International Journal of Computers
http://www.iaras.org/iaras/journals/ijc

Communication Reimagined: Hands in Motion and the Future of Gesture Recognition Systems

ENACHI ANDREI, TURCU CORNELIU, CULEA GEORGE, SGHERA BOGDAN-CONSTANTIN, UNGUREANU ANDREI-GABRIEL
Faculty of Electrical Engineering and Computer Science,
Ştefan cel Mare University of Suceava, Suceava 720229,
ROMANIA
Department of Energetics and Computer Science,
Vasile Alecsandri University of Bacău, Bacău 600115,
ROMANIA

Abstract: - By integrating real-time recognition systems into modern communication frameworks, it changes the way humans interact with technology in one way or another. This paper presents the development and design of an innovative gesture recognition solution using state-of-the-art technology in an intuitive, simple and efficient interface for communication. Using specialized equipment and software, but also taking into account recent discoveries and advances in artificial intelligence, the study aims to cover accessibility, human-computer interaction and redefining non-verbal communication for members of deaf-mute communities. From sign language to gesture-based device control, recognition systems have the potential to revolutionize industries such as medicine and even gaming. It also addresses the different challenges such a system faces in terms of accuracy, real-time image sequences processing, adaptation to various environmental conditions and solutions to overcome them by ensuring a balance between precision and functionality. Finally, recognizing gestures is an important step for the future, where communication is breaking through traditional communication to the point where gestures are the ones that speak more than words.

Key-Words: hand gesture, OpenCV, MediaPipe, Tkinter, artificial intelligence, webcam, real-time, machine learning, data sets.

Received: April 6, 2024. Revised: December 9, 2024. Accepted: January 11, 2024. Published: March 14, 2025.

1 Introduction

An important step in the field of artificial intelligence is gesture recognition. This implies the inclusion of as many users as possible in a system but also the analysis and natural identification of hand gestures or sign language. In this work is proposed a system that has the ability to recognize in real time the gestures made by users in different environments but also the interaction and inclusion of people from deaf-mute communities.

The experience gained on the basis of this study is vast and growing and also offers the possibility of overcoming traditional communication and easing the facilitation of information by converting sign language into text. These key points are very important for members of the deaf-mute community, as they allow easier interaction with the digital environment as a result of the robustness of the proposed system and also its adaptability under various conditions.

The proposed system uses two important libraries, namely OpenCV and MediaPipe, to analyze and detect the images captured by the camera. This system will ensure technological accessibility without involving sophisticated hardware devices, which will lead to an optimal solution for future applications in this field. The images are captured by the camera continuously, in real time, in order to determine the key elements of the hand, such as the wrist or its joints, followed by the classification of the gestures made by the user based on the algorithms that make up the system.

It was also mentioned earlier that the last particularity that the system addresses is the conversion from sign language to writing, which involves advanced algorithms and models for better results. Various tests have been performed on the robustness of the system in several working environments and the detailed information about the system component, the implementation and the results obtained will be detailed in the future sections of this paper.

A significant impact of the system are the contributions it makes, in terms of communication between all people and in particular members of the deaf-mute community. Also, the results reflect the growing possibilities in the field of motion recognition belonging to artificial intelligence, providing aspects related to development and social inclusion in a digitized era. Initially, the dataset was designed with only five gestures and then twenty-five new gestures were trained for a total of thirty different gestures. This study differs from the other existing approaches because it utilizes the combination of the two libraries mentioned above with the possibility of adapting to different working environments under the incidence of light intensity and contrast, independent of dedicated hardware. The importance of this study lies in its accessibility and possibility of integration on different equipment and platforms, as opposed to sensor or camera-based approaches. Based on the set of gestures and feedback from the user, this system is an important and accessible tool for members of the deaf-mute community as a practical solution for sign language conversion in an interactive way.

- Problem statement and questions:

Currently, in this field, recognition systems face various difficulties in terms of accessibility, processing of datasets in real time easily and also the level of accuracy provided. On the basis of the research and articles browsed, several particular situations have been identified, as strictly dependent such as brightness and contrast conditions in a working environment, limitation in terms of available hardware capacity, but also the clarity and calibration of the system for continuous use. Another important aspect, is that many of these systems fail to meet the needs of members of the deaf-mute community, thus reducing the interaction with the technology. In this paper limitations will be addressed after solving the following questions:

(1) By using a webcam and combining the two libraries OpenCV and MediaPipe, what results will the gesture recognition system achieve?

