

## Review Article

# A Computer Vision-Based System for Recognition and Classification of Urdu Sign Language Dataset for Differently Abled People Using Artificial Intelligence

Hira Zahid <sup>1</sup>, Sidra Abid Syed <sup>2</sup>, Munaf Rashid <sup>3</sup>, Samreen Hussain <sup>4</sup>, Asif Umer <sup>5</sup>, Abdul Waheed <sup>6,7</sup>, Shahzad Nasim<sup>8</sup>, Mahdi Zareei <sup>9</sup>, and Nafees Mansoor <sup>10</sup>

<sup>1</sup>Biomedical Engineering Department and Electrical Engineering Department, Ziauddin University, Karachi 74600, Pakistan

<sup>2</sup>Biomedical Engineering Department, Sir Syed University of Engineering and Technology, Karachi 75300, Pakistan

<sup>3</sup>Software Engineering Department, Sir Syed University of Engineering and Technology, Karachi 74600, Pakistan

<sup>4</sup>Electronic Engineering Department, Dawood University of Engineering & Technology, Karachi 74800, Pakistan

<sup>5</sup>Department of Computer Science & IT, Hazara University, Mansehra 23100, Pakistan

<sup>6</sup>Department of Computer Science, Women University Swabi, Swabi 23430, Pakistan

<sup>7</sup>School of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Republic of Korea

<sup>8</sup>Begum Nusrat Bhutto Women University, Faculty of Management Information Science and Technology, Sukkur, Pakistan

<sup>9</sup>Tecnologico de Monterrey, School of Engineering and Sciences, Monterrey 64849, Mexico

<sup>10</sup>Department of Computer Science and Engineering, University of Liberal Arts Bangladesh (ULAB), Dhaka, Bangladesh

Correspondence should be addressed to Abdul Waheed; [abdul@netlab.snu.ac.kr](mailto:abdul@netlab.snu.ac.kr) and Nafees Mansoor; [nafees@nafees.info](mailto:nafees@nafees.info)

Received 18 September 2022; Revised 21 February 2023; Accepted 13 April 2023; Published 26 June 2023

Academic Editor: Junaid Shuja

Copyright © 2023 Hira Zahid et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Communication between normal people and deaf people is the most difficult part of daily life worldwide. It is difficult for a normal person to understand a word from the deaf one in their daily routine. So, to communicate with deaf people, different countries developed different sign languages to make communication easy. In Pakistan, for deaf people, the government developed Urdu Sign Language to communicate with deaf people. Physical trainers and experts are difficult to provide everywhere in society, so we need such a computer/mobile-based system to convert the deaf sign symbol into voice and written alphabet that the normal person can easily get the intentions of the deaf one. In this paper, we provided an image processing and deep learning-based model for Urdu Sign Language. The proposed model is implemented in Python 3 and uses different image processing and machine techniques to capture the video and transform the symbols into voice and Urdu writing. First, we get a video from the deaf person, and then the model crops the frames into pictures. Then, the individual picture is recognized for the sign symbol such as if the deaf showed a symbol for one, then the model recognizes it and shows the letter which he/she wants to tell. Image processing techniques such as OpenCV are used for image recognition and classification while TensorFlow and linear regression are used for training the model to behave intelligently in the future. The results show that the proposed model increased accuracy from 80% to 97% and 100% accordingly. The accuracy of the previously available work was 80% when we implemented the algorithms, while with the proposed algorithm, when we used linear regression, we achieved the highest accuracy. Similarly, when we used the TensorFlow deep learning algorithm, we achieved 97% accuracy which was less than that of the linear regression model.

## 1. Introduction

Because many people are born with impairments, none of us are flawless. We cannot ignore them in society due to a variety of issues. Even the government has a quota set aside

for disabled people. Researchers are attempting to develop digital solutions to overcome limitations associated with special personas and enable them to participate in functioning societies [1]. Everything in our world is flawed, and idealism has no place in it, as many scientific data and

statistics demonstrate. Similarly, man is neither perfect nor ideal, and some people are born differently than others. Since then, they have been dubbed “handicapped.” They are diverse, yet they each have their own set of needs. Deafness affects an estimated 72 million individuals globally [2], including roughly 10 million people in Pakistan. There are different communication styles of deaf people all around the world. Visual communication was used to convey information since the dawn of time. In general, several new sorts of trademarks, languages, and sign language are being adopted all over the world. Without the need for paper or pencil, the deaf community and the community at large may communicate efficiently using a variety of sign languages. Different nations, such as the United States, have their own sign languages, such as American Sign Language, British Sign Language, Spanish Sign Language, and probably sign languages across the world. There are still more sign languages than winks, and numerous varieties of American Sign Language (ASL) are exploited in communication. There are around 60 sign languages that are acknowledged and practiced [3]. ASL is a comprehensive and complicated language, according to the National Institute on Deafness and Other Communication Disorders (NIDCD). Hand gestures, as well as face expressions and muscular movements, are all examples of this. This is not simply a hand gesture translation in English; it can also handle grammar and pronunciation rules, as well as different races and dialects. Furthermore, there are several reports in various languages, including Chinese, American, and Indian, showing that a significant amount of work has been put into the worldwide sign language identification system. Local and regional languages and cultures play an essential role in the evolution of sign language, as they do in the evolution of any spoken language, regardless of origin. Many experts, on the other hand, have questioned why there is no common sign language for signatories [4]. It is identical to wondering why there is not a widely recognized language spoken across the world. Pakistani Sign Language has been used by deaf people in Pakistan to communicate among each other (PSL). It follows linguistic norms, just like all other sign languages, and it contains syntax, letters and words, gestures, and complicated sentences, just like spoken Urdu. It also has its own set of symbols and a constantly changing grammar, much like every other sign language system on the planet [5]. Due to its growth throughout time, PSL has evolved into a comprehensive language. Urdu is the official language of Pakistan and is spoken by many people across South Asia. In Urdu, the most widely used writing systems are Nastaliq and Naskh. Nastaliq style is commonly utilized in ancient Urdu literature and journalism. Persian, Pashto, Punjabi, Balochi, and Seraiki are among the ethnic languages that employ the Nastaliq writing style. Urdu belongs to the Indo-European language family. Since its inception, the Indo-European Urdu language has had its origins in India. In the Indian subcontinent, it is now one of the most frequently spoken languages. Urdu is one of India’s 23 official languages and one of Pakistan’s two languages [6]. In addition, Dubai boasts a sizable Urdu-speaking community. It is spoken by the majority of the world’s inhabitants. This is a written





























version of Urdu that is based on the Arabic script and is developed from the Persian script. Urdu is written from right to left, much like Arabic. Sign languages have formed the backbone of distinct deaf cultures as a practical way of communication for deaf people all over the world. Other symptoms are used by listeners who are unable to communicate vocally owing to a disability or impairment, such as augmented and alternative communication, or whose family members are deaf, such as children of deaf adults. Through picture-to-speech technology, those who are deaf or hard of hearing, as well as blind people, can benefit from this effort. For the blind, having a computer recognize a character in an image and translate it to a sound can be lifesaving. Urdu, unlike Arabic and Persian, has a larger number of separate letters. Urdu’s script is more complicated than Arabic or Persian. The Sind Welfare Association of the Deaf presented the fundamental structure of the Urdu alphabet for deaf people as shown in Figure 1. Similarly, in Figure 2, the Sindh Welfare Association of the Deaf provided the basic structure of number system for deaf people and implemented it in most of the schools. In Figure 3, the association provided structure for the deaf people about English alphabets using two hands while in Figure 4, the association presented English alphabets of sign language using single hand. In this case, gesture recognition has shown to be useful, helping deaf patients to interact with us more effectively and efficiently. Identification of the sign has taken a long time and effort all over the world. In the case of Urdu Sign Language, however, no such work has been undertaken. Around 0.2 million Pakistanis who are deaf or hard of hearing lack access to assistive and rehabilitative technologies. Gestures can be divided into two categories. Static gestures are those in which the hand, body, and face do not move. Signals that do not alter are known as static signals [7].

The perceivable gesture happens within a set duration that the performer physically orchestrates during static motions. The finger and hand positions are recognized and examined one after the other [8]. Other sign languages, such as British Sign Language (BSL), American Sign Language (ASL), Arabic Sign Language (ArSL), and Spanish Sign Language (SSL), are used in different regions of the globe [9]. Each of these sign languages evolves on its own. In general, gestures in sign language consist of theoretical fictional hand motions, such as the thumb, which is frequently employed for the word “OK,” or the principles of the specific sign language. Spell each word one at a time.

To recognize gestures or hand movements, two basic methods are employed [10]. There are two methods: one is based on computer vision, which uses an image evaluation technique to translate images of the singer into text, and the other is based on machine learning.

Wearing sensor-equipped gloves [11] is the third approach. The present status of sign language recognition (SLR) is over 30 years behind the voice recognition technology due to a variety of causes. The capture and detection of two-dimensional video data is far more complicated than the analysis of linear audio signals, which is one of the key reasons for this. Sindh Welfare Association of the Deaf

## Sindh Welfare Association of the Deaf (Regd.) *SINGLE - HAND SYSTEM (NUMBERS)*

<b>1</b> 	<b>2</b> 	<b>3</b> 	<b>4</b> 	<b>5</b> 
<b>6</b> 	<b>7</b> 	<b>8</b> 	<b>9</b> 	<b>1 0</b> 
<b>1 1</b> 	<b>1 2</b> 	<b>1 3</b> 	<b>1 4</b> 	<b>1 5</b> 
<b>1 6</b> 	<b>1 7</b> 	<b>1 8</b> 	<b>1 9</b> 	<b>2 0</b> 
<b>3 0</b> 	<b>4 0</b> 	<b>5 0</b> 	<b>6 0</b> 	<b>7 0</b> 
				
<b>8 0</b> 		<b>9 0</b> 		<b>1 0 0</b> 

**SWAD Office: H-360, 1st Floor, Sector 32/A, Zia Colony, Korangi No.1, Karachi East**  
**E-mail: swadeaf@yahoo.com, swadeaf@gmail.com**

FIGURE 1: Urdu alphabets for deaf persons (Sindh Welfare Association of the Deaf).

provided standard alphabets for disabled people, which will be followed as a standard for Urdu speakers. In Figure 3, the association used the two-hand system for deaf people to communicate with other people without any misguidance or hesitation. They provided advance sign language for communication of English words such as if deaf want to speak "MAN," then he will provide the sign of M then sign for A and then sign for N. So, in this way, the deaf person can share his thoughts with healthy people. All such sign languages need computer-based systems to recognize such signs and

then abbreviate/speak the correct words on behalf of the deaf ones. In Figure 4, the welfare association provided standard single hand English alphabets for deaf people. Instead of using double hand English alphabets, if any deaf person does not have one hand and he wants to communicate on single hand, then there is standard system for such disabled person to communicate with other healthy persons. Furthermore, oral communication's lexical and grammatical features are yet to be completely investigated, and no ordinary terms are accessible. Furthermore, there is no standard definition for

## Sindh Welfare Association of the Deaf (Regd.)

The Alphabet as Spoken in the Sign Language  
Single - Hand System Urdu



**SWAD Office: H-360, 1st Floor, Sector 32/A, Zia Colony, Korangi No.1, Karachi East**  
**E-mail: swadeaf@yahoo.com, swadeaf@gmail.com**

FIGURE 2: Number system for deaf persons (Sindh Welfare Association of the Deaf).

such indicators. In the early 1990s, research on the categorization and identification of sign language achieved a pinnacle. Techniques for collecting data are critical for identifying key aspects of various SLR studies [12]. Much research has looked towards data gloves or cyber gloves to extract the features of mechanical and nonmechanical

components of signals due to the heavy dependence on sensor-based SLR systems. Unfortunately, using such sensors is inconvenient and restricted to the signer. Furthermore, because of the high cost of sensors, sensor-based SLR systems are not viable to deploy. The weight and capacity to manage variations under changing illumination and barriers

# Sindh Welfare Association of the Deaf (Regd.)

## THE ALPHABET IN SIGN LANGUAGE (ENGLISH) DOUBLE - HAND SYSTEM



SWAD Office: H-360, 1st Floor, Sector 32/A, Zia Colony, Korangi No.1, Karachi East  
E-mail: swadeaf@yahoo.com, swadeaf@gmail.com

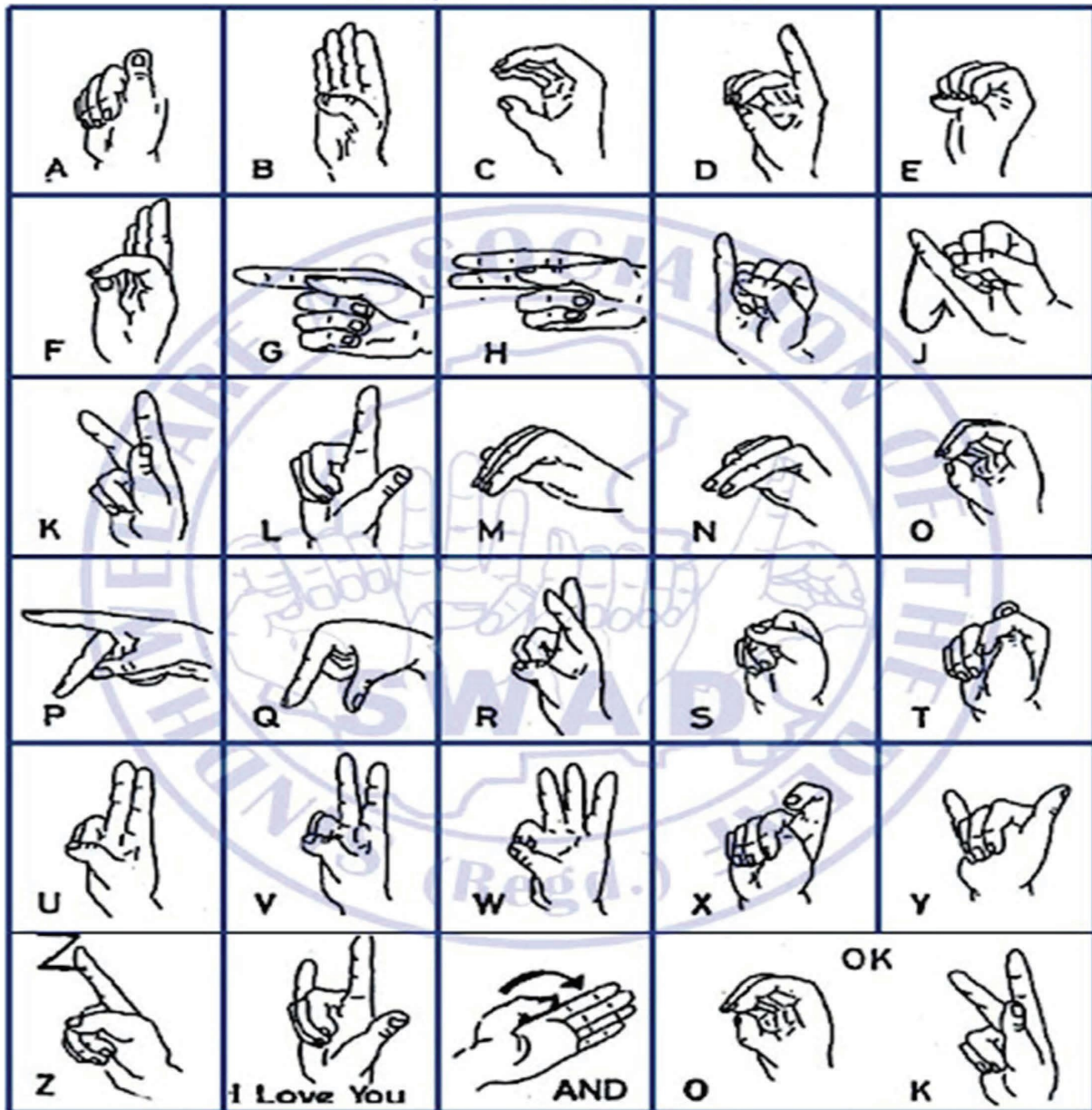
FIGURE 3: English alphabets for deaf persons (Sindh Welfare Association of the Deaf).

in the crowd, dynamic anomalies, and feature extraction phase of vision-based SLR systems, on the other hand, has tremendously affected researchers. To categorize the main characteristics of the various SLR jobs, population sampling approaches are critical. Many studies have employed electronic gloves or cyber gloves to collect data on the mechanical and nonmechanical components of symptoms due

to the substantial dependence on sensor-based SLR systems. The usage of these sensors, on the other hand, is cumbersome and severely limiting for the signatory [13]. Furthermore, practical implementations of sensor-based SLR systems are problematic because of the high cost of sensors. On the other hand, because of their weight and through, dynamic heterogeneous environment, and capacity to

# Sindh Welfare Association of the Deaf (Regd.)

## THE ALPHABET IN SIGN LANGUAGE (ENGLISH) SINGLE - HAND SYSTEM



**SWAD Office: H-360, 1st Floor, Sector 32/A, Zia Colony, Korangi No.1, Karachi East**  
**E-mail: swadeaf@yahoo.com, swadeaf@gmail.com**

FIGURE 4: Single-hand English alphabets for deaf persons (Sindh Welfare Association of the Deaf).

control changes in distribution phase under varying illumination and limitations, visionary SLR systems have wowed researchers [14]. The SLR solution's standard feature automatically enables signer-dependent actions, which means that all signers are educated before working with the patient. Signer independence, also known as cross-signer verification, entails the normalizing of features to eliminate confuser interactivity. The line

between certain signatures and the camera is rarely evident, as is the currency of signature and magnification [15]. SLR's early phases were likened to speech recognition in that they concentrated on symptoms. Although several SLR approaches for identifying continuous phrases have been established, the recognition accuracy for short dictionaries employing epithelial movement between the symbols is barely 90%. Researchers are now

working on developing an image-based system that can receive live stream photos and tell or write the sign language alphabets used by the deaf.

### 1.1. Contributions

- (i) In this paper, we proposed an image-based sign language classification and recognition model to capture the image of the sign of the deaf person and recognize it and then tell the number in written and voice as well.
- (ii) The proposed model is loaded into the system, and then the configuration is loaded or set with parameters like max number of hands and confidence level. The file with an index of names is loaded to show the human understandable output on screen. Open Computer Vision comes now and gets images from files or cameras (live feed). Then, the frame is prepared for detection and recognition.
- (iii) We used Python 3 language for programming, Open Computer Vision, TensorFlow, and Keras for detection of images, and Open Computer Vision, TensorFlow, Keras, Pandas, and sklearn for module training.
- (iv) The proposed model accurately detects and recognizes the sign language Urdu alphabets.

The following is a breakdown of the paper's structure. Section 2 is devoted to a review of the literature. The problem is outlined in Section 3. The proposed solution is shown in Section 4. The simulation and methodology are provided in Section 5, and results are covered in Section 6. The paper's discussion and conclusion are included in Section 7.

## 2. Literature Review

The disabled persons or special persons such as deaf people cannot communicate with normal person efficiently. Therefore, sign language can help people with disabilities to communicate. In sign language, several sorts of motions with various shapes are utilized [16]. Similarly, sign languages vary by geography, and there are currently 138 recognized sign languages. The British-American Sign Language is based on English, while Chinese and Indian Sign Languages are also growing in popularity. Because sign language is focused on forms and concepts while spoken and written languages are based on words and grammar, sign language grammar is mostly based on written and spoken language grammar. As a result, the two languages have different grammatical structures. Information technology has had a significant impact on human life. To assist humanity in solving various challenges, many technologies, techniques, and instruments have been invented [17]. Information technology has been utilized to overcome the communication gap between deaf and hearing people. The idea behind these IT-based technologies is to help deaf individuals interact more effectively with persons who have impairments and vice versa. IT-based technologies like these might be useful in overcoming this communication divide.

Many developed countries solved the issue of miscommunication of deaf with normal people by using information technology (IT). By using IT, most of the issues of miscommunication are solved efficiently in this modern era [18]. In Pakistan, there is no such software or other IT technologies to solve the Urdu communication among deaf and normal people. As a result, we require such a computer-based system to communicate with deaf people and to decrease human involvement during the deaf people's understanding of everyday life activities. Using popular literature, reduce and recommend communication gaps for Pakistan's deaf population and find essential processes for building an architectural framework based on information technology programmed that can also help bridge the gap between the deaf and the country's public [19]. Mobile phone computing, gesture-based environments, and cloud technologies are all part of modern technology. Artificial Intelligence technologies such as Leap Motion Kinect, Google Glass, and Leap Motion Controller are using to capture the gesture of hands and other disabled parts of human body [20]. As previously said, technological advancements may be leveraged to help deaf individuals. Communication is a huge issue for deaf children since they are unable to interact with society because they are unable to speak normally. The learning environment for deaf students at educational institutions is not necessarily the same as for hearing students. As a result, sign language is one of the most effective ways for deaf individuals to communicate. Signature conversations cannot be understood by an expert or colleagues using a gesture-based interface, which causes communication challenges between the deaf and the public [21]. Communication among deaf and normal people is the ultimate goal of the modern research. Sensors, gadgets, and image-based methods are all utilized. HCI, robotics, game-based learning, and login software recognition system have all received more attention in recent research. Programs and systems for language recognition are employed. The PC vision community uses a variety of procedures and algorithms [22]. Efforts are already been made by the researchers in the field of communication among deaf and normal people. So, most of the best algorithms/software are the results of such research studies in the field of gesture recognition and communication for disabled people [19]. The game encourages users to engage with a virtual environment, allowing them to learn sign language in a novel and enjoyable way. It is regarded as a language of Portuguese Sign Language [23] to use sign language signals to consolidate similar aims and purposes. ISL translation system and Indian Sign Language (ISL) translation device for learning sign language pictures or continuous video images (of the public) are captured using a microphone or USB camera and can be interpreted by an application. It is envisaged that the obtained expressions will be translated, scaled, and disseminated. Image capture, identification of the binary type of hand, and finding shape and function are variables in the approach and interpretation processes. The message is displayed and sent to the recipient in the built-in form by the GUI program. It compels normal individuals to converse freely with deaf people [24]. With the use of image-based

approaches, the spoken prefixes are transformed to sign languages utilizing a computer-based system. With different degrees of success, this translation project has an exceptional employment offer. In [25], the authors presented sign languages of different countries such as United States, Greece, South Africa, Arabic countries, Spain, Italy, Japan, Netherlands, and United Kingdom. When it comes to people who cannot hear properly, sign language (LS) is a valuable resource. East Optical motion bidding is another name for the approach. This strategy is used by those who do not comprehend adequately. Language serves as the primary means of communication. Every country has its own set of signs. China, the United States, India, and Pakistan, for example, all have their own sign languages. Indian Sign Language and Pakistani Sign Language are two different sign languages. Many developing countries hold seminars on the subject. To close the gap, they organize a variety of project activities, including information technology. For those who are deaf and people who are not deaf, several surveys have been undertaken throughout Central and South Asia. Nonetheless, this approach is being investigated in Pakistan because there is no systematic information on it. In [18], the authors proposed grammar, content, and delivery tools used in Pakistan for disabled people communication as the main communication of deaf take place on the basis of language structure. However, the primary point has been made thus far. The goal of this study is to talk about the difficulties that need to be addressed to close the gap between the general and deaf communities. They provide several suggestions for constructing a bridge. Sign language follows a distinct set of norms than spoken and written languages. Sign language is a kind of communication that is used to communicate. It is built on forms, and written language is built on certain fundamental word construction and grammatical rules [26]. Information technology has a significant influence on our lives; individuals create many of the items we use. In India, exceptional labor has been done to pave the path for extraordinary individuals. The most significant challenge they encounter is their inability to communicate with others. People have taken up the role of virtual or effective language [27]. Pakistani academics are also focused on developing different assistive technologies for persons with impairments, such as a sensory glove and transliteration of American Sign Language [19]. The above automated device, dubbed "talking hand," was built to promote communication between the public and those who were impacted. The author employed artificial neural networks to receive sensory information utilizing gloves in the suggested method. The technique was utilized for 24 letters of the English alphabet and two punctuation marks by the author. As a result, deaf people who use the software can use it to communicate [28]. The authors did a similar study on the connection between deaf and handicapped individuals in [29], with the gloves equipped with sensors that detect spoken messages via finger movement. This is a handy program that converts the alphabet to text and speech [30]. The authors employed a two-based jump motion device technique to create letters in Pakistani Sign Language. The "communication module" is a module that trains one system

for translation while the other gathers information using a jump motion device. The authors of [31] developed a gadget that translates physical signals to digital data and issues the necessary instructions for deaf people. The gadget uses symbols to transmit data to the computer alliance. This is a visual art that has been utilized in Pakistan before, such as a glove for deaf people. In Pakistan, there is a lot of effort being done for deaf persons who desire to communicate with regular people. The authors of [32] created a tool called an Ambiguous Classifier. Gloves are used to identify fingers, and this instrument indicates the deaf person's symptoms or the color of the operation. This approach has a 95% accuracy rate. Many academics are attempting to bridge the gap between deaf and hearing persons, and the authors of [18] developed a machine learning-based model that might serve as a foundation for comprehending signal-based communication. As a result, the authors presented a model for deaf individuals in Pakistan that assists them in translating English or Urdu texts or speaking in Pakistani Sign Language [18]. The authors proposed a study in [10] with the goal of producing a tool to assist persons with impairments as well as building a program to allow deaf people to conduct normal discussions with other deaf people. Enter sign language as text and make sure the other person can comprehend it. The authors of [33] used the GoogLeNet pretrained architecture on the ILSVRC dataset, which is based on convolutional neural networks and uses ASL datasets from Macy's Store. The University of Surrey employed transfer learning for this purpose. They created a solid model that identifies the letters and works well with newcomers. Fully generalizable translators may be created for any ASL publications. Deep sensor technology is fast gaining popularity, as are other instruments used in this process, which have shown to be successful, such as colorful advertisements like custom-designed gloves. Its purpose is to make the identifying process easier and more efficient. According to [34], certain signal units are simple to classify and identify. To date, automatic gesture language recognition systems have not been able to make use of today's deep detection technologies. Previously, only single camera technology was employed. There are simply pixels in basic picture datasets with no depth or contour information, although classifying images of ASL letter movements using CNN have had some success [33] but utilizing GoogLeNet architects who are already trained [35], a repeated deep structure has been suggested. Continuous sign language recognition uses a convolutional neural network. We have devised a step-by-step solution. How to educate our deep neural network's structure is the focus. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have made significant progress in signaling [36] and sign recognition [37]. In sign recognition, dynamic mundane dependency learning yielded significant results [38].

The authors used image processing and machine learning techniques to produce a detailed literature evaluation on Urdu Sign Language in [39]. Based on [40, 41] and the preceding literature assessment, the following difficulties in the field of Urdu Sign Language remain to be addressed:



- (i) Image processing and deep learning-based computerized system is required for deaf people to communicate with the normal people in the society.
- (ii) We have a shortage of people who are expert in Urdu Sign Language and who can take part as the middle entity between deaf and normal persons.
- (iii) To remove the physical expert and perform that work by computer or any other electronic device, we need an efficient intelligent system to understand and communicate with deaf people for the betterment of the human society.

### 3. Problem Formulation and Gap Analysis

As in the literature review such as [39], the authors presented that there is no such computer-based system for Urdu sign language that could make possible the communication between deaf and normal person. So, in this research, we intend to work on such computer-based system that could help the deaf ones' communication normal ones. In Pakistan, the Urdu sign language is used by deaf people, but the normal people did not understand their communication, as sign language is not studied by everyone. So, we intend to propose such system that pick the sign of the deaf one and sound the word in voice by the computer to understand by the normal one.

### 4. Proposed Solution

In front of the camera, the deaf person shows the symbol he/she wants to communicate, and then the camera captures the picture of the symbol and sends it to another machine where image processing technique is used to classify and recognize the image. After recognition, the deep learning technique is used to train the model for future correspondence and send it to the main CPU for preprocessing.

In this section of the paper, we provide a detailed explanation of the proposed model with the help of flowcharts and diagrams.

In Figure 5, the proposed model is provided. As shown in the figure, the proposed model is based on Python 3, image processing model "Open Computer Vision," and deep learning models such as "linear regression" and "TensorFlow" using convolutional neural network. After preprocessing, the image is shown with the desired value that the deaf person wants to communicate. The following entities are used in developing the proposed model as shown in Figure 5.

**4.1. Convolutional Neural Network (CNN).** We used convolutional neural network in the proposed model. A convolutional neural network is a feed-forward neural network that processes input in a grid-like structure and is commonly used to evaluate visual pictures. ConvNet is another name for it. To recognize and categorize items in a picture, a convolutional neural network is employed. Multiple layers of artificial neurons make up convolutional neural networks. Artificial neurons are mathematical functions that calculate the weighted sum of various inputs and output an activation

value, like their biological counterparts. Each layer of a ConvNet creates numerous activation maps when you input a picture into it. The important aspects of the picture are highlighted using activation maps. Each neuron takes a patch of pixels as input, multiplies their color values by their weights, adds them all together, and then runs them through the activation function. Basic characteristics such as horizontal, vertical, and diagonal edges are often detected by the CNN's first (or bottom) layer.

The first layer's output is sent into the second layer, which extracts more complicated characteristics like corners and edge combinations. Each neuron's action is determined by its weight. When given pixel data, CNN's artificial neurons pick out numerous visual features. Basic characteristics such as horizontal, vertical, and diagonal edges are generally detected by the CNN's first (or bottom) layer. The first layer's output is sent into the next layer, which extracts more complicated characteristics like corners and edge combinations. The layers recognize higher-level characteristics such as objects, faces, and more as you progress further into the convolutional neural network. The CNN model is shown in Figure 6. The following hidden layers are used by the CNN architecture to classify and recognize the images or text.

**4.1.1. Convolution Layer.** This is the initial stage in obtaining useful information from a photograph. The convolution action is performed by many filters in a convolution layer. Every image is seen as a pixel value matrix. Also, we get the image we wanted as a result.

**4.1.2. ReLU Layer.** The rectified linear unit is abbreviated as ReLU. After the feature maps have been extracted, they must be moved to a ReLU layer. ReLU conducts an element-by-element procedure, setting all negative pixels to zero. It causes the network to become nonlinear, and the result is a corrected feature map.

**4.1.3. Pooling Layer.** Pooling is a downsampling process that decreases the feature map's dimensionality. To create a pooled feature map, the corrected feature map is now sent via a pooling layer. The pooling layer employs a variety of filters to recognize various aspects of the picture, including edges, corners, the body, feathers, eyes, and the beak.

**4.1.4. Fully Connected Layer.** Feed-forward neural networks are what the fully connected layer is all about. Fully connected levels are the network's final layers. The output from the final pooling or convolutional layer, which is flattened and then fed into the fully connected layer, is the input to the fully connected layer. We used ResNet model of the CNN methodology.

**4.2. ResNet.** Residual neural network (ResNet) by Kaiming et al. introduced a novel architecture with "skip connections" and features heavy batch normalization. Such skip connections are also known as gated units or gated recurrent units and have a strong similarity to recent successful

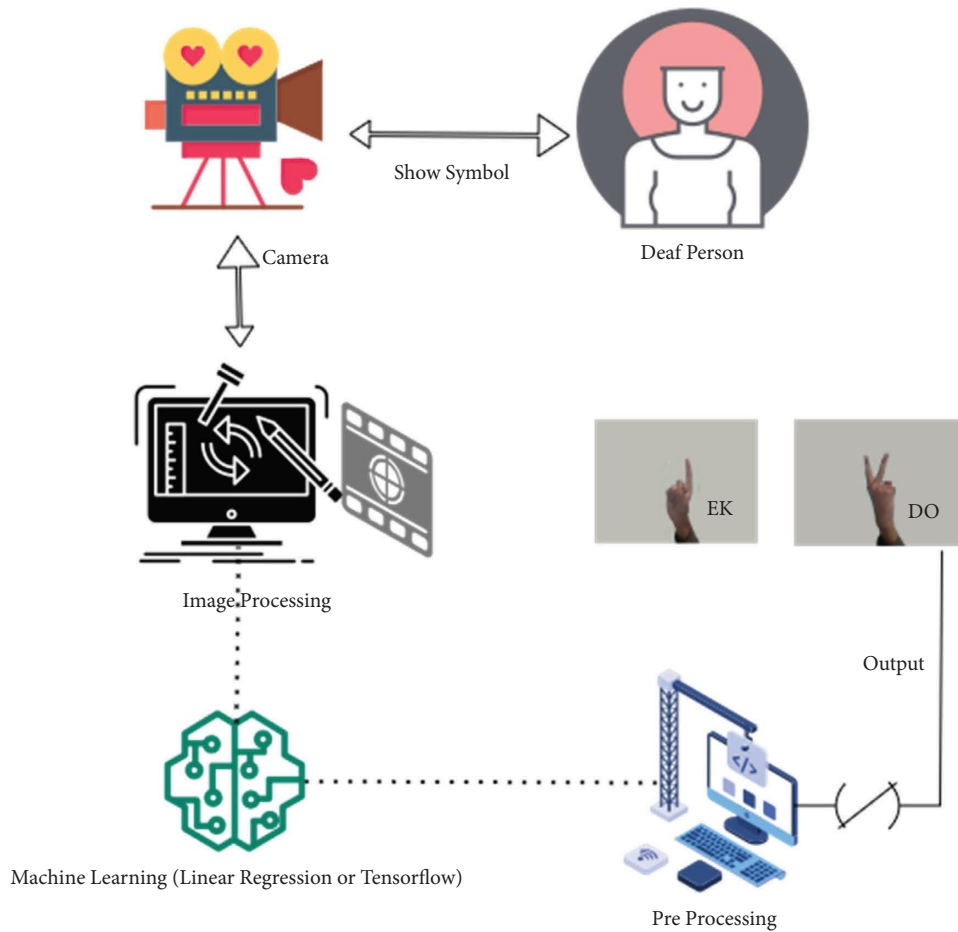


FIGURE 5: Proposed image processing-based solution.

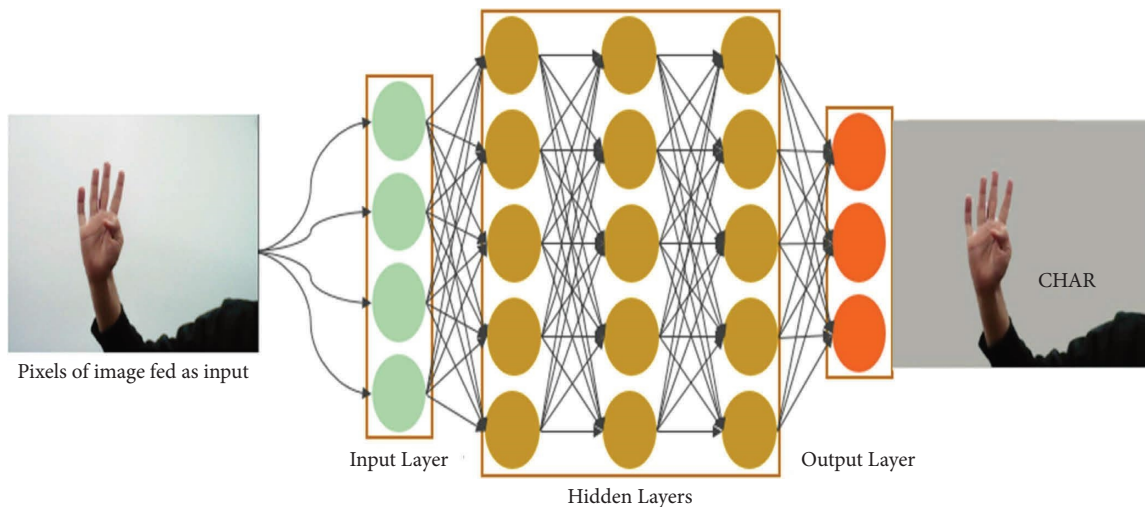


FIGURE 6: CNN architecture for the proposed model.

elements applied in RNNs. Thanks to this technique, they were able to train a NN with 152 layers while still having lower complexity than VGGNet. It achieves a top-5 error rate of 3.57% which beats human-level performance on this dataset. AlexNet has two parallel CNN lines trained on two GPUs with

cross-connections, GoogLeNet has inception modules, and ResNet has residual connections. The library of CCN which is called ResNet is used to train the sign language model to acquire better results. Figure 7 shows the ResNet architecture for the image extraction and classification.

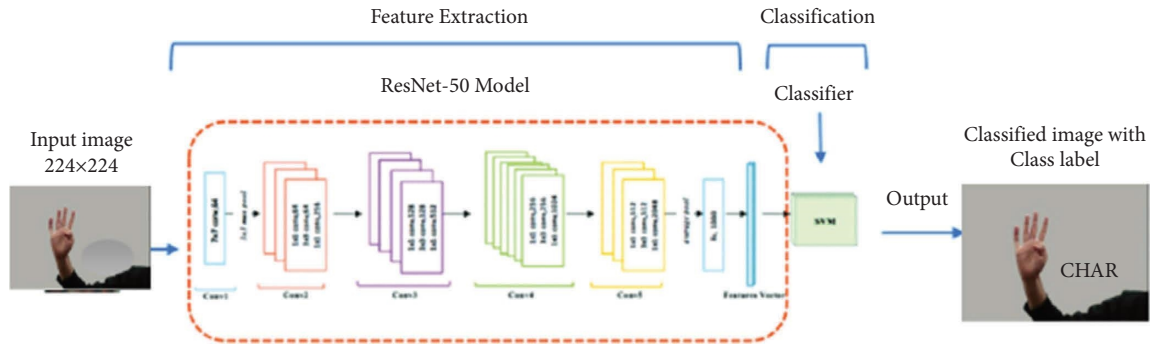


FIGURE 7: ResNet architecture.

**4.3. Camera.** A camera is a visual instrument for capturing images. The camera comprises a sealed box (camera body) with a tiny hole (aperture) that allows light to be captured onto a light-sensitive surface at its most basic level (usually photographic film or a digital sensor). To regulate how light falls on a light-sensitive surface, cameras use a variety of methods. The camera's lenses concentrate the light that enters it. The aperture can be made smaller or larger. The length of time a light-sensitive surface is exposed to light is determined by a shutter mechanism. In the art of photography, the steel image camera is an essential tool. Photographic, digital imaging, and photo printing images can be recreated afterwards. Film, videography, and cinematography are among the creative genres covered by the motion picture camera.

**4.4. Programming Language.** Python was the programming language that we utilized. Python is a popular general-purpose programming language with a lot of flexibility. Python is an object-oriented programming language, making it ideal for quick application development. Python's basic syntactic value is easy to comprehend, which reduces the cost of software security. Modules and packages are supported by Python, which fosters program modularity and code reuse. Python language and large standard library for all main platforms are accessible in free source code or binary format.

**4.5. Open Computer Vision.** As an image processing-based model, we employed Open Computer Vision. Open Computer Vision is a collection of programming functions for computer vision that is primarily intended for real-time use. In a word, it is an image processing library. It is primarily utilized to perform all image-related operations. More than 2500 complex algorithms are included in the library, including a comprehensive range of both traditional and contemporary machine learning and computer vision techniques.

**4.6. TensorFlow.** The TensorFlow deep learning model was employed. TensorFlow is a framework for graphing and analyzing complicated complexities. Multiple rows known

as model tensors are needed to do this. It is a tool for TensorFlow training. Deep learning-based PSL dataset is utilized for both research and development at the same time.

Table 1 shows the abbreviations used in the research.

## 5. Simulation and Methodology

To implement the concept of Urdu Sign Language symbols by computer, we used Core i5 laptop with Windows 10. We used 8 GB RAM and 256 GB SSD card. Urdu Sign Language's different symbols were provided as input to the dataset for training the model. The authors used Python-based programming environment and embedded the image processing and deep learning techniques such as OpenCV and TensorFlow. The NumPy IDE is used for the programming setup. NumPy stands for "Numerical Python" and is a Python module. It was used to calculate numerical values. Two datasets are used as number system symbols provided by Sindh welfare association and alphabet symbols used for deaf people. It has many logarithmic processing operations, including differentiation and integration. Multidimensional array is used to store the information sign language in the computer memory. The model is imported into the system, followed by the setup, which includes parameters such as the maximum number of hands and the confidence level. To display the human-readable output on the screen, a file with a name index is loaded. Open Computer Vision comes now and gets images from files or cameras (live feed). Then, the frame is prepared for detection and recognition. The model detects and draws the landmarks on that frame. The prediction model takes these landmarks as input and outputs the prediction as class id which is then later matched with the index of name file. Table 2 is used for simulation setup of the proposed model, and the parameters are referenced from [38–40].

Table 2 is used for simulation setup for the evaluation and implementation of the proposed system. Programming language is defined and provided; if anyone wants to work further in the research area, then he/she can know about the programming language used in the research. Other libraries of Python which are used in the proposed research are provided in Table 2. After completion of the proposed system, we feed live video to the model, and images shown in Figure 6 were drawn. As

TABLE 1: Abbreviation table.

Abbreviation	Full form
PSL	Pakistan Sign Language
ReLU	Rectified linear unit
CNN	Convolutional neural network
ASL	American Sign Language
NIDCD	National Institute on Deafness and Other Communication Disorders
BSL	British Sign Language
ArSL	Arabic Sign Language
SSL	Spanish Sign Language
SLR	Sign language recognition
IT	Information technology
ISL	Indian Sign Language
HCI	Human-computer interaction
PSL	Portuguese Sign Language
PC	Personal computer

TABLE 2: Simulation setup components.

S. No.	Field	Component
1	Programming language	Python 3
2	Modules for detection	Open Computer Vision, TensorFlow, Keras
3	Modules for training	Open Computer Vision, TensorFlow, Keras, Pandas, sklearn

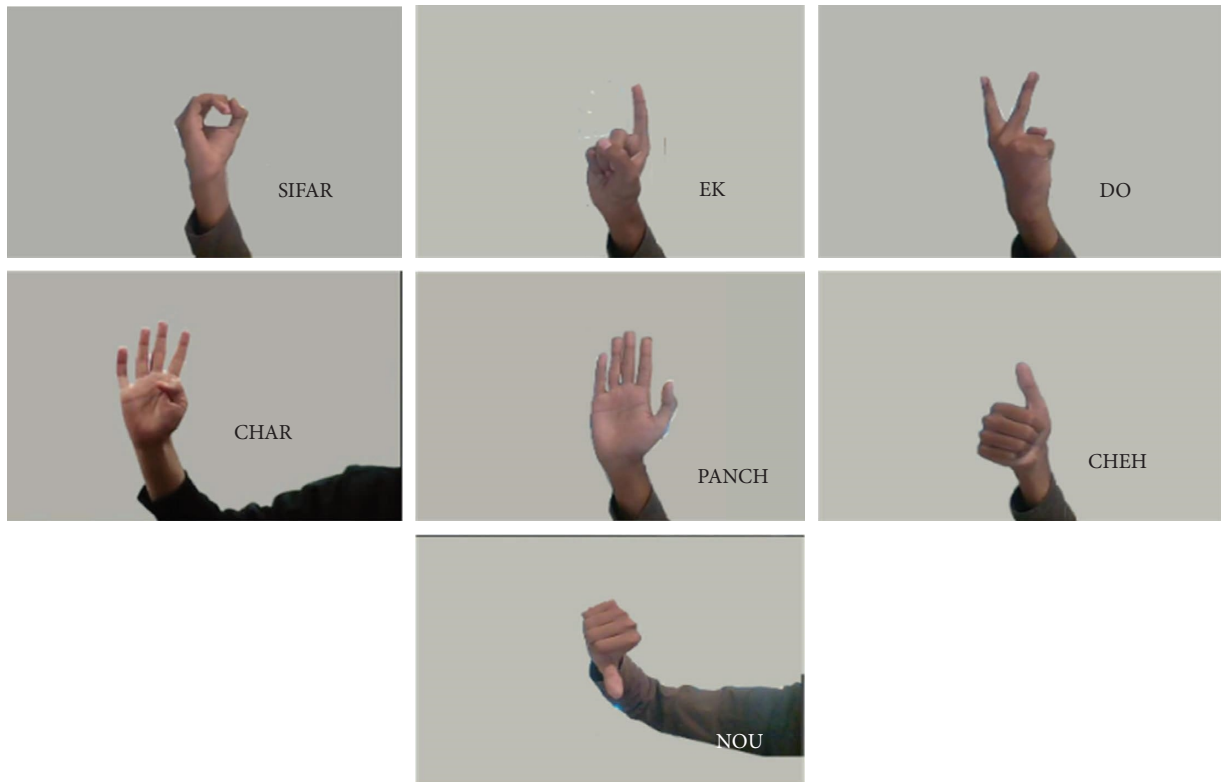


FIGURE 8: Live video feed results from the proposed model.

shown in Figure 8, the proposed model is evaluated based on live feed video, and when we showed the Urdu Sign Language symbols one by one, the proposed model recognized the symbols and provided them in written form on real-time basis.

## 6. Results

Furthermore, we trained the model in different simulations runs, and the following results were obtained as shown in Figures 9–12.

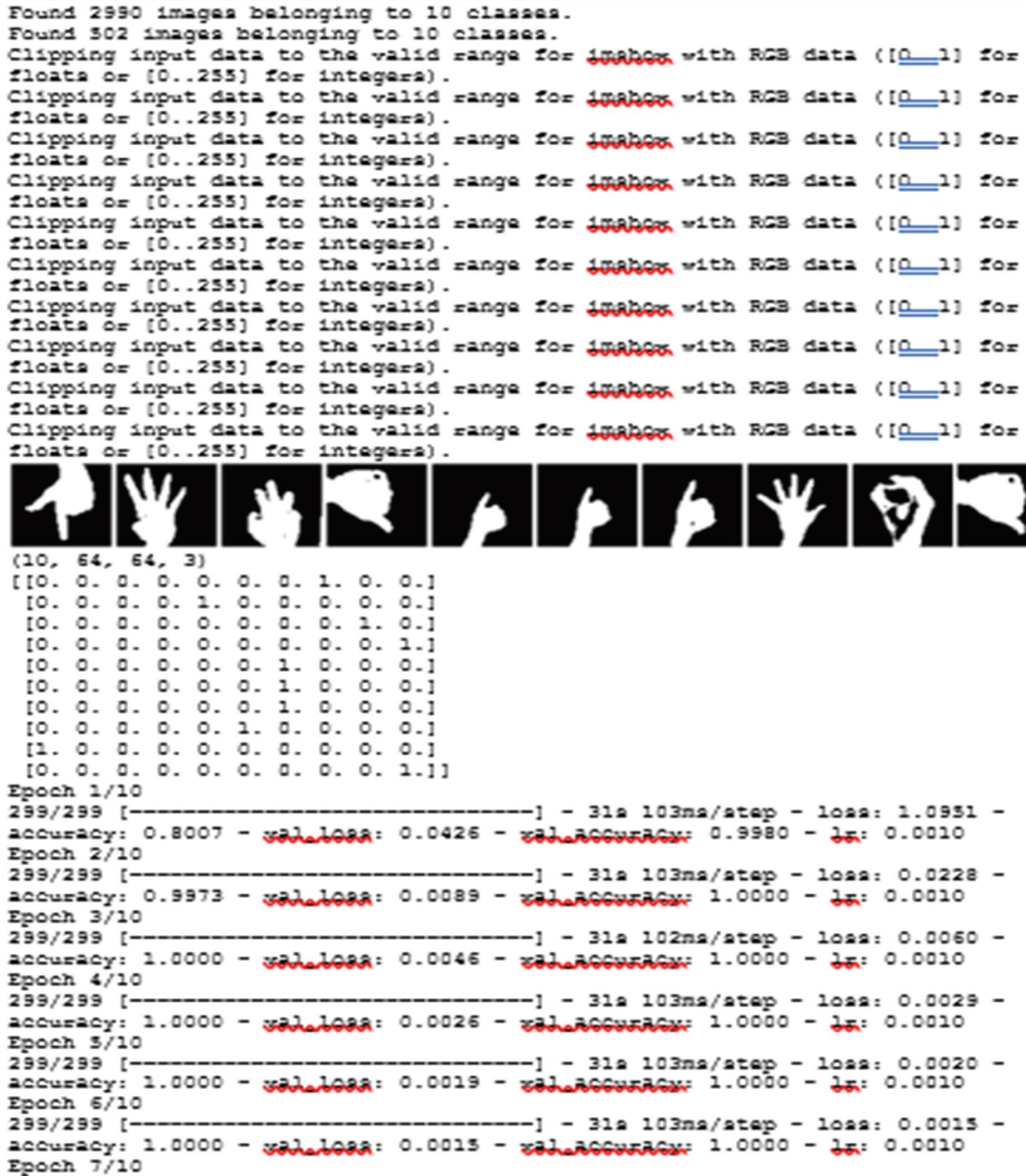


FIGURE 9: Result of the proposed model 1.

