International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 12 | Dec 2024 www.irjet.net p-ISSN: 2395-0072

Novel Multi-Modal CNN - Fuzzy Based Hand Gestures Recognizing System

Shrivarshan N K[1], Sahana Mahesh[2], Anbuchelvan A P[3]

[1,2&3] *Student, School of Computer Science Engineering and Information Systems, SCORE Vellore Institute of Technology, Vellore, Tamil Nadu, India*

Abstract - This paper presents a novel multi-model neural network and fuzzy-based system for hand gesture recognition, leveraging the combined strengths of Convolutional Neural Networks (CNN) and Fuzzy Inference Systems (FIS). The proposed approach aims to enhance gesture detection accuracy by integrating a fuzzy logic-based edge detection mechanism with a CNN model for improved feature extraction and classification. The Fuzzy Inference System is designed to detect edges in hand gesture images by applying fuzzy rules and membership functions, which reduces the complexity and noise in the input data. The edge-detected images are then fed into the CNN for training, resulting in a more efficient recognition process with improved accuracy and reduced computational costs. Experimental results demonstrate that the proposed system outperforms traditional CNN-based approaches in terms of recognition accuracy and robustness across varying lighting conditions and hand shapes. This innovative combination of CNN and fuzzy logic provides a reliable and efficient solution for hand gesture recognition, with potential applications in human-computer interaction, sign language interpretation, and virtual reality.

Keywords – Convolutional Neural Networks, Edge Detection, Fuzzy Inference Systems, Hand gesture analysis, Human Computer Interaction

1. INTRODUCTION

Hand gesture recognition is a critical area of research in human-computer interaction (HCI), offering intuitive and non-invasive methods for users to interact with digital environments. Traditional approaches to hand gesture recognition primarily rely on deep learning techniques, such as Convolutional Neural Networks (CNNs), which have demonstrated significant success in extracting complex features from images. However, CNN-based models often face challenges in handling noise, variations in lighting, and complex backgrounds, which can adversely affect recognition accuracy.

To address these limitations, this paper proposes a novel hand gesture recognition system that combines the capabilities of CNNs with a Fuzzy Inference System (FIS). The FIS is employed to perform edge detection on input images by applying fuzzy logic-based rules, thereby reducing noise and enhancing the most salient features of hand gestures. The pre-processed, edge-detected images are then utilized to train a CNN model, which focuses on learning discriminative features essential for accurate gesture classification. This hybrid approach aims to improve the robustness and accuracy of hand gesture recognition systems, especially in challenging real-world environments.

The remainder of the paper is structured as follows: Section II discusses related work in the field of hand gesture recognition. Section III describes the proposed multi-model CNN-Fuzzy system, detailing the architecture and the fuzzy logic-based edge detection process. Section IV presents the experimental setup and results, highlighting the performance of the proposed system compared to conventional methods. Finally, Section V concludes the paper with a discussion on future research directions.

2. LITERARY SURVEY

Hand gesture recognition (HGR) has been an active research area due to its applications in human-computer interaction, sign language recognition, and automation. Various techniques and models have been developed to enhance the accuracy and robustness of HGR systems, involving different sensors, machine learning algorithms, and deep learning models.

Sharma et al. (2023) proposed a time-distance parameter-based HGR system using multiple ultra-wideband (UWB) radars to capture hand gestures. Their approach demonstrated the potential for UWB radars in accurately detecting and recognizing hand movements by leveraging time and distance parameters. This method is beneficial for environments where conventional cameras may face challenges due to lighting or obstructions.

Another study by Sharma et al. (2023) explored the application of machine learning techniques in HGR. They conducted an extensive review of different machine learning methods, including support vector machines, decision trees, and deep learning models, highlighting the advantages and limitations of each approach in

recognizing hand gestures. This study provides a comprehensive understanding of the strengths and weaknesses of various machine learning models for HGR.

