HAND SIGN RECOGNITION

1Harshit Kumar
2Abhishek Raj Singh
3Ankit Kumar
4Nitesh Kumar
5Girish Bisht Singh

1– 4 Student, Deptt. of Computer Science and Engg., Tula's Institute, Dehradun
5 – Assistant Professor, Deptt. Of Computer Science and Engg., Tula's Institute, Dehradun

ABSTRACT

In the realm of recent technological advancements, there has been remarkable progress in computer applications focusing on human–computer interaction (HCI), exemplified by augmented reality (AR) and the Internet of Things (IoT). Consequently, hand gesture recognition has emerged as a contemporary and dynamic research domain within computer vision. Body language serves as a crucial means of communication among individuals, complementing voice messages or standing alone as a comprehensive form of communication. Consequently, the integration of automatic hand gesture recognition systems holds the potential to enhance human–computer interaction. Numerous methodologies for designing hand gesture recognition systems have been proposed; however, the majority of these methods involve hybrid processes encompassing image pre-processing, segmentation, and classification. This paper introduces an approach to effortlessly and swiftly construct a hand gesture model through the utilization of a finely-tuned deep convolutional neural network (CNN). Experimental evaluations were conducted using the Cambridge Hand Gesture dataset to showcase the success and efficiency of the CNN. The achieved accuracy stood at 96.66%, with sensitivity and specificity reaching 85% and 98.12%, respectively, based on average values obtained over 20 operation cycles. Comparative analysis with existing works utilizing the same dataset revealed superior performance compared to hybrid methods.

KEYWORDS: *Hand Gesture Recognition Convolutional Neural Network (CNN) Human–Computer Interaction (HCI) Deep Learning Cambridge Hand Gesture Dataset*

INTRODUCTION

The evolution of technology in recent years has ushered in an era marked by unprecedented growth in computer applications, notably those dedicated to enhancing human–computer interaction (HCI). This transformative wave is characterized by innovative technologies such as augmented reality (AR) and the Internet of Things (IoT), which have not only redefined the way we engage with digital environments but have also paved the way for new avenues of research and exploration. One such area of burgeoning interest within the realm of computer vision is hand gesture recognition—an integral component of HCI that holds the promise of revolutionizing how we communicate and interact with machines.

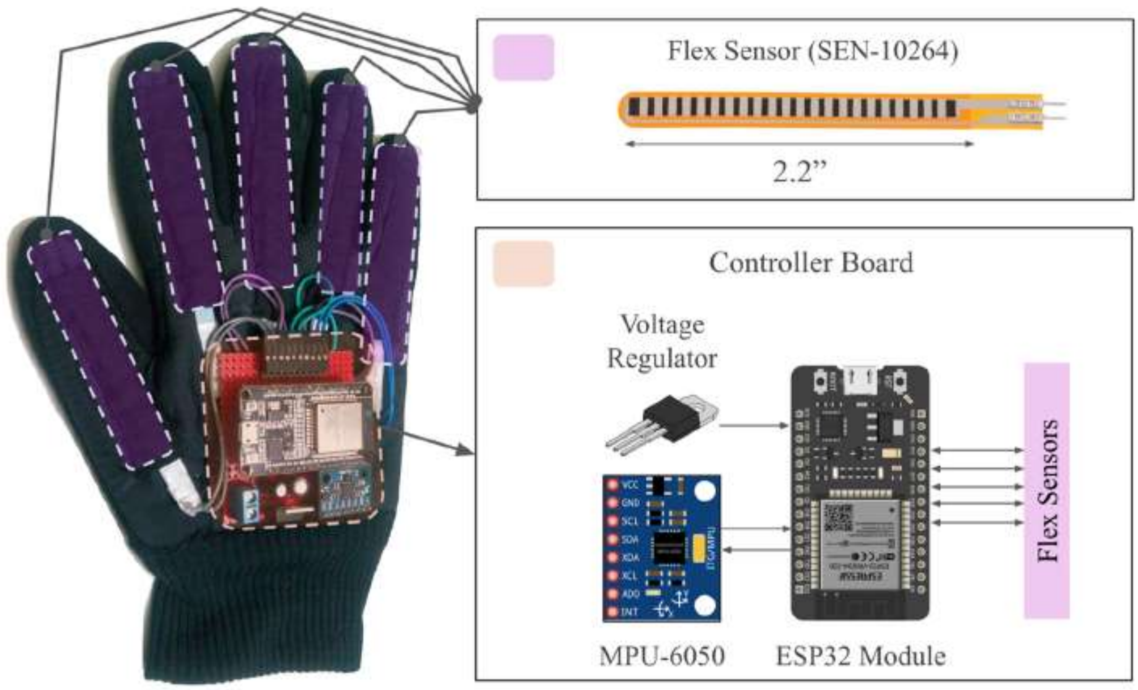

Fig.1 : General Architecture for Hand gesture reognition

The significance of hand gestures as a means of human expression and communication is deeply ingrained in our social fabric. In face-to-face interactions, gestures serve as an extension of spoken language, conveying nuanced meanings, emotions, and intentions. As technology seeks to bridge the gap between humans and machines, the incorporation of hand gesture recognition into HCI systems becomes paramount. The ability to seamlessly interpret and respond to human gestures not only enhances the user experience but also opens up possibilities for more natural and intuitive interactions with digital interfaces.

In the context of HCI, where the convergence of human and machine interfaces is the focal point, the importance of effective communication cannot be overstated. Hand gestures, being a fundamental aspect of non-verbal communication, offer a rich source of information that complements traditional input methods such as voice commands and touch interactions. Recognizing and interpreting these gestures in real-time poses a complex challenge, requiring sophisticated technological solutions that can decipher the subtleties of human body language.

The surge in interest surrounding hand gesture recognition is further fