A Convolutional Neural Network Framework for Robust Hand Gesture Recognition

Rameesa A B[1]*, Bismin V Sherif*[2]

[1]*Dept. of Computer Applications, MES College Marampally, Kerala, India*

[2]*Dept. of Computer Applications, MES College Marampally, Kerala, India*

Abstract-Hand gesture recognition has become increasingly relevant in enhancing human-computer interaction across various applications, from virtual reality to assistive technologies. This paper introduces a novel approach using Convolutional Neural Networks (CNNs) to accurately recognize hand gestures and uniquely convert the recognized gestures into audio output. The system employs advanced preprocessing and training techniques to ensure high accuracy. The effectiveness of the proposed CNN model is rigorously compared with traditional models, including K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN). Experimental results demonstrate the superior performance of the CNN-based approach, offering a robust and innovative solution for real-time gesture recognition and audio feedback.

Keywords: Hand Gesture Recognition, Convolution Neural Network, k-Nearest Neighbors, Random Forest, SVM

1. INTRODUCTION

The method of automatically recognizing and deciphering hand movements is known as hand gesture recognition, or HGR. To deduce the intended meaning or action, it entails recording, examining, and comprehending the hand movements and configurations.In a variety of fields, such as virtual reality, robotics, sign language translation, and human-computer interface, hand gesture recognition, or HGR, is essential. A new method for HGR using Convolutional Neural Networks (CNN) is presented in this research and is compared to more conventional classifiers like Support Vector Machines (SVM), Artificial Neural Networks (ANN), k-Nearest Neighbors (KNN), and Random Forest. In hand gesture detection tests, CNNs have shown remarkable performance, with high accuracy rates and robustness to variations in hand positions, backdrops, and lighting conditions. The suggested CNN-based approach eliminates the need for human feature engineering by utilizing deep learning's power to automatically identify discriminative features from unprocessed pixel data. After a thorough testing process on a benchmark dataset of hand gesture images taken in various lighting scenarios, backgrounds, and hand pose scenarios, each classifier's effectiveness is assessed using performance metrics like accuracy, precision, recall, and F1-score. Here im use Image dataset which includes 10 classes such as call_me, rock, rock_on, fingers_crossed, peace, paper, okay, thumbs, up, scissor. The dataset have 5000+ images, each class have 500 above images of the above mentioned gestures different feature. The results of the trial highlight the better performance. Increasing the accuracy of correctly detected motions is the goal. Comparing the actual label of the image with the anticipated label allows one to assess the model's accuracy.

2. RELATED WORKS

Dandu Amarnatha Reddy et al. [1] conducted a research study titled "Hand Gesture Recognition Using Local Histogram Feature Descriptor." The system focused on hand gesture recognition for enhancing human-computer interaction, employing a vision-based approach with three main stages: preprocessing, feature extraction, and classification. In the preprocessing stage, the goal was to localize the hand region within the image frame. The authors utilized the Laplacian of Gaussian filtering technique in conjunction with a zero-crossing detector to identify the edges of the hand region in hand gesture images. A novel feature extraction technique, the Local Histogram Feature Descriptor (LHFD), was proposed in this paper. This method involves extracting features by computing the local histogram of the grayscale *gesture image, utilizing the entire hand region. Importantly, the proposed method demonstrated invariance to scaling and illumination changes. The evaluation of the proposed technique was conducted on two standard datasets: the Massey University Gesture Dataset (MUGD) and Jochen Triesch Static Hand Posture Database. The recognition performance of the proposed technique was reported as 99.5% for the Massey University Gesture Dataset and 95% for the Triesch dataset. The evaluation utilized a multi-class support vector machine (SVM) classifier for classification purposes.In summary, the study presented a comprehensive approach to hand gesture recognition, introducing a novel feature extraction method (LHFD) that exhibited strong performance on standard datasets, showcasing its effectiveness in real-world applications.*

