number of predictions. This is presented in Equation (1).

$$\text{Accuracy} = \frac{Correct\ Detection}{Total\ number\ of\ predictions} \tag{1}$$

In this experiment, the repeatability of the algorithm is evaluated. The system is validated by processing the gestures of all letters and numbers. For validation, the most accurate gesture is selected and processed. The results are presented in Table 2.

Table 2. Accuracy using Mobile Net.

Letters and Number	Accuracy
A to Z	98.46
0 to 9	98.9

5.5. Experiment 4—Algorithm Performance

The performance of the system is further evaluated using a confusion matrix. This is an N x N matrix, where 'N' denotes the number of target classes. The matrix compares the actual values and the values predicted by the machine learning algorithm. This provides a comprehensive picture of the performance of the machine learning algorithm. This also highlights errors in detecting gestures. The actual or expected values are shown in the columns, while the values predicted by the algorithm after detecting the gesture are listed

in rows. The confusion matrix for letters 'A to Z' is shown in Figure 13, while the confusion matrix for numbers '0 to 9' is presented in Figure 14.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0.1	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0.1	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0.9	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0.9

Figure 13. Confusion matrix of letters A to Z.

	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
2	0	0	0.97	0.02	0	0	0.01	0	0	0
3	0	0	0	0.99	0	0	0	0	0.01	0
4	0	0	0	0	0.99	0	0	0	0	0.01
5	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	0.97	0	0.03	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0.02	0	0	0	0	0.98

Figure 14. Confusion matrix of numbers 0 to 9.

The reason for incorrect gesture recognition is due to some similar gestures. Similar gestures are shown in Figure 15. The accuracy of the detection of similar gestures is both dependent on the machine learning algorithm as well as on how correct the user's gesture is.
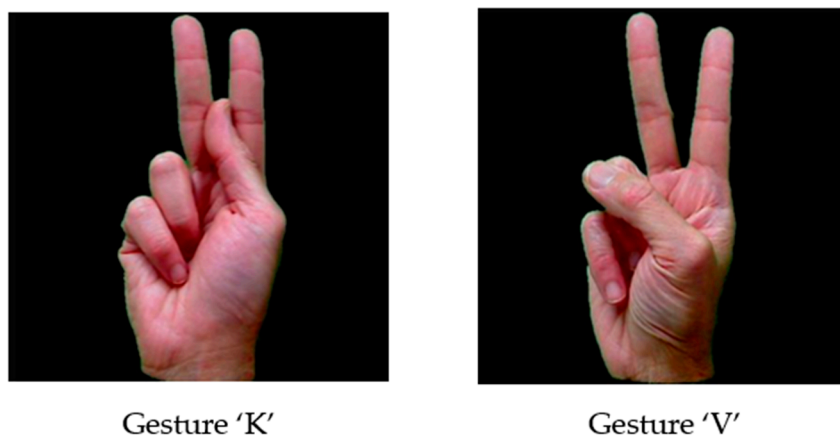
Gesture 'K'   Gesture 'V'

Figure 15. Similar gestures.

6. Discussion

DnM people are an integral part of society, and the system presented in this manuscript aims to help them communicate with NDnM people. In this manuscript, machine learning is applied to develop a two-way system. The system is computer-based, where the DnM interface is provided so that their hand gestures are acquired and converted into text and then voice. The NDnM people are provided with an audio interface in which their voices are acquired and then converted into text. The system uses deep learning to detect and process hand gestures, which are implemented using a Leap device. Language detection is achieved through supervised machine learning. The system also provides speech-to-text-to-speech conversion, which is a useful feature for NDnM people. Due to this option, NDnM people do not need any training or are not required to know sign language.

Many existing similar systems only focus on DnM people, where the interface is provided for them so that their hand gestures are detected. These systems expect NDnM people to learn about the detection of hand gestures and the interfaces provided to them. It is not easy for NDnM people to learn sign language, hence reducing the effectiveness of this system. The proposed system provides a complete solution to the problems faced by both DnM and NDnM people.

The system is presented through hardware and software block diagrams in the manuscript. The system is validated through four experiments in which system performance is evaluated. The system performance is evaluated using several experiments, namely detection of similar hand gestures, low-resolution gesture images, inaccurate gestures, speed of making hand gestures, etc.

The system is designed with a Leap motion device, and the results are encouraging. Using a COTS device means that there is no need for extensive validation and verification to convert a prototype into a product. The other advantage is that there is no calibration required compared to other custom-made devices, which are normally based on sensors connected to a glove and require continuous maintenance to make sure the sensors are working correctly. Upgradation of custom-made devices is also a challenge where the availability of components has been a problem recently.

Another feature of the proposed system is that the end user is not expected to undergo rigorous training. Ease of use is a key feature of this research. Another key feature is that DnM individuals can also use this system to communicate with each other. In this mode, the system will acquire the data, which will help improve the performance of the machine learning algorithm. The software application is easy to install and doesn't require any licensing. The software processes new data and stores them in the database.

7. Conclusions

An automated system based on machine learning is presented in this article, which provides an interface for communication between DnM and NDnM people. The system is

based on a modular approach in which hand gestures from DnM people are converted into speech. The speech from NDnM people is converted into text. The system also provides a learning mode in which DnM people can communicate with each other using the system. In this mode, the system acquires data during ongoing communication between DnM users. This mode is also used for evaluating system performance by comparing the detection accuracy between computers and humans. The system is validated through a series of experiments, and it is concluded that the system can detect hand gestures correctly, and the overall accuracy is mostly more than 90%. There are a few scenarios in which the accuracy is between 80% and 90% for similar gestures. The system also stores the acquired data and uses it for processing. Due to this feature, system accuracy will increase due to more data being added to the dataset.

Another novel feature is that the system supports multiple languages. More work is being carried out to evaluate this feature by adding more data to the database. The software is developed in LabVIEW, which provides a high-quality graphical user interface (GUI) and quick development. In the future, the GUI can be modified to add more features. Multiple language gestures can be added to the database to support multi-language support. The challenge will be to process similar gestures from different languages.

References

1. Vaidya, O.; Gandhe, S.; Sharma, A.; Bhate, A.; Bhosale, V.; Mahale, R. Design and development of hand gesture based communication device for deaf and mute people. In Proceedings of the IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 4–6 December 2020; pp. 102–106. [CrossRef]
2. Marin, G.; Dominio, F.; Zanuttigh, P. Hand gesture recognition with Leap Motion and Kinect devices. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 1565–1569. [CrossRef]
3. Saleem, M.I.; Otero, P.; Noor, S.; Aftab, R. Full duplex smart system for Deaf & Dumb and normal people. In Proceedings of the Global Conference on Wireless and Optical Technologies (GCWOT), Mlaga, Spain, 6–8 October 2020; pp. 1–7. [CrossRef]
4. Deb, S.; Suraksha; Bhattacharya, P. Augmented Sign Language Modeling (ASLM) with interaction design on smartphone—An assistive learning and communication tool for inclusive classroom. *Procedia Comput. Sci.* **2018**, *125*, 492–500. [CrossRef]
5. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
6. Rishi, K.; Prarthana, A.; Pravena, K.S.; Sasikala, S.; Arunkumar, S. Two-way sign language conversion for assisting deaf-mutes using neural network. In Proceedings of the 8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, Coimbatore, India, 25–26 March 2022; pp. 642–646. [CrossRef]
7. Anupama, H.S.; Usha, B.A.; Madhushankar, S.; Vivek, V.; Kulkarni, Y. Automated sign language interpreter using data gloves. In Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 472–476. [CrossRef]
8. Kawai, H.; Tamura, S. Deaf-and-mute sign language generation system. *Pattern Recognit.* **1985**, *18*, 199–205. [CrossRef]
9. Bhadauria, R.S.; Nair, S.; Pal, D.K. A Survey of Deaf Mutes. *Med. J. Armed Forces India* **2007**, *63*, 29–32. [CrossRef] [PubMed]
10. Sood, A.; Mishra, A. AAWAAZ: A communication system for deaf and dumb. In Proceedings of the 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 7–9 September 2016; pp. 620–624. [CrossRef]
11. Yousaf, K.; Mehmood, Z.; Saba, T.; Rehman, R.; Rashid, M.; Altaf, M.; Shuguang, Z. A Novel Technique for Speech Recognition and Visualization Based Mobile Application to Support Two-Way Communication between Deaf-Mute and Normal Peoples. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 1013234. [CrossRef]

12. Raheja, J.L.; Singhal, A.; Chaudhary, A. Android Based Portable Hand Sign Recognition System. *arXiv* **2015**, arXiv:1503.03614.