Kang et al. (2023) developed a hand gesture recognition system based on surface electromyography (sEMG) signals using a binarized neural network. This approach aimed to reduce computational complexity while maintaining high accuracy by binarizing the network's weights and activations. The study demonstrated that sEMG-based systems could be effective in scenarios requiring real-time performance with minimal computational resources.

Huang et al. (2023) focused on real-time automated detection of hand gestures of older adults in home and clinical settings. They employed a deep learning-based approach that utilized temporal and spatial information from video sequences, achieving high recognition accuracy in real-time scenarios. This study highlights the importance of adapting HGR systems for specific demographic groups, such as the elderly, where gesture recognition can play a crucial role in monitoring and assistance.

Sahoo et al. (2023) proposed DeReFNet, a dual-stream dense residual fusion network designed for static hand gesture recognition. The network integrates dense connections and residual learning to capture both global and local features effectively. Their method showed improved recognition rates compared to other conventional deep learning models, demonstrating the effectiveness of integrating multiple feature streams in gesture recognition tasks.

Miah et al. (2023) introduced a multistage spatial attention-based neural network for hand gesture recognition. Their approach utilized spatial attention mechanisms to focus on significant regions of the input images, enhancing the model's ability to discriminate between similar gestures. The study demonstrated the efficacy of attention mechanisms in improving the recognition accuracy of deep learning models.

Damaneh et al. (2023) employed a convolutional neural network (CNN) with feature extraction methods using ORB descriptors and Gabor filters for static hand gesture recognition in sign language. The integration of ORB and Gabor filters enhanced the model's ability to capture essential features, leading to improved recognition performance compared to conventional CNNs.

Gao et al. (2023) developed a hand gesture teleoperation system for dexterous manipulators in space stations using monocular hand motion capture. This study demonstrated the feasibility of applying hand gesture recognition in remote and constrained environments, where traditional input devices may not be practical.

Bora et al. (2023) proposed a real-time Assamese sign language recognition system using Mediapipe and deep learning. Their method achieved high accuracy and real-time performance, demonstrating the potential of deep learning frameworks like Mediapipe for low-resource language applications.

Qahtan et al. (2023) conducted a comparative study evaluating sign language recognition systems based on wearable sensory devices using a single fuzzy set. Their research highlighted the importance of fuzzy logic in handling uncertainty and imprecision in gesture recognition, providing insights into the potential advantages of fuzzy-based systems over traditional methods.

Sarkar et al. (2023) presented a human activity recognition model using sensor data combined with a spatial attention-aided CNN and a genetic algorithm. This model showed enhanced performance in recognizing complex gestures by optimizing feature selection and model parameters through genetic algorithms.

Das et al. (2023) proposed a hybrid approach for Bangla sign language recognition using deep transfer learning models with a random forest classifier. This method integrated the strengths of deep learning and traditional machine learning to achieve high recognition accuracy, particularly in low-resource language contexts.

These studies collectively highlight the diversity of approaches in hand gesture recognition, ranging from sensor-based systems and machine learning algorithms to deep learning frameworks incorporating attention mechanisms and fuzzy logic. The integration of various techniques, such as spatial attention, fuzzy logic, and multi-sensor data, demonstrates the evolving nature of HGR technologies aimed at improving accuracy, robustness, and adaptability in diverse applications.

3. PROPOSED METHOD

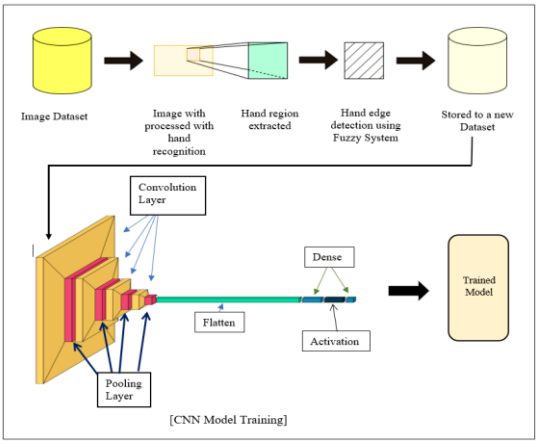

Figure -1: An outline of proposed architecture.

3.1 Fuzzy Methodology

a. Fuzzy Membership Functions:

Membership functions specify how each value in an input set fits into a particular fuzzy set. In this methodology, the input variable (gradient magnitude) is categorized using triangular membership functions.

b. Fuzzy Rules:

The inference system's rule base is made up of fuzzy rules, which define the relationship between input and output variables. In this system, the rules are defined based on the gradient magnitude:

- *Rule 1:* The edge strength is low if the gradient magnitude is low.
- *Rule 2:* The edge strength is medium if the gradient magnitude is medium.
- *Rule 3:* The edge strength is high if the gradient magnitude is high.

c. Fuzzy Inference:

The fuzzy inference system receives an input (gradient magnitude) and applies the fuzzy rules to determine the degree of activation for each fuzzy set in the output variable.This process involves two steps:

- *Fuzzification:* The input value is assigned to the relevant fuzzy sets based on their membership functions.
- *Rule Evaluation:* The system combines the activated rules to determine the output fuzzy set activations.

d. Defuzzification:

The activated output fuzzy sets are combined to produce a crisp output value. This process is known as defuzzification. In this case, centroid defuzzification is used, which calculates the center of mass of the activated fuzzy sets to determine the final output value.

3.2 CNN Architecture

a. Convolutional Layers:

The model begins with a convolutional layer consisting of 32 filters of size 5x5, utilizing a ReLU (Rectified Linear Unit) activation function. 32 unique features are extracted from the input images by this layer. Additional convolutional layers are added, each with a different number of filters and sizes, to capture multiple features and aspects of the input image.

b. Max Pooling Layers:

After each convolutional layer, a max-pooling layer with a pool size of 2x2 and a stride of 2x2 is added.

c. Flatten Layer:

After the final max-pooling layer, a flatten layer is introduced to convert the 2D feature maps into a 1D feature vector, preparing the data for fully connected layers.

d. Fully Connected Layers:

The architecture includes a dense (fully connected) layer with 512 units, which serves as a high-level feature extractor from the flattened feature vector, followed by a ReLU activation function to introduce non-linearity. The final dense layer consists of 10 units corresponding to the 10 classes in the classification task, followed by a softmax activation function to convert the output into class probabilities.

e. Training Configuration:

The batch size is set to 128. The model is trained over 10 epochs, allowing the entire dataset to be passed through the network 10 times, optimizing the weights for accurate predictions.

4. IMPLEMENTATION

4.1 Image Dataset:

We have used Hagrid dataset, which is a dataset containing 12gb of images of various gestures like call, like, dislike, etc.

4.2 *Hand Recognition:*

In this project, the MediaPipe framework was employed for hand detection and bounding box extraction. MediaPipe offers a comprehensive suite of pre-trained machine learning models and processing pipelines specifically designed for various computer vision tasks, including hand tracking. Leveraging MediaPipe's Hand Tracking solution, the system efficiently localizes and tracks the user's hand in real-time video streams or images. By utilizing this technology, the project benefits from MediaPipe's robustness and accuracy in detecting hand landmarks and estimating hand poses. Moreover, MediaPipe's lightweight and optimized implementation enable seamless integration with the overall system architecture, ensuring real-time performance and responsiveness. This choice of technology simplifies the hand detection process, allowing the project to focus on subsequent stages such as feature extraction and gesture recognition.

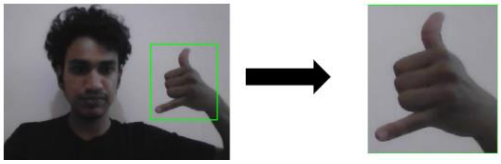

Figure -2: Depicts the process of extracting the region of interest from the original image.