In [2] develop a study on "Gaze-aware hand gesture recognition for intelligent construction." proposing an innovative framework serving as a human–robot interface. This framework aims to address limitations in existing hand gesture approaches for robot–worker collaboration, encompassing three key components: visual detection and tracking, machine-of-interest generation, and hand gesture recognition. The study includes a validation test to evaluate precision and recall performance, demonstrating that the proposed framework is effective for facilitating interaction between workers and multiple construction machines. Implemented on a Windows 10 64-bit operating system, the framework utilizes Python 3.6 with support from PyTorch and Tensorflow platforms, incorporating essential algorithms, functions, and tools. Notably, this research marks the first integration of gaze tracking and gesture recognition for collaborative interactions with construction equipment. The gaze-aware hand gesture recognition framework achieved a precision of 93.8% and a recall of 95.0% during the validation test, highlighting its suitability for one-to-many collaboration in construction applications.

Ji-Won Lee and Kee-Ho Yu [3] proposed a study titled "Wearable Drone Controller: Machine Learning-Based Hand Gesture Recognition and Vibrotactile Feedback." The research introduced a wearable drone controller integrating hand gesture recognition and vibrotactile feedback. The control system utilized an inertial measurement unit (IMU) positioned on the back of the hand to detect hand motions for drone navigation. The recorded motions were categorized through machine learning employing the ensemble method, achieving a classification accuracy of 97.9%. Additionally, the controller incorporated vibrotactile feedback by relaying information about the distance to obstacles in the drone's heading direction. This feedback was delivered to the user through a vibration motor attached to the wrist. In simulated experiments with a participant group, the hand gesture control exhibited robust performance. The vibrotactile feedback proved beneficial in enhancing the user's awareness of the drone's operational environment, particularly in scenarios with limited visual information. To evaluate the proposed controller, a subjective assessment was conducted with participants to gauge its convenience and effectiveness. Subsequently, a real drone experiment validated the applicability of the controller as a natural interface for drone operation. The average accuracy in direct mode was approximately 96%, exhibiting a marginal decrease of 2.6% compared to the simulation. Conversely, the average accuracy in gesture mode was around 98%, indicating a slight improvement of 1.4% over the simulation results.

E. Stergiopoulou and N. Papamarkos [4] presented a study on "Hand Gesture Recognition using a Neural Network Shape Fitting Technique." The paper introduces an innovative approach to hand gesture recognition based on extracting hand gesture features and employing a neural network shape fitting method. Initially, a skin color filtering process is applied in the YCbCr color space to swiftly isolate the hand region, ensuring noiseless segmented images despite variations in skin color and lighting conditions. The subsequent stages involve fitting the shape of the hand and recognizing the finger configuration. The developed hand gesture recognition system, implemented in Delphi, underwent testing with hand images from diverse individuals, considering variations in morphology, slope, and size. The system was trained to identify 31 hand gestures, involving combinations of raised and not raised fingers, facilitating human-computer communication without the need for specialized hardware. Extensive testing of the proposed system with a substantial number of input images yielded a highly promising recognition rate.

The success of the system is demonstrated through rigorous testing.

Nahla Majdoub Bhiri et al.[5] proposed a method " Hand gesture recognition with focus on leap motion: An overview, real world challenges and future directions." Researchers have investigated a variety of sensors for data collecting in the development of Hand Gesture Recognition (HGR) systems, giving consideration to elements including accuracy, precision, and cost-effectiveness. It has been determined by comparative study that the LMC (Leap Motion Controller) is the best option since it provides a better price-performance ratio for HGR applications. Thus, it has been noticed that LMC is widely being adopted in a variety of fields, such as virtual reality interfaces, robotics, educational technology, home assistance devices, sign language interpretation, and medical applications. Even with the widespread application of LMC, problems remain in the field of HGR, which drives continued research and development to overcome these barriers and enhance the capabilities of gesture recognition technology.