As shown in Figures 7–10, the proposed model accurately recognized the Urdu Sign Language symbols from the live video, and hence it can be told that the proposed model is an efficient model for deaf people to communicate with normal people. Furthermore, the accuracy of the proposed is evaluated and provided in the following figures.

As shown in Figure 13, the proposed model’s accuracy with linear regression was 100% while using TensorFlow, the

accuracy was 97% as some of the pictures were not recognized correctly.

As compared with the latest work, our proposed model performs very well as the previous work provided the accuracy up to 80% and we improved the accuracy up to 100%, and the proposed model can easily convert the Urdu Sign Language symbols into voice and alphabets. In Figure 14, RMSE of the proposed model is provided. When the model initially started its training, the root mean square error was

```

299/299 [-----] - 31a 104ms/step - loss: 0.0012 -
accuracy: 1.0000 - val_loss: 0.0011 - val_accuracy: 1.0000 - lr: 0.0010
Epoch 8/10
299/299 [-----] - 31a 103ms/step - loss: 9.7736e-04
- accuracy: 1.0000 - val_loss: 9.6149e-04 - val_accuracy: 1.0000 - lr: 0.0010
Epoch 9/10
299/299 [-----] - 31a 105ms/step - loss: 8.1869e-04
- accuracy: 1.0000 - val_loss: 8.1612e-04 - val_accuracy: 1.0000 - lr: 0.0010
Epoch 10/10
299/299 [-----] - 32a 106ms/step - loss: 7.1409e-04
- accuracy: 1.0000 - val_loss: 7.2608e-04 - val_accuracy: 1.0000 - lr: 0.0010
loss of 0.00029752490809187293; accuracy of 100.0%
{'loss': [1.0950770378112793, 0.02283531054854393, 0.00596899027004838,
0.002935325726866722, 0.001999481115490198, 0.0015140236353364323,
0.0011748004471883178, 0.0009775563376024365, 0.0008186884806491435,
0.0007140882662497461], 'accuracy': [0.8006688932445984, 0.9973244071006775,
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0], 'val_loss': [0.04260633513331413,
0.008947278372943401, 0.004601104184985161, 0.00264947721734643,
0.0018566565122455359, 0.0014645023038610816, 0.0011406627018004656,
0.0009614907903596759, 0.000816119194496423, 0.0007260802667587996],
'val_accuracy': [0.9980079531669617, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0,
1.0], 'lr': [0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001,
0.001]}
loss of 0.001286243787035346; accuracy of 100.0%
Model: "sequential_7"

```

Layer (Type)	Output Shape	Param #
conv2d_21 (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d_21 (MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_22 (Conv2D)	(None, 31, 31, 64)	18496
max_pooling2d_22 (MaxPooling2D)	(None, 15, 15, 64)	0
conv2d_23 (Conv2D)	(None, 13, 13, 128)	73856
max_pooling2d_23 (MaxPooling2D)	(None, 6, 6, 128)	0
flatten_7 (Flatten)	(None, 4608)	0
dense_28 (Dense)	(None, 64)	294976
dense_29 (Dense)	(None, 128)	8320
dense_30 (Dense)	(None, 128)	16512
dense_31 (Dense)	(None, 10)	1290

```

Total params: 414,346
Trainable params: 414,346
Non-trainable params: 0

```

FIGURE 10: Result of the proposed model 2.

```
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
predictions on a small set of test data--
```

```
Three Six Four One Eight Seven Nine Three Seven One
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
Clipping input data to the valid range for image with RGB data ([0..1] for
floats or [0..255] for integers).
```



```
Actual labels
Three Six Four One Eight Seven Nine Three Seven One (10,
64, 64, 3)
```

```
{'accuracy': [0.8006688952445984,
0.9973244071006775,
1.0,
1.0,
1.0,
1.0,
1.0,
1.0,
1.0,
1.0],
'loss': [1.0950770378112793,
0.02283531054854393,
0.00396899027004838,
0.002935325726866722,
0.001999481115490198,
0.0015140236355364323,
0.0011748004471883178,
0.0009775363376024365,
0.0008186884806491435,
0.0007140882662497461],
'lr': [0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001],
'val_accuracy': [0.9980079531669617,
1.0,
1.0,
```

FIGURE 11: Result of the proposed model 3.

```
1.0,
1.0,
1.0,
1.0,
1.0,
1.0,
1.0],
'val_loss': [0.04260633513331413,
0.008947278372943401,
0.004601104184583161,
0.00264947721734643,
0.0018566365122455359,
0.0014643023038610816,
0.0011406627018004656,
0.0009614907903596759,
0.000816119194486423,
0.0007260802667587996]}
```

FIGURE 12: Result of the proposed model 4.

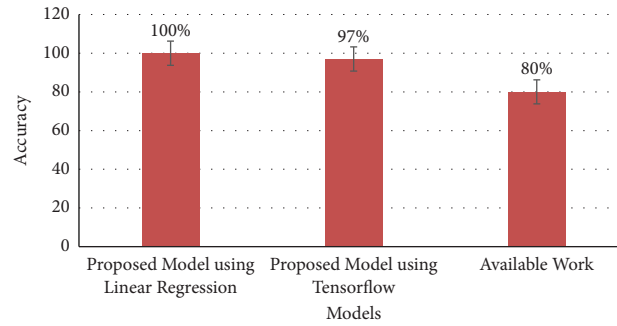


FIGURE 13: Accuracy of the proposed model.

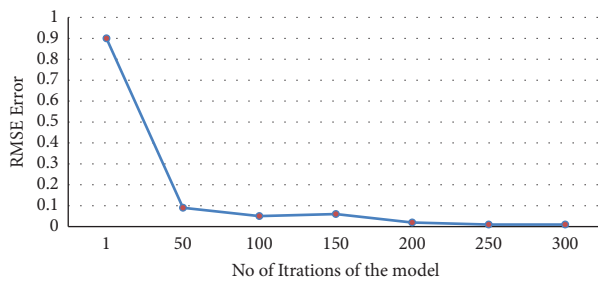


FIGURE 14: Root mean square error.

high, while after some iterations, the error decreased as shown in Figure 14.

## 7. Conclusion

In this paper, we provided an image processing and deep learning-based model for Urdu Sign Language. The proposed model is implemented in Python 3 and used different image processing and machine techniques to capture the video and transform the symbols into voice and Urdu writing. First, we get a video from the deaf person, and then the model crops the frames in pictures. Then, the individual picture is recognized for the sign symbol such as if the deaf showed a symbol for one, then the model recognizes it and shows the letter which he/she wants to tell. Image processing technique such as OpenCV is used for image recognition and classification while TensorFlow and linear regression are used for training the model to behave intelligently in the future. The results show that the proposed model increased accuracy from 80% to 97% and 100% accordingly. The accuracy of the previously available work was 80% when we implemented the algorithms, while with the proposed algorithm, when we used linear regression, we achieved the highest accuracy. Similarly, when we used the TensorFlow deep learning algorithm, we achieved 97% accuracy which was less than that of the linear regression model. In future work, we need to implement and develop a mobile-based model to help the deaf people in normal societies.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This project was sponsored by the Department of Computer Science and Engineering, University of Liberal Arts Bangladesh (ULAB), Dhaka, Bangladesh

## References

- [1] E. Emerson and C. Hatton, *Health Inequalities and People with Intellectual Disabilities*, Cambridge University Press, Cambridge, UK, 2014.
- [2] A. Kuenburg, P. Fellinger, and J. Fellinger, "Health care access among deaf people: table 1," *Journal of Deaf Studies and Deaf Education*, vol. 21, no. 1, pp. 1–10, 2016.
- [3] A. Kumar, S. Kumar, S. Singh, and V. Jha, "Sign Language recognition using convolutional neural network," in *ICT Analysis and Applications*, pp. 915–922, Springer, 2022.
- [4] S. Vachmanus, A. A. Ravankar, T. Emaru, and Y. Kobayashi, "Multi-modal sensor fusion-based semantic segmentation for snow driving scenarios," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 16839–16851, 2021.
- [5] R. Elakkiya, "Retracted article: machine learning based sign language recognition: a review and its research Frontier," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 7, pp. 7205–7224, 2021.
- [6] S. Sharma and S. Singh, "Vision-based hand gesture recognition using deep learning for the interpretation of sign language," *Expert Systems with Applications*, vol. 182, Article ID 115657, 2021.
- [7] A. Khelalef, F. Ababsa, and N. Benoudjit, "An efficient human activity recognition technique based on deep learning," *Pattern Recognition and Image Analysis*, vol. 29, no. 4, pp. 702–715, 2019.
- [8] R. Mishra and R. Subban, "Face detection for video summary using enhancement based fusion strategy," *International Journal of Renewable Energy Technology*, vol. 3, no. 15, pp. 69–74, 2014.
- [9] A. Middleton, S. D. Emery, and G. H. Turner, "Views, knowledge, and beliefs about genetics and genetic counseling among deaf people," *Sign Language Studies*, vol. 10, no. 2, pp. 170–196, 2010.
- [10] A. Wadhawan and P. Kumar, "Sign language recognition systems: a decade systematic literature review," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 785–813, 2021.



- [11] M. Oudah, A. Al-Naji, and J. Chahl, "Elderly care based on hand gestures using kinect sensor," *Computers*, vol. 10, no. 1, p. 5, 2020.
- [12] L. Meng and R. Li, "An attention-enhanced multi-scale and dual sign language recognition network based on a graph convolution network," *Sensors*, vol. 21, no. 4, p. 1120, 2021.
- [13] L. Quesada, G. López, and L. Guerrero, "Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 4, pp. 625–635, 2017.
- [14] N. Mohamed, M. B. Mustafa, and N. Jomhari, "A review of the hand gesture recognition system: current progress and future directions," *IEEE Access*, vol. 9, pp. 157422–157436, 2021.
- [15] R. Elakkiya and K. Selvamani, "Subunit sign modeling framework for continuous sign language recognition," *Computers and Electrical Engineering*, vol. 74, pp. 379–390, 2019.
- [16] D. K. L. Lee and P. Borah, "Self-presentation on Instagram and friendship development among young adults: a moderated mediation model of media richness, perceived functionality, and openness," *Computers in Human Behavior*, vol. 103, pp. 57–66, 2020.
- [17] S. Kafle and M. Huenerfauth, "Predicting the understandability of imperfect English captions for people who are deaf or hard of hearing," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 12, no. 2, pp. 1–32, 2019.
- [18] N. S. Khan, A. Abid, and K. Abid, "A novel natural language processing (NLP)-based machine translation model for English to Pakistan sign language translation," *Cognitive Computation*, vol. 12, no. 4, pp. 748–765, 2020.
- [19] A. Abbas and S. Sarfraz, "Developing a prototype to translate text and speech to Pakistan Sign Language with bilingual subtitles: a framework," *Journal of Educational Technology Systems*, vol. 47, no. 2, pp. 248–266, 2018.
- [20] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 131–153, 2019.
- [21] A. Mindess, *Reading between the Signs: Intercultural Communication for Sign Language Interpreters*, Nicholas Brealey, London, UK, 2014.
- [22] A. Rashid and O. Hasan, "Wearable technologies for hand joints monitoring for rehabilitation: a survey," *Microelectronics Journal*, vol. 88, pp. 173–183, 2019.
- [23] S. Hermawati and K. Pieri, "Assistive technologies for severe and profound hearing loss: beyond hearing aids and implants," *Assistive Technology*, vol. 32, 2019.
- [24] M. Mahesh, A. Jayaprakash, and M. Geetha, "Sign language translator for mobile platforms," in *Proceedings of the 2017 International Conference on Advances in Computing, Communications, and Informatics (ICACCI)*, pp. 1176–1181, IEEE, Mangalore, India, September, 2017.
- [25] B. L. Ludlow, "Virtual reality: emerging applications and future directions," *Rural Special Education Quarterly*, vol. 34, no. 3, pp. 3–10, 2015.
- [26] S. Žilič Fišer, I. Kožuh, and I. Kožuh, "The impact of cultural events on community reputation and pride in Maribor, the European Capital of Culture 2012," *Social Indicators Research*, vol. 142, no. 3, pp. 1055–1073, 2019.
- [27] S. Sankar Kumar, J. Jenitha, I. Narmadha, and A. Suganya, "An embedded module as "Virtual Tongue" for voiceless," *International Journal of Information Sciences and Techniques*, vol. 4, no. 3, pp. 155–163, 2014.
- [28] M. Naseem, S. Sarfraz, A. Ali, and H. Ali, "Developing a prototype to translate Pakistan Sign Language into text and speech while using convolutional neural networking," *Journal of Education and Practice*, vol. 10, 2019.
- [29] K. Kim, J. Choi, and S.-M. Lee, "Why does bundled product in telecommunication service market matter: evidence from South Korea," *International Journal of u-and e-Service, Science and Technology*, vol. 9, no. 3, pp. 209–226, 2016.
- [30] N. Raziq and S. Latif, "Pakistan sign language recognition and translation system using leap motion device," in *Proceedings of the International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 895–902, Springer, Asan-si, Korea, November, 2016.
- [31] A. Fatima and K. Huma, *Image Based Pakistan Sign Language Recognition System*, 2011.
- [32] S. Kausar, M. Y. Javed, and S. Sohail, "Recognition of gestures in Pakistani sign language using fuzzy classifier," in *Proceedings of the 8th Conference on Signal Processing, Computational Geometry and Artificial Vision*, pp. 101–105, World Scientific and Engineering Academy and Society (WSEAS), Rhodes, Greece, August, 2008.
- [33] B. Garcia and S. A. Viesca, "Real-time American sign language recognition with convolutional neural networks," *Convolutional Neural Networks for Visual Recognition*, vol. 2, pp. 225–232, 2016.
- [34] D. S. Quentin, H. Wannous, and J. P. Vandeborrel, "Skeleton-based dynamic hand gesture recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–9, Las Vegas, NV, USA, June, 2016.
- [35] O. Koller, N. C. Camgoz, H. Ney, and R. Bowden, "Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2306–2320, 2020.
- [36] J. Lien, N. Gillian, M. E. Karagozler et al., "Soli: ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–19, 2016.
- [37] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," *IEEE Access*, vol. 5, pp. 23626–23636, 2017.
- [38] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 115, 2016.
- [39] H. Zahid, M. Rashid, S. Hussain, F. Azim, S. A. Syed, and A. Saad, "Recognition of Urdu sign language: a systematic review of the machine learning classification," *PeerJ Computer Science*, vol. 8, p. e883, 2022.
- [40] M. P. Kane, S. Fernandes, R. Fonseca, S. Desai, A. Shetye, and A. Sharma, "Sign Language apprehension using convolution neural networks," in *Proceedings of the 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–7, IEEE, Kharagpur, India, October, 2022.
- [41] E. Fatima, W. Naeem, and I. Abbas, "The influence of gender on the discourse markers in pakistani sign language," *Pakistan Journal of Scientific Research*, vol. 4, no. 2, pp. 1201–1207, 2022.