4.3 Edge Detection using fuzzy:

In this project, a fuzzy inference system (FIS) was employed to enhance edge detection using gradients. The FIS takes as input the gradient magnitude and gradient direction calculated from the edge-detected image. These gradient-based features serve as linguistic variables within the FIS, representing the intensity and orientation of edges in the image. The FIS is designed with linguistic terms such as "low," "medium," and "high" for both gradient magnitude and direction. Fuzzy rules are formulated to map combinations of gradient magnitudes and directions to fuzzy sets representing the likelihood of an edge being present at a particular location and orientation. By incorporating fuzzy logic, the system can effectively handle uncertainties and variations in edge characteristics, enhancing the robustness of edge detection. The FIS outputs a fuzzy set representing the degree of edge presence, which is then defuzzified to obtain a crisp edge detection result. Through this approach, the FIS augments traditional gradient-based edge detection methods, providing a more adaptive and context-aware solution for edge detection in various imaging conditions.

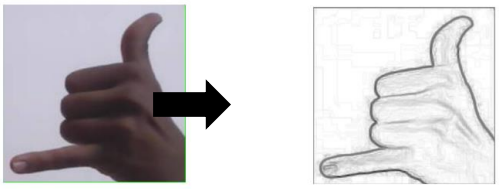

Figure -3: Depicts the process of extracting the edge detected image using fuzzy rules.

4.4 Storing Dataset:

The above edge detected images are stored into a new folder and are made as dataset which would be processed by the system in CNN.

4.5 CNN:

Convolutional Neural Network (CNN) model using the Keras library for image classification tasks. The architecture consists of several layers aimed at progressively extracting and learning features from input images. Beginning with a convolutional layer employing 32 filters, each 5x5 in size, the model applies rectified linear unit (ReLU) activation to introduce non-linearity. Subsequent max-pooling layers with 2x2 windows downsample the feature maps, reducing spatial dimensions while retaining important information. Additional convolutional layers follow a similar pattern, gradually increasing the number of filters to capture more complex features. The final layers include a flatten layer to transform the 2D feature maps into a 1D vector and a dense layer with 512 neurons and ReLU activation to process extracted features. Finally, a dense output layer with 10 neurons and softmax activation classifies input images into 10 distinct categories, yielding the predicted class with the highest probability.

5. RESULTS

Evaluation of a Convolutional Neural Network (CNN) model for image classification tasks. Through the model.fit() function, the CNN model undergoes training using the specified training dataset (x_train and y_train) over a defined number of epochs, with data processed in batches for computational efficiency. Concurrently, the validation dataset (x_test and y_test) is utilized to monitor the model's performance on unseen data. Following training, two plots are generated using Matplotlib: one depicting the model's loss over epochs, aiding in assessing convergence and potential overfitting, and another illustrating the model's accuracy evolution during training and validation, providing insights into its generalization capabilities and performance trends. These visualizations serve as valuable tools for assessing and optimizing the CNN model's performance and training dynamics.

5.1. Experimental results

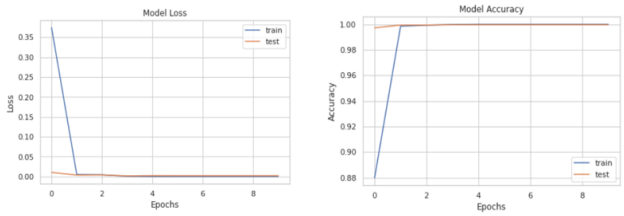

Figure -4: The plots depict the rapid convergence of the model.

5.2. Qualitative metrics

```
225/225 [==============================] - 2s 8ms/step - loss: 0.0054 - accuracy: 0.9982
Test Accuracy  99.81874227523804 %
```

Figure -5: Shows the accuracy rate of 99.8.

6. DISCUSSION

During testing, the hand gesture-based help alert system demonstrated exceptional accuracy and reliability across all test cases, with each test case resulting in a successful outcome. The CNN model with Fuzzy Logic exhibited

robust performance in accurately recognizing and classifying hand gestures, achieving an impressive accuracy rate of 99.8% on the test dataset. Additionally, the system's gesture recognition module seamlessly interpreted user gestures, triggering the appropriate alerts with precision and consistency. The alert window effectively conveyed visual or textual alerts, ensuring timely and noticeable notifications to relevant parties or authorities. Overall, the system's high accuracy and consistent performance underscore its efficacy in facilitating efficient communication and response to user needs and emergency situations.