(2) Based on the tests performed in different environments and working conditions, what is the actual level of accuracy that the system provides under particular lighting scenarios?

(3) Out of the distinct set of thirty gestures, will the algorithms be able to achieve considerable and accurate results in gesture recognition?

(4) Will the system interface realized using Tkinter be sufficiently friendly and accessible in its interaction with users?

(5) As an addition to the presented system, what other new technologies can be incorporated to increase the level of accuracy and user feedback?

(6) How will the system behave and affect the recognition process in the way it can continuously capture both static and dynamic images?

(7) Will the response time of the system in any way affect the experience of users within the deaf-mute community?

Objectives:

The main objective of this work is to obtain a sufficiently robust gesture recognition system, using state-of-the-art technologies, integrated in identification and classification steps with accurate results obtained in a short time. Also, the system as mentioned above will be tested in different light and contrast conditions. A very important aspect is also the implementation and improvement of the whole system, based on artificial intelligence algorithms, thus having the ability to identify and store complex movements of the user's hand. This will improve the response time of the system, as a result of user interaction. As mentioned, the interface has been realized in a user-friendly and accessible way for all users and members of the deaf-mute community through Tkinter.

Also, based on the level of technology that has been reached to meet the needs of all users the system was designed to be flexible and scalable regardless of the conditions of the working environments in which it is engaged. Previously mentioned tests that have been performed on the system, confer promising results on the accuracy, but also on aspects related to its improvement measured under the metrics of artificial intelligence. Additionally, information is obtained from within the deaf-mute community which makes it possible to meet the needs and further develop the system in a user-friendly and successful way.

This paper is structured as follows: the second section presents existing recognition systems and the literature in the field; the third section describes the entire architecture of the recognition system also a detailed memento of the collection stages, data training, optimization methods and the performances obtained by the entire system; the fourth section illustrates and analyzes the results obtained by the system; the fifth section summarizes the conclusions and key points identified in the presented work and outlines the future research directions.

2 Related Work

An important step in the field of artificial intelligence is attributed to the system of recognizing hand gestures in a natural way and represented through applications with friendly interfaces. The research works in this field have been strictly focused on gesture analysis using various technologies which includes the use of specialized equipment that come equipped with sensors and cameras. The sensor-based approach is not very efficient due to its acquisition costs and strict limitation by computer cables, here sensor gloves are included. Also, fine cameras like Microsoft's have been widely used for three-dimensional gesture recognition. However, they too have had limitations when it comes to testing in various work environments, interfering with recognition accuracy.

The latest papers, also rely on using only algorithms for gesture identification and classification, but also cameras. By this, it ensures and covers user needs in an efficient response time portable on any device. By combining the two main libraries OpenCV and MediaPipe, the learning model has the full capability of recognizing gestures performed by users under high accuracy. Also, establishing an optimal accuracy still remains a persistent goal in current research as well as working environments and generalization of the user hand model. Another aspect to mention, would be the fact that the accessibility of systems is not designed to meet the needs of all users such as members of the deaf-mute community.

K. D. Kumar et al. propose a sign language optimization model by combining convolutional neural networks with manually worked datasets. By integrating these datasets, it improves the accuracy of the model subjected to complex scenarios with noise dominated by variations between different gestures. Based on the proposed model, remarkable performances were obtained in the recognition of gestures from datasets with background noises, proving also a high level of robustness in recognizing complex gestures.

S. Rithesh Manikandan et al. propose a system for the recognition of Indian sign language-oriented to people from deaf-mute communities. They believe that the implementation of this solution provides the possibility of translation and accessibility of people in real time with technology. The results they obtained are good due to the high accuracy obtained on a limited set of gestures. The system can also be easily integrated into everyday life, both on portable devices and on specialized equipment such as servers.

S. V. Nimbalkar et al. created an interpreter for the Indian sign language identification by implementing deep learning algorithms. Also, the displaying of the results is done through a simple interface with the possibility of rendering the results as text or voice. The obtained model achieves remarkable performances for both static and dynamic gestures. To improve the accuracy, augmentation techniques and dynamic adjustment of hyperparameters have been implemented.

W. Qin et al. propose a method for recognizing and translating sign language using a visual transformation network. A model is used to realize machine translation as well as to correct gestures and grammatical mistakes in sentences. The results obtained by the model are accurate, forming coherent sentences on the basis of the signs made by the user. It can also be assigned to a more complex set of gestures or to another sign language standard.