13. Soni, N.S.; Nagmode, M.S.; Komati, R.D. Online hand gesture recognition & classification for deaf & dumb. In Proceedings of the International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016; pp. 1–4. [CrossRef]

14. Chakrabarti, S. State of deaf children in West Bengal, India: What can be done to improve outcome. *Int. J. Pediatr. Otorhinolaryngol.* **2018**, *110*, 3742. [CrossRef] [PubMed]

15. Ameur, S.; Khalifa, A.B.; Bouhlel, M.S. Chronological pattern indexing: An efficient feature extraction method for hand gesture recognition with Leap Motion. *J. Vis. Commun. Image Represent.* **2020**, *70*, 102842. [CrossRef]

16. Ameur, S.; Khalifa, A.B.; Bouhlel, M.S. A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with Leap Motion. *Entertain. Comput.* **2020**, *35*, 100373. [CrossRef]

17. Boppana, L.; Ahamed, R.; Rane, H.; Kodali, R.K. Assistive sign language converter for deaf and dumb. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; pp. 302–307. [CrossRef]

18. Suharjito; Anderson, R.; Wiryana, F.; Ariesta, M.C.; Kusuma, G.P. Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output. *Procedia Comput. Sci.* **2017**, *116*, 441–448. [CrossRef]

19. Patwary, A.S.; Zaohar, Z.; Sornaly, A.A.; Khan, R. Speaking system for deaf and mute people with flex sensors. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics, ICOEI 2022, Tirunelveli, India, 28–30 April 2022; pp. 168–173. [CrossRef]

20. Sharma, A.; Yadav, A.; Srivastava, S.; Gupta, R. Analysis of movement and gesture recognition using Leap Motion Controller. *Procedia Comput. Sci.* **2018**, *132*, 551–556. [CrossRef]

21. Salem, N.; Alharbi, S.; Khezendar, R.; Alshami, H. Real-time glove and android application for visual and audible Arabic sign language translation. *Procedia Comput. Sci.* **2019**, *163*, 450–459. [CrossRef]

22. Zhang, Y.; Min, Y.; Chen, X. Teaching Chinese Sign Language with a Smartphone. *Virtual Real. Intell. Hardw.* **2021**, *3*, 248–260. [CrossRef]

23. Samonte, M.J.C.; Gazmin, R.A.; Soriano, J.D.S.; Valencia, M.N.O. BridgeApp: An assistive mobile communication application for the deaf and mute. In Proceedings of the 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 16–18 October 2019; pp. 1310–1315. [CrossRef]

24. Sobhan, M.; Chowdhury, M.Z.; Ahsan, I.; Mahmud, H.; Hasan, M.K. A communication aid system for deaf and mute using vibrotactile and visual feedback. In Proceedings of the 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 21–22 September 2019; pp. 184–190. [CrossRef]

25. KN, S.K.; Sathish, R.; Vinayak, S.; Pandit, T.P. Braille assistance system for visually impaired, blind & deaf-mute people in indoor & outdoor application. In Proceedings of the 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 17–18 May 2019; pp. 1505–1509. [CrossRef]

26. Villagomez, E.B.; King, R.A.; Ordinario, M.J.; Lazaro, J.; Villaverde, J.F. Hand gesture recognition for deaf-mute using FuzzyNeural network. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics—Asia (ICCE-Asia), Bangkok, Thailand, 12–14 June 2019; pp. 30–33. [CrossRef]

27. Tao, Y.; Huo, S.; Zhou, W. Research on communication APP for deaf and mute people based on face emotion recognition technology. In Proceedings of the 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Weihai, China, 14–16 October 2020; pp. 547–552. [CrossRef]

28. Shareef, S.K.; Haritha, I.V.S.L.; Prasanna, Y.L.; Kumar, G.K. Deep learning based hand gesture translation system. In Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 1531–1534. [CrossRef]

29. Dhruv, A.J.; Bharti, S.K. Real-time sign language converter for mute and deaf people. In Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 24–26 September 2021; pp. 1–6. [CrossRef]

30. Rosero-Montalvo, P.D.; Godoy-Trujillo, P.; Flores-Bosmediano, E.; Carrascal-Garcia, J.; Otero-Potosi, S.; Benitez-Pereira, H.; Peluffo-Ordonez, D.H. Sign language recognition based on intelligent glove using machine learning techniques. In Proceedings of the 2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM), Cuenca, Ecuador, 15–19 October 2018; pp. 1–5. [CrossRef]