Weina Zhou and Kun Chen [6] Conducted a research on " A light weight hand gesture recognition in a complex background." introduced a two-stage Hand Gesture Recognition (HGR) system designed to address challenges in real-world scenarios, particularly the recognition of hand gestures in complex backgrounds. The proposed system focuses on achieving a balance between lightweight design and high recognition accuracy. In the initial stage, the authors employed an accurate segmentation process to separate the hand from the complex background. This segmentation network utilized a combination of a dilated residual network (DRN), an atrous spatial pyramid pooling module (ASPP), and a simplified decoder. The DRN and ASPP were utilized for precise segmentation, while the simplified decoder refined the segmented hand regions. The second stage of the HGR system introduced double-channel Convolutional Neural Networks (CNNs) to enhance recognition performance. This involved taking both the segmented hand image and the original RGB image as input to the CNNs, with the extracted features from both images fused to produce the final HGR results. The authors highlighted the significance of HGR in machine vision, emphasizing its applications in intelligent driving, machine control, and virtual reality. The proposed vision-based HGR system aimed to overcome challenges related to hand recognition in complex backgrounds. The two-stage method consisted of hand segmentation and hand gesture recognition, utilizing an encoder-decoder framework in the first stage, including DRN, ASPP module, and a simplified decoder. The study utilized the OUHANDS dataset, known for its complex backgrounds, featuring ten different gesture classes from 23 subjects with background disturbances. The proposed model demonstrated notable advancements, achieving a duration accuracy of 91.17% with a model size of 1.8 MB. These results surpassed other state-of-the-art models in hand gesture recognition, highlighting the effectiveness of the two-stage HGR system in complex background scenarios.

In [7] develop a study "Comparing EMG-based hand gesture recognition (HGR) systems employing supervised and reinforcement learning." While many HGR methods rely on supervised machine learning (ML), the application of reinforcement learning (RL) for EMG classification remains underexplored. The performance of HGR systems based on ML and RL methods for user-general HGR on large datasets is an ongoing research challenge. This study compares supervised learning, featuring k-nearest neighbors (K-NN), support vector machine (SVM), artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNNs) like long-short-term memory (LSTM), with a reinforcement learning approach. The HGR systems consist of pre-processing, feature extraction using CNN, classification, and post-processing stages. Both supervised and RL models were evaluated for six hand gestures. The learning agents in both cases use CNN for feature extraction. The proposed models were tested on a validation set of 306 users for the supervised method and 40 users for the RL method. The supervised learning method achieved the best results with an accuracy of 90.49% (+9.7%).

JP.Vasconez et al. [8] conducted a study on "Hand Gesture Recognition Using EMG-IMU Signals and Deep Q-Networks." They introduced an HGR system based on the DQN algorithm for classifying 11 distinct hand gestures, encompassing both static and dynamic gestures. The research involved the evaluation and comparison of two sensors, namely the Myo armband

and G-force sensors. EMG and IMU signals were utilized to derive feature vectors from these sensors. The proposed models underwent validation on 43 users and testing on an additional 42 users. The Myo armband sensor exhibited superior performance, achieving a classification accuracy of up to 97.50%±1.13% and 88.15%±2.84% for static gestures, and 98.95%±0.62% and 90.47%±4.57% for dynamic gestures in terms of classification and recognition, respectively. The study demonstrated that the DQN effectively learned a policy from online experience to classify and recognize gestures based on EMG and IMU signals, surpassing results obtained by similar methods using only EMG. Furthermore, the Myo armband sensor outperformed the G-force sensor in terms of accuracy and data distribution.

In [9] develop a method " Hand Gesture Recognition using Image Processing and Feature Extraction Techniques." The suggested ORB feature extraction method has been thoroughly tested on the same dataset using a variety of pre-processing techniques, such as Histogram of Gradients, LBP, and PCA. Prominent classifiers like as KNN, SVM, Random Forest, Naïve Bayes, Logistic Regression, and Multi-Layer Perceptron (MLP) have been used to evaluate these approaches. Notably, PCA performs better for MLP, Random Forest, and SVM classifiers whereas the suggested strategy performs better than other pre-processing techniques for Naïve Bayes, Logistic Regression, and KNN classifiers. The current system has the ability to recognize dynamic motions in videos in real-time, but it is restricted to static gesture images, even though it achieves high accuracy in gesture detection. Moreover, RGBD images from Kinect Sensors can be recognized by the system through adaptation.