P. Khaire et al. propose the development of a system for the automatic translation of sign language into text or audio sequences using artificial intelligence techniques. This technique is optimized so that it can be easily integrated on mobile devices with real-time gesture recognition capability. The proposed system provides fast feedback to the user through a simple and user-friendly interface, making it accessible on multiple platforms.

R. Verma et al. create a system for automatic gesture detection and sign language recognition by implementing machine learning techniques. In order to ensure a high level of robustness, this system is subjected to different lighting conditions and different gesture capture angles. As a result, the system registers outstanding performance on a diverse set of gestures and the possibility of implementing advanced image preprocessing techniques in order to significantly reduce the classification errors.

A. Singh et al. develop a system for recognizing dynamic gestures present in the Indian sign language alphabet. Also, an optimized algorithm in combination with convolutional neural networks is used to identify and recognize complex gestures and the position of the user's hand. The system has a good accuracy in recognizing continuous gestures with the possibility of integrating virtual assistants.

T. G. Moape et al. use artificial neural networks for African sign language recognition and translation. They also conducted a study on dialects from African regions to build a robust model. The model finally achieves a high accuracy for gesture recognition with the possibility to compose coherent sentences independent of the variations of gestures performed by the user.

For deaf-mute community members, the gesture recognition system facilitates communication and interaction with other users through sign language. Research in previous years has also explored, how to make sign language more accessible to users who do not know or understand it. The data sets that the system incorporates are not complex in terms of volume, but are limited to a strict number of identified and precisely categorized gestures. In the proposed recognition system, these limitations are overcome by implementing OpenCV and MediaPipe libraries. This provides an accessible solution involving a complex number of users, including members of the deaf-mute community. Finally, the gestures rendered by users in front of the camera, are converted into text opening a new way of digital communication [1-8].

3 System Architecture

3.1 Proposed system for detection of the hand gestures

The proposed system for hand gesture recognition and identification has been designed in such a way that the user's movements recorded by the camera are captured naturally in different working environments. As mentioned later, it incorporates two sufficiently powerful libraries OpenCV and MediaPipe for recognizing and classifying key hand features such as wrists, joints and degrees of freedom. To achieve high accuracy the system architecture works in a few simple steps. First the camera is initialized through which both dynamic and static images and frames are captured. The dynamic frames are used for accurate identification of users' hand gestures by using the point history classifier. The MediaPipe library is used to track the gestures made by the users, through which the key characteristics of the hand are stored and displayed on the screen, among which are the fingers, hands, wrists and joints used to distinguish each gesture. Two different classifiers are used to classify gestures. The first classifier is used for static sequences, by tracking in particular, the position of the fingers and is called the key point classifier and the other one for dynamic work sequences, called historical point classifier. When a gesture is recognized, it is displayed accordingly on the Tkinter user interface screen. This interface also provides real-time feedback based on the gestures transmitted to the camera by the user. The system has been designed to be able to give prompt responses for both static and dynamic sequences. The results, cover this by the fact that the two libraries feature efficient algorithms in easy resource utilization. The recognition of the thirty distinct gestures includes both static and dynamic sequences, as an example can be seen in Figure.1.

The obtained dataset is fully labeled so that the classifiers can ensure an increased accuracy in the hand gesture recognition process. It also represents a wide opportunity for communication and inclusive integration of deaf-mute community members, through direct interaction and in an interactive manner.

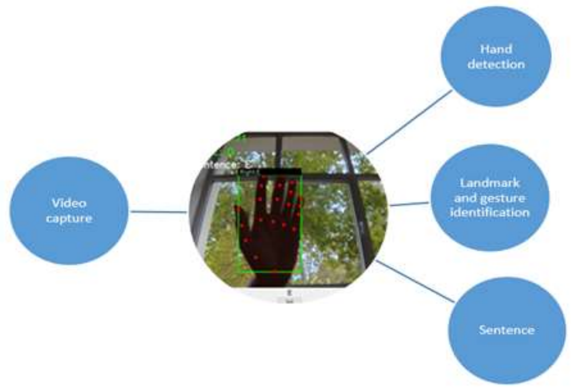

Fig. 1. Block scheme of the recognition system

The simple gestures made by the users in front of the camera, are in fact the way of communication which brings a benefit in various domains. In this way, the accessibility that users have to the system can be highlighted. As an advantage of this system, through its ability to recognize in real time the data sequences transmitted from the camera, it becomes accessible in several workspaces. It is worth mentioning, that it still encounters recognition difficulties in low brightness and low contrast conditions, but also when the user's hand is covered by, for example, gloves or other unknown objects. This affects the accuracy of the whole system in gestures recognizing phase. Nevertheless, the system represents an efficient and accurate approach, for future hand gesture recognition applications in the field of artificial intelligence.