31. Janeera, D.A.; Raja, K.M.; Pravin, U.K.R.; Kumar, M.K. Neural network based real time sign language interpreter for virtual meet. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 1593–1597. [CrossRef]

32. Gupta, A.M.; Koltharkar, S.S.; Patel, H.D.; Naik, S. DRISHYAM: An interpreter for deaf and mute using single shot detector model. In Proceedings of the 8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, Coimbatore, India, 25–26 March 2022; pp. 365–371. [CrossRef]

33. Lan, S.; Ye, L.; Zhang, K. Attention-augmented electromagnetic representation of sign language for human-computer interaction in deafand-mute community. In Proceedings of the 2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), Singapore, Singapore, 4–10 December 2021; pp. 47–48. [CrossRef]

34. Telluri, P.; Manam, S.; Somarouthu, S.; Oli, J.M.; Ramesh, C. Low cost flex powered gesture detection system and its applications. In Proceedings of the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 15–17 July 2020; pp. 1128–1131. [CrossRef]

35. Jamdar, V.; Garje, Y.; Khedekar, T.; Waghmare, S.; Dhore, M.L. Inner voice—An effortless way of communication for the physically challenged deaf & mute people. In Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 24–26 September 2021; pp. 1–5. [CrossRef]

36. He, Y.; Kuerban, A.; Yu, Q.; Xie, Q. Design and implementation of a sign language translation system for deaf people. In Proceedings of the 2021 3rd International Conference on Natural Language Processing (ICNLP), Beijing, China, 26–28 March 2021; pp. 150–154. [CrossRef]

37. Xia, K.; Lu, W.; Fan, H.; Zhao, Q.A. Sign Language Recognition System Applied to Deaf-Mute Medical Consultation. *Sensors* **2022**, *22*, 9107. [CrossRef] [PubMed]

38. Siddiqui, A.; Zia, M.Y.I.; Otero, P. A universal machine-learning-based automated testing system for consumer electronic products. *Electronics* **2021**, *10*, 136. [CrossRef]

39. Siddiqui, A.; Zia, M.Y.I.; Otero, P. A Novel Process to Setup Electronic Products Test Sites Based on Figure of Merit and Machine Learning. *IEEE Access* **2021**, *9*, 80582–80602. [CrossRef]

40. Ronchetti, F.; Quiroga, F.; Estrebou, C.A.; Lanzarini, L.C.; Rosete, A. LSA64: An argentinian sign language dataset. In Proceedings of the Congreso Argentino de Ciencias de La Computacion (CACIC), San Luis, Argentina, 3–10 October 2016; pp. 794–803.

41. Joze, H.R.V.; Koller, O. MS-ASL: A large-scale data set and benchmark for understanding American sign language. In Proceedings of the 30th British Machine Vision Conference 2019, BMVC 2019, Cardiff, UK, 9–12 September 2019.

42. Jie, H.; Zhou, W.; Li, H.; Li, W. Attention-Based 3D-CNNs for Large-Vocabulary Sign Language Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 2822–2832. [CrossRef]

43. Kagirov, I.; Ivanko, D.; Ryumin, D.; Axyonov, A.; Karpov, A. TheRuSLan: Database of Russian sign language. In Proceedings of the LREC 2020—12th International Conference on Language Resources and Evaluation, Conference Proceedings, Marseille, France, 11–16 May 2020; pp. 6079–6085.

44. Sincan, O.M.; Keles, H.Y. AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods. *IEEE Access* **2020**, *8*, 181340–181355. [CrossRef]

45. Dongxu, L.; Opazo, C.R.; Yu, X.; Li, H. Word-Level Deep Sign Language Recognition from Video: A New Large-Scale Dataset and Methods Comparison. 2020. Available online: https://dxli94.github.io/ (accessed on 10 November 2022).

46. Tavella, F.; Schlegel, V.; Romeo, M.; Galata, A.; Cangelosi, A. WLASL-LEX: A Dataset for Recognising Phonological Properties in American Sign Language. *arXiv* **2022**, arXiv:2203.06096.

47. Engineer Ambitiously—NI. Available online: https://www.ni.com/en-gb.html (accessed on 23 October 2022).

48. Kaggle Dataset. Available online: https://www.kaggle.com/datasets/alexalex1211/aslamerican-sign-language (accessed on 23 October 2022).