S.Tiwari et al. [10] propose a method on the development of a" hand gesture-based volume controller." utilizing the OpenCV module for gesture detection and control. The system captures images and videos from the webcam and adjusts the volume based on user gestures, providing a hands-free approach to volume control. This technology is particularly beneficial in scenarios where direct device access is challenging or when users prefer discreet volume adjustments. A key advantage highlighted by the authors is the accessibility of the hand gesture volume controller, enabling individuals with physical disabilities to manage device volume without relying on physical buttons or remote controls. They emphasize the potential revolutionary impact of this technology on human-device interactions. The research involves the use of a camera or sensor to capture user hand gestures, with a focus on addressing challenges related to background images or videos that may affect gesture recognition quality. The authors emphasize the use of open-source software and hardware to enhance accessibility and ease of replication for those interested in building similar volume controllers using hand gestures. To achieve their goals, the authors utilized essential packages such as ImUtils, OpenCV-Python, SciPy, TensorFlow, NumPy, and MediaPipe. The gesture recognition process using an Artificial Neural Network (ANN) typically involves steps such as data collection, data preprocessing, feature extraction, ANN training, testing, and evaluation. The authors evaluated their proposed system using a dataset comprising 50 different hand gestures, including actions such as decreasing volume, increasing volume, reaching minimum and maximum volume, and mute. The reported results demonstrated a success rate exceeding 95%, showcasing the effectiveness of the hand gesture volume controller.

3. PROPOSED ARCHITECTURE

Hand Gesture Recognition (HGR) is an essential procedure that is used in many different fields, such as virtual reality, robotics, sign language translation, and human-computer interfaces. It is the process of automatically recognizing and interpreting hand gestures to infer their intended meaning or action. This paper introduces a robust technique to HGR using Convolutional Neural Networks (CNN) along with the facility for text to voice converter and compares it with traditional classifiers like random forest, SVM, K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN).

The architecture of the proposed approach is shown in figure 1. The different modules involved in the proposed approaches are input module, preprocessing module, CNN module, output module, Text to voice converter module and a comparison module. The input module is the initial stage of the hand gesture *recognition system, responsible for capturing and reading the input image. This module typically utilizes a camera or other image acquisition device to obtain visual data of hand gestures. The captured image is then fed into the system, serving as the raw data for subsequent processing stages. Ensuring high-quality image capture is crucial, as the clarity and accuracy of the input image directly impact the performance of the entire recognition system.*

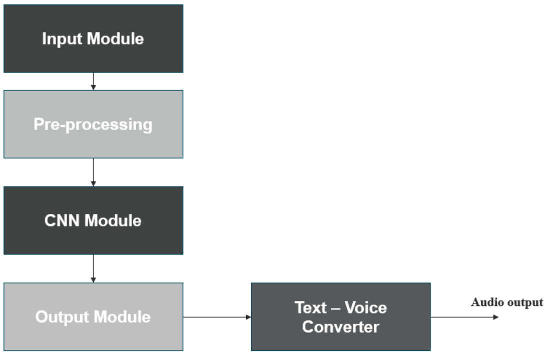

Figure 1 Proposed System Architecture

The preprocessing module ensures that the input data is properly formatted, normalized, and split into training and testing sets, making it ready for training the CNN model. The preprocessing module plays a vital role in preparing the input image data for effective processing by the CNN model. This stage involves several critical steps. First, the input images are formatted to ensure consistency in size and resolution. Next, normalization is applied to scale pixel values, often between 0 and 1, to enhance the model's training efficiency. Finally, the dataset is divided into training and testing sets, ensuring that the model has sufficient data to learn from and that its performance can be evaluated accurately. Proper preprocessing is essential for achieving high model accuracy and reliability.