As future research directions, the main objectives will be to enrich and extend the number of gestures in the vocabulary, which will increase the robustness of the system in different working environments and to implement new technologies prior to artificial intelligence.

3.2 Data collection and preprocessing

For the realization of the whole system, the first step is the work data collection stage. The dataset from which the hand gesture recognition model is composed, contains various user frames transmitted and captured by the camera. As mentioned, the dataset includes both static and dynamic gestures, consisting of various sign language alphabets and also the collection of frames from members of the deaf-mute communicators was used. By adding the working dataset from members of the deaf-mute community, some of their specific needs were identified and covered. Again, tests were performed in different working environments under different lighting and contrast conditions to identify the degree of robustness of the proposed system. By adding various gestures, the system is able to understand the complexity of real scenes. Another important aspect is the preprocessing stage, which directly influences the accuracy of the system. The images captured by the camera, are then transmitted to the system, which converts the images to grayscale, in order to optimize the computation and complexity of data. The tracking tool for key points, fingers, wrist and hand joints, is also provided by the MediaPipe library.

These values, are then passed through a normalization process, due to the fact that the dimensions and anatomical particularities of the users' hands are different, representing another strength of the input data in the database.

3.3 Model training

The foundation of the gesture recognition system is found in the models, trained on the datasets that are used in hand motion recognition. As already mentioned, two special classifiers have been used for recognition, namely the historical point classifier and the historical key classifier for both static and dynamic gestures captured by the camera. The first classifier uses as input values, data or gestures that have already passed through the normalization stage. These normalized values, have labeled values, that are found when displaying the result on the interface screen. Mandatory, optimization techniques are also used to remove the possibility of overlap or misclassification by cross-validation and to observe if the model behaves differently when trying to recognize a gesture that is not in the vocabulary. The second classifier, resorts only to dynamic data sequences captured by the camera. A few frames of the hand recorded in a previous time interval are analyzed and then the new data is trained, to recognize new gestures or to validate existing gestures in the vocabulary. The performance of the gesture recognition system was evaluated using accuracy, F1-score and Recall function as tools. The first tool ensures that the proposed model has the ability, to correctly classify among all gestures and predict the results as accurately as possible. The sensitive or recall function, counts the ability of the system only to correctly recall all recognized gestures. The harmonic mean between the two tools mentioned above is called F1-score. It is intended to identify unrecognized or misclassified gestures and to balance the recognition power of the system. Another very important tool is the confusion matrix, which validates the performance that the system offers in several different classes of gestures. It also helps to identify misclassified gestures as well as to visualize and correct the weaknesses of the whole system [9-17].

3.5 Model optimization and performance evaluation

The development of the gesture and sign language recognition system, was also subject to a preliminary stage of optimization, in order to increase the degree of accuracy, robustness and to improve user feedback. This chapter illustrates the approaches considered for the gesture recognition system, which after implementation have been of enormous benefit for the members of deaf-mute community and are summarized below.

- Hyperparameter tuning:

This concerns the adjustment of hyperparameters within the two classifiers key point history classifier and point history classifier respectively. Also, different types of systematic hyperparameter search, adjusted and added more layers, improved learning level and activation functions were explored. These represent the most effective improvement methods in terms of classification accuracy and gesture identification as well as fast response time on the feedback from users.

- Data augmentation:

Despite the fact that the datasets have been extended, augmentation techniques have also been added to mainly improve the robustness of the system. The methods used to achieve this were by rotating, scaling and flipping the images, which resulted in the model's ability, to analyze and identify new data not present in the vocabulary. This aspect is very important due to the fact that it also involves dependency on the degree of tilt or rotation of the hand, which provides the system with feedback for accommodating in real life situations.

- Model pruning and quantization:

To make the proposed system more efficient and adding possibility of adaptation to other equipment or platforms, two techniques were necessary to implement. Pruning, means simplifying the system by removing some neurons or integrating within a neural network, without reducing the size of the system's datasets, that will improve the system's accuracy. Also, converting model weights to smaller units will allow the system to respond in a timely manner and with minimal operating resources, these represents aspects of the quantization.