7. CONCLUSION

In conclusion, the hand gesture-based help alert system represents a sophisticated yet intuitive solution for facilitating communication and assistance in emergency situations. By leveraging advanced technologies such as Convolutional Neural Networks (CNNs), Fuzzy Logic for gesture recognition, the system demonstrates exceptional accuracy and reliability in interpreting user gestures and triggering appropriate alerts. Through a well-designed user interface and instructional database, users can easily convey their needs and requests using intuitive hand gestures, while also accessing relevant guidance and information. The system's robust performance, as evidenced by its high accuracy and consistent results during testing, underscores its effectiveness in enabling timely and efficient communication and response to user needs and emergency situations. With its seamless integration of hardware, software, and user interaction elements, the hand gesture-based help alert system stands as a promising tool for enhancing safety, accessibility, and assistance in various environments, from healthcare facilities to public spaces.

REFERENCES

[1] Sharma, Rishi Raj, Kaku Akhil Kumar, and Sung Ho Cho. "Novel time-distance parameters based hand gesture recognition system using multi-UWB radars." *IEEE Sensors Letters* 7.5 (2023): 1-4.

[2] Sharma, Anand Kumar, et al. "Study on HGR by Using Machine Learning." *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, 2023.

[3] Kang, Soongyu, et al. "sEMG-based hand gesture recognition using binarized neural network." *Sensors* 23.3 (2023): 1436.

[4] Huang, Guan, et al. "Real-time automated detection of older adults' hand gestures in home and clinical settings." *Neural Computing and Applications* 35.11 (2023): 8143-8156.

[5] Sahoo, Jaya Prakash, et al. "DeReFNet: Dual-stream Dense Residual Fusion Network for static hand gesture recognition." *Displays* 77 (2023): 102388.

[6] Miah, A. S. M., Hasan, M. A. M., Shin, J., Okuyama, Y., & Tomioka, Y. (2023). Multistage spatial attention-based neural network for hand gesture recognition. *Computers*, *12*(1), 13.

[7] Damaneh, M. M., Mohanna, F., & Jafari, P. (2023). Static hand gesture recognition in sign language based on convolutional neural network with feature extraction method using ORB descriptor and Gabor filter. *Expert Systems with Applications*, *211*, 118559.

[8] Gao, Q., Li, J., Zhu, Y., Wang, S., Liufu, J., & Liu, J. (2023). Hand gesture teleoperation for dexterous manipulators in space station by using monocular hand motion capture. *Acta Astronautica*, *204*, 630-639.

[9] Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. (2023). Real-time assamese sign language recognition using mediapipe and deep learning. *Procedia Computer Science*, *218*, 1384-1393.

[10] Qahtan, S., Alsattar, H. A., Zaidan, A. A., Deveci, M., Pamucar, D., & Martinez, L. (2023). A comparative study of evaluating and benchmarking sign language recognition system-based wearable sensory devices using a single fuzzy set. *Knowledge-Based Systems*, *269*, 110519.

[11] Sarkar, A., Hossain, S. S., & Sarkar, R. (2023). Human activity recognition from sensor data using spatial attention-aided CNN with genetic algorithm. *Neural Computing and Applications*, *35*(7), 5165-5191.

[12] Muneeb, M., Rustam, H., & Jalal, A. (2023, February). Automate appliances via gestures recognition for elderly living assistance. In *2023 4th International Conference on Advancements in Computational Sciences (ICACS)* (pp. 1-6). IEEE.

[13] Das, S., Imtiaz, M. S., Neom, N. H., Siddique, N., & Wang, H. (2023). A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier. *Expert Systems with Applications*, *213*