A convolutional neural network (CNN) is a type of deep learning algorithm specifically designed for image processing and recognition tasks. convolutional neural network consists of multiple layers like the input layer, convolutional layer, pooling layer, fully connected layer, and output layers. The CNN module is the core of the hand gesture recognition system, leveraging the power of deep learning for image analysis. A Convolutional Neural Network consists of several specialized layers that work together to extract features and classify images. The input layer receives the preprocessed image, which is then passed through convolutional layers where filters detect essential features such as edges and textures. Pooling layers reduce the dimensionality of the data, making the computation more efficient. The fully connected layer combines these features to form a high-level understanding of the image. Finally, the output layer produces the classification result, identifying the hand gesture. Each layer in the CNN contributes to the model's ability to learn and recognize complex patterns in the input images.

The output module is a critical component of the CNN architecture, responsible for generating the final predictions based on the processed data. After the input image has passed through the various convolutional, pooling, and fully connected layers, the output layer produces a probability distribution over the possible gesture classes. This module typically uses a softmax activation function to convert the raw scores into probabilities, allowing the system to identify the most likely gesture. The accuracy and reliability of the output module are paramount, as they directly influence the system's effectiveness in real-world applications.

The text to voice converter module adds an interactive and user-friendly dimension to the hand gesture recognition system. Once the CNN model predicts the gesture class, this module translates the classification result into audible speech. This conversion is achieved using text-to-speech (TTS) technology, which synthesizes a human-like voice to announce the recognized gesture. This feature is particularly beneficial in applications where visual feedback may not be practical or sufficient, such as assisting individuals with visual impairments or providing hands-free interaction in various environments.

4. RESULTS AND DISCUSSION

In the context of hand gesture detection, CNNs have shown exceptional performance, exhibiting high accuracy rates and robustness to changes in hand locations, backdrops, and lighting conditions. The suggested CNN-based approach does away with the need for human feature engineering by using deep learning to automatically extract discriminative features from unprocessed pixel data. This work uses the image dataset which includes 10 classes such as call_me, rock, rock_on, fingers_crosssed, peace, paper, scissor, okay, thumbs, up. The sample dataset is shown in figure 2.

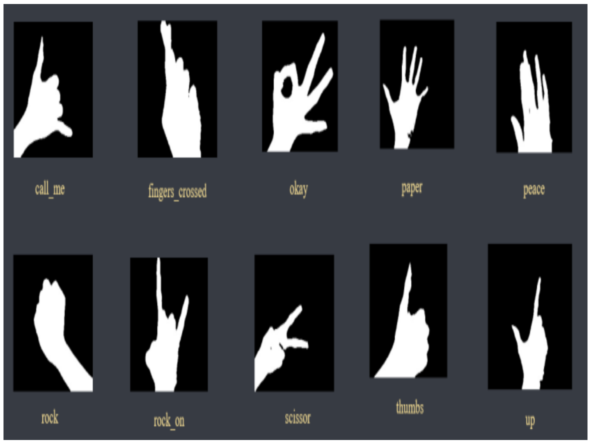

Figure 2 Data set classes

A comprehensive testing approach is carried out on a benchmark dataset comprising hand gesture photographs recorded under various settings in order to evaluate the effectiveness of each classifier. The accuracy, precision, recall, and F1-score are among the performance indicators used to assess the effectiveness of each classifier. The experimental results highlight the higher performance of the CNN-based method.

In this study, Convolutional Neural Networks (CNN) are introduced and contrasted with more traditional classifiers such as Random Forest, ANN, SVM, k-Nearest Neighbors (KNN), and Convolutional Neural Networks. CNNs have performed remarkably well in hand gesture identification experiments, with excellent accuracy rates and resilience to changes in hand locations, backgrounds, and lighting. The comparison of different approaches are given in table 1. Given their superior scalability, accuracy, and resilience over existing methods, CNNs are anticipated to play a significant role in the advancement of hand gesture detection in the future. As deep learning and computing power increase, CNNs should become ever more capable and versatile tools for solving challenging image recognition problems.