- Real-Time performance optimization:

The system was designed to render in real time, by reducing the latency level. For fast and efficient processing of the data sequences, techniques to reduce the size of the recorded frames were implemented, as well as parallel processing steps, thus providing almost instantaneous results and user feedback. The system has also gone through these stages in order to acquire the ability to adapt to different platforms and equipment, with minimal requirements. Preliminary results obtained by the recognition system, indicate that there is still room for improvement: initial vocabulary of the system was trained to identify only a specific set of sign language gestures. Implementing new gestures or sentences through different scenarios will increase the robustness of the whole system. For this, it will be necessary to adopt a new training dataset and to adjust the model used for gesture classification. Another important aspect, may be the addition of more sign language alphabets, by collecting different datasets, which will lead to a global extension of the system. After the successful adoption of the sign language to text translator, another enhancement would be text to speech conversion. This would lead to improved communication with members of the deaf-mute community, but also provides a main pillar for people who do not know or want to learn sign language. Another important feature to be mentioned would be the possibility for users to customize the model, by adding their own datasets or sign language sequences, leading to increased accuracy, both locally and digitally stored.

The level of computation, has undergone large changes, so that the system can maintain its accuracy, regardless of the complexity of the dataset and therefore of its users. This continuous approach to the recognition model, does not impose any requirement on users to update or modify it. By optimizing the system, it gives the possibility, to transform it into a widely used application, with a strong communication benefit for the members of the deaf-mute community. By achieving the high accuracy and implementation opportunities mentioned above, this system can become an indispensable global communication resource. The system's overall performance recorded is 92%, reflecting the highest accuracy obtained in recognizing gestures across various scenarios, in working environments under different light intensity and contrast. These results were also influenced by the previously mentioned optimization and refinement techniques. By implementing the LSTM layer, the system has the ability to identify time sequences for both static and dynamic gestures. The effect of overlapping due to confusion between gestures of different classes has been diminished due to LSTM's capability of sequential image analysis. Also, the model has the ability to generalize to new gestures regardless of the variations between classes or the different illumination conduction or viewing and capture angles.

3.6 Comparative analysis with existing approaches

In this section is presented a comparative analysis of the gesture recognition system with other existing approaches in the field of artificial intelligence. This analysis emphasizes as objectives: the accuracy, the robustness of the system, the way it is implemented, the equipment it uses and the strengths and weaknesses. It has already been previously stated that the gestures recognized in the specialized studies of today's researches are based on methods involving either equipment with sensors and cameras or machine learning algorithms. The acquisition costs of hardware equipped with sensors and cameras are very high as in the case of gloves. Also, they are limited because the transfer of information they make is strictly dependent on the computer or laptop via transmission cables. The sensors, offered by Microsoft, have been a promoter in this field for their ability to capture images and virtualize them in three-dimensional space. They confer increased accuracy in terms of gesture detection, regardless of the brightness and contrast conditions of the various working environments being limited only by the hardware component [19].

The proposed system uses both a simple webcam and algorithms provided by MediaPipe and OpenCV libraries for gesture recognition. This eliminates the dependency on hardware, but also opens a new horizon by interacting with new users. By using OpenCV, real-time gesture preprocessing and gesture recognition are achieved providing instant feedback in real time.

In terms of accuracy, the system performs much better than existing approaches. By integrating the two classifiers, the system achieves a very high rate of accuracy when recognizing different gestures. While sensor-based approaches in different working environments provide a slightly higher accuracy rate, the proposed system proves to be scalable and responsive to the requirements of its users. It also proves high performance in terms of its adaptability to different work environment situations such as different lighting and contrast conditions. Unlike other existing approaches, it is portable to different platforms and camera-independent equipment.

The optimization steps ensure, as a novelty element, the possibility for the system to evaluate in a different way new gesture that are not in the vocabulary [20].

Compared to other later approaches, models using only convolutional neural networks had varying accuracy between 80-90% due to temporal dependencies. MLP (Multi-Layer Perceptron) models or SVMs have an accuracy of up to 80%. So, by implementing a LSTM (Long-Short Therm Memory) layer, the accuracy has surpassed the 90% threshold and provides increased performance and high robustness to the proposed system for recognizing complex gestures.

In conclusion, in the case of the proposed gesture recognition system, it represents an alternative solution to existing approaches. By bringing together the real-time performance of the system, it makes it possible to extend the domain to other applications, by facilitating and easing the communication with members of the deaf-mute community through a user-friendly interface.

Although, there are still areas that require some simple adjustments; by increasing the number of gestures in the vocabulary will activate the full potential and the system will have the ability to recognize in any position or work independent of environment and in various conditions different types of gestures.