Table 1 Comparison of different approaches

Model	Accuracy	Precision	Recall	F-score
CNN	0.9724	0.9725	0.9723	0.9722
KNN	0.9151	0.9218	0.9151	0.9166
Random Forest	0.9103	0.9128	0.9103	0.9103
ANN	0.7455	0.7481	0.7454	0.7414
SVM	0.7397	0.7421	0.7397	0.7385

5. CONCLUSION

With a focus on the use of convolutional neural networks (CNNs) and a comparison with more conventional classifiers like support vector machines (SVM), artificial neural networks (ANN), k-nearest neighbor (KNN), and random forests, this paper provides a thorough overview of hand gesture recognition (HGR) techniques. It is clear from thorough testing and analysis that CNNs perform better in hand gesture identification tasks, obtaining high accuracy rates and resilience to a range of environmental conditions. Used the CNN for bulding the model and it achieved promising results in accurately predicting hand gestures from different classes. It is evident from the outcome analysis that our model produces higher output with more efficiency.In the future, expect more research and model improvement to open the door to more complex and precise hand gesture classification systems.

A number of themes are developing in hand gesture detection as technology progresses, suggesting fascinating prospects for the future, Improved Deep Learning Models, Edge Computing for Real-Time Processing, Fusion of Multiple Sensors, Enhanced Gesture Vocabulary, and others are some of the major upcoming advances in hand gesture detection.

REFERENCE

[1] Danda Amarnatha Reddy, java prakash Sahoo, Samit Ari “Hand Gesture Recognition using local histogram Feature Descriptor,” Department of EC, NIT(2018),doi:10.1109/1C0EI.2018.8553849

[2] X.Wang, D.Veeramani, Z.Zhu“ Gaze-aware hand gesture recognition for intelligent construction,” University of Wisconsin-Madison, Engineering Applications of Artificial Intelligence (2023) doi: https: //doi.org/10.1016/j.engappai.2023.106179

[3] *Ji-won Lee, Kee-Ho Yu "Wearable Drone Controller: Machine Learning -Based Hnad Gesture Recognition and vibrotactile feedback," Joenhuk National University (2023), doi: https://doi.org. 10.3390/s23052666*

[4] *E.Stergiopoulou, N.Papamarkos "Hand gesture recognition using a neural network Shape fitting technique," Engineering Application of Artificial Intelligence(2009) doi: https://org./10.1016 /j.engapp ai.2009.03.008*

[5] *Nahla Majdoub Bhiri , Safa Ameur , Ihsen Alouani , Mohamed Ali Mahjoub , Anouar Ben Khalifa " Hand gesture recognition with focus on leap motion: An overview, real world challenges and future directions," Expert Systems with Applications(2023) doi https://doi.org/ 10.1016 /j.eswa.2023.120125*

[6] *Weina Zhou, Kun Chen "A Lightweight Hand Gesture Recognition In Complex Backgrounds," Shanghai Maritime University, Shanghai 201306, China (2022) doi: https://doi.org/ 10.1016/ j.displa. 2022.102226*

[7] *JP. Vasconez, LI. Barona Lopez, AL. Valdivieso Caraguay, ME. Benalcazar, "A Comparison of EMG- based hand gesture recognition systems based on supervised and reinforcement learning," Engineering Application of Artificial Intelligence (2023), doi: https://doi.org/ 10.1016/j.engappai .2023.106327*

[8] *JP. Vasconez, LI. Barona Lopez, AL. Valdivieso Caraguay, ME. Benalcazar "Hand gesture recognition using EMG-IMU Signals and Deep Q-Networks," Artificial Intelligence and Computer Vision Research Lab, Escuela Politécnica Nacional, Quito 170517, Ecuador (2022), doi: https://doi .org/10.3390/s22249613*

[9] *Ashish Sharma , Anmol Mittal , Savitoj Singh , Vasudev Awatramani "Hand Gesture Recognition using Image Processing and Feature Extraction Techniques," Procedia Computer Science(2020) doi:https://doi.org/10.1016 /j.procs.2020.06.022*

[10] *S. Tiwari, A. Mishra, D. Kukreja and A. L. Yadav, "Volume Controller using Hand Gestures," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10308134*