4 Results and discussion

This section presents the results that were obtained after the implementation and testing phase of the proposed gesture recognition system. Its key features was its accuracy, its testing in different working environments and also the benefits it brought from the interaction with members of the deaf-mute community.

The dataset on which the system has been trained is composed of three different sets of thirty different gestures each, as can be seen in Figure.4. In total, an accuracy of 92% was recorded, whereby the gestures recognized by the two classifiers are correctly labeled by a complete example shown in Figure.2.

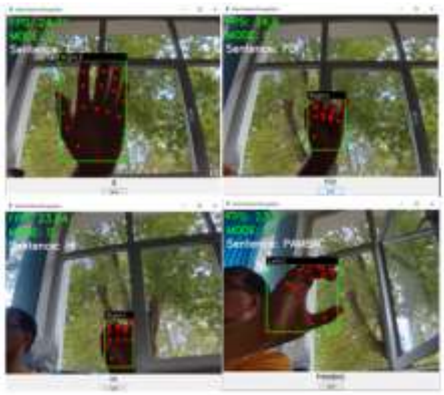

Fig. 2. Results of the proposed system System's confusion matrix

The accuracy recorded throughout the tests in different working environments under light incidence and contrast, highlight the fact that through the optimization and tuning steps, a robust system has been successfully achieved. It is also worth mentioning, that in working environments under low light incidence, the accuracy decreases to 88% and also if users cover their hand with an unknown object, the algorithm suffers difficulties in the recognition stage as can be seen in Figure.3.

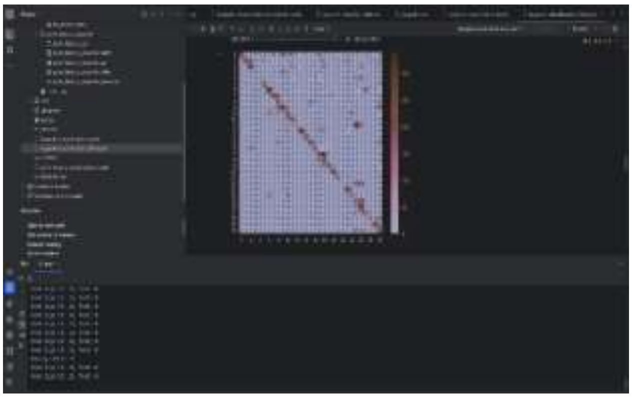

Fig. 3. System’s confusion matrix

In the above image, are the zones that need attention on the recognition side and adding new features or fine tuning in the optimization process. The system also has the possibility of capturing dynamic frames with a value of 25-30 fps, without consuming additional resources or needing state-of-the-art equipment to support gesture recognition. The response time from the moment the gesture has been executed by a user and the feedback provided by the system is very short in terms of milliseconds. Thus, the live transmission of frames and the feedback provided, contribute to the communion with members of the deaf-mute community.

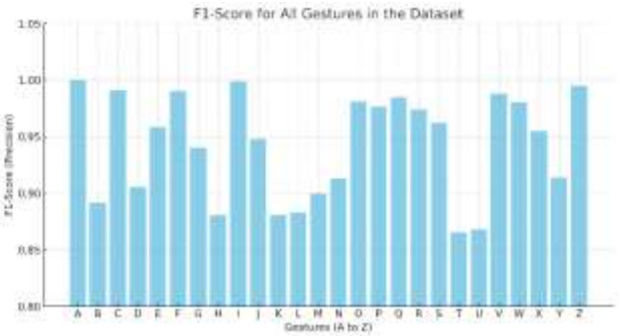

Fig. 4. Overall F1-Score for all gestures

Based on the experiments that were carried out to improve the performance of the gesture recognition system, an LSTM layer was added to the recognition stage. The main purpose of this layer is to remove the temporal and spatial dependency of complex or dynamic gestures. The model recorded a significant accuracy of 94% compared to architectures using only convolutional neural networks. Also, the information captured from the image is stored over a longer time interval in order to handle the temporal or spatial variation present in the dataset.

The last feature to be mentioned concerns the system's ability to convert recognized gestures into text. This is an important component of the system, which is strictly user-dependent and encompasses both members of the deaf-mute community and people who want to learn sign language. In the testing phase, the interaction with members of the deaf-mute community took place and it was found that the system had no difficulties in identifying the native gestures made by them in front of the camera. By this, not only the communication barrier is covered, but also an inclusion of all users in digitized environments. Thus, the system allows users of the deaf-mute community to interact in a free and easy way. After successful testing and completion of all gesture recognition, however, some possible limitations of the proposed system were discovered. The lighting and contrast conditions, the covering of the hands with unknown objects or gloves make the algorithm encounter difficulties in recognizing key hand features.

Even if the gesture vocabulary is limited to only thirty different gestures, by extending it, these limitations can be overcome and it also gives the system the ability to generalize better on new datasets that are not part of the vocabulary. The level of scalability is limited by the number of users. For it to be scalable, it is necessary to address optimization and preprocessing steps for new gestures.

Finally, the proposed and developed system, brings a consistent contribution in the field of artificial intelligence by improving the communication due to the use of sign language. By small considerable adjustments it can constitute a powerful and accessible tool for communication across different devices and platforms.

5 Conclusions and future work

The development of the proposed hand recognition system, presents an important benefit in the field of artificial intelligence and in particular facilitate the interaction and inclusion of deaf-mute community members, by tracing a new bidirectional communication path. By utilizing the two libraries, the system's ability is extended based on its component datasets and live webcam frames, also constituting a foundation for future applications in the field. Moreover, the speech-to-text conversion also brings an important contribution to the users evaluated on the basis of user feedback.

The technology embedded in the system, is not only tested in a closed experimental environment, but through various tests in different working scenarios and has the ability to adapt to different conditions and real-life situations. Despite the success, as a result of increased accuracy, the proposed system can be improved. In terms of future research directions, a first step is to expand the number of gestures stored in the vocabulary, which will cover the needs of the system and users in new scenarios, allowing a better understanding of new gestures. By adding the LSTM layer in the recognition stage, the system registers a high performance for both static and dynamic gesture recognition, constituting a robust and simple to implement solution compared to existing approaches.

Several optimization and preprocessing techniques will also be used, with the suggestion of adding more layers, for deeper understanding of key hand gestures, even if the hand is covered by unknown objects or gloves. Several sign language specific alphabets can also be added which will make the system globally integrated and accessible to all users worldwide. The reverse conversion of text into sign language as well as its vocalization, can also be considered an aspect of the system improvement.

This would add more creativity and interactivity to the whole system, by being more efficient and providing timely feedback. The last and most important key aspect also mentioned at the beginning of this paper, remains the removal of any communication barriers between digitization and deaf-mute community members, by implementing a cloud preprocessing, with the aim of efficiency and extended portability in terms of the proposed system.

By further fine-tuning and implementing new technologies to the proposed system, it will become a powerful and necessary tool in maintaining and easing the communication barriers, by being accessible and responding to the future inclusive requirements of the users.

References:

[1] K. D. Kumar, K. Ragul, G. P. Pravin Kumar, and G. Kajeeth Kumar, "Enhancing Sign Language Recognition through Deep CNN and Handcrafted Features," in *2024 2nd International Conference on Networking and Communications (ICNWC)*, Chennai, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICNWC60771.2024.10537400.

[2] S. Rithesh Manikandan, G. L. Prasuna, G. Sivanesh, and K. Divya Lakshmi, "Gesture Guide: Empowering Deaf and Mute Individuals with ISL Recognition and Guidance System using Hand signs," in *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India: IEEE, Mar. 2024, pp. 462–465. doi: 10.1109/ICICV62344.2024.00078.

[3] S. V. Nimbalkar, S. N. Vaidya, M. M. Gade, P. S. Hagare, and P. N. Shendage, "Empowering Deaf with Indian Sign Language Interpreter using Deep Learning," in *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)*, Pune, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/MITADTSoCiCon60330.2024.10575064.

[4] W. Qin, X. Mei, Y. Chen, Q. Zhang, Y. Yao, and S. Hu, "Sign Language Recognition and Translation Method based on VTN," in *2021 International Conference on Digital Society and Intelligent Systems (DSInS)*, Chengdu, China: IEEE, Dec. 2021, pp. 111–115. doi: 10.1109/DSInS54396.2021.9670588.

[5] P. Khaire, S. Mahajan, P. Pawar, S. Pawar, Dr. S. Dedgaonkar, and Prof. Chaitali Shewale, "SilentSync-AI," in *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)*, Pune, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/MITADTSoCiCon60330.2024.10575259.

[6] R. Verma, J. Singh Chohan, S. Kumar Joshi, J. Kaur, H. S. Bhatia, and H. Mohammed Abbas, "Machine Learning Approach for Sign Language Recognition System Development," in *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Ballari, India: IEEE, Apr. 2024, pp. 01–05. doi: 10.1109/ICDCECE60827.2024.10549550.

[7] A. Singh, A. Wadhawan, M. Rakhra, U. Mittal, A. A. Ahdal, and S. K. Jha, "Indian Sign Language Recognition System for Dynamic Signs," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India: IEEE, Oct. 2022, pp. 1–6. doi: 10.1109/ICRITO56286.2022.9964940.

[8] T. G. Moape, A. Muzambi, and B. Chimbo, "Convolutional Neural Network Approach for South African Sign Language Recognition and Translation," in *2024 Conference on Information Communications Technology and Society (ICTAS)*, Durban, South Africa: IEEE, Mar. 2024, pp. 101–106. doi: 10.1109/ICTAS59620.2024.10507130.

[9] N. Gulati, A. Rajput, and A. Singh, "Sign Language Recognition using Convolutional Neural Network," in *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India: IEEE, Jan. 2024, pp. 876–881. doi: 10.1109/Confluence60223.2024.10463257.

[10] H. Adhikari, Md. S. Bin Jahangir, I. Jahan, Md. S. Mia, and Md. R. Hassan, "A Sign Language Recognition System for Helping Disabled People," in *2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI)*, Dhaka, Bangladesh: IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/STI59863.2023.10465011.

[11] R. Gupta, K. Gupta, C. Pandit, and P. Singh, "Sign Language Recognition using VGG16 and ResNet50," in *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi,

India: IEEE, Feb. 2024, pp. 996–1001. doi: 10.23919/INDIACom61295.2024.10498253.

[12] V. Subashini, B. Someshwaran, S. Sowmya, and S. A. Kumar, "Sign Language Translation Using Image Processing to Audio Conversion," in *2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, Krishnankoil, Virudhunagar district, Tamil Nadu, India: IEEE, Mar. 2024, pp. 1–6. doi: 10.1109/INCOS59338.2024.10527627.

[13] R. V. Vardhan, B. Girinath, L. Murali, D. J. Ferdin, and V. K. Aswathaman, "Automatic Sign Language Recognition Using Convolutional Neural Networks," in *2024 International Conference on Science Technology Engineering and Management (ICSTEM)*, Coimbatore, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICSTEM61137.2024.10560930.

[14] Z. A. Abdulkarim Shwany, S. K. Ismail, K. H. Khalid Khoshnaw, T. S. Mustafa, M. Karabatak, and N. A. Othman, "A Real-Time Hand Sign Language Recognition System for Threatening Situations Using Deep Learning," in *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, San Antonio, TX, USA: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ISDFS60797.2024.10527337.

[15] P. Das, T. Ahmed, and Md. F. Ali, "Static Hand Gesture Recognition for American Sign Language using Deep Convolutional Neural Network," in *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh: IEEE, 2020, pp. 1762–1765. doi: 10.1109/TENSYMP50017.2020.9230772.

[16] A. Kumar, S. Utekar, and D. Viji, "Hand Sign Gesture Recognition and Translator Using Convolution Neural Networks (CNN) and k-Nearest Neighbors (k-NN)," in *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, Bengaluru, India: IEEE, Mar. 2024, pp. 1–7. doi: 10.1109/ICDCOT61034.2024.10515667.

[17] P. Dhanalakshmi, M. V. Sree, T. Hindu, T. Papaiah, and M. Karna, "Hand Gesture Recognition System using Feedback CNN," in *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India: IEEE, Jun. 2024, pp. 167–173. doi: 10.1109/ICAAIC60222.2024.10575066.

[18] M. Shabana Parveen, R. G. Keerthana, S. Shanmathi, and M. Shajitha, "Sign Language Detection Using Open Cv," in *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Chennai, India: IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/IC3IoT60841.2024.10550232.

[19] T. Kumaragurubaran, S. Senthil Pandi, M. Tharun, and M. V. Devi, "Unlocking Sign Language Communication: A Deep Learning Paradigm for Overcoming Accessibility Challenges," in *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Chennai, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/IC3IoT60841.2024.10550356.

[20] O. B. Soultane, Z. Amine, and Y. Salih-Alj, "A Smart Sign Language Interpreter for Medical Environments Using Deep Learning: Morocco Case Study," in *2024 2nd International Conference on Mechatronics, Control and Robotics (ICMCR)*, Jeju, Korea, Republic of: IEEE, Feb. 2024, pp. 21–26. doi: 10.1109/ICMCR60777.2024.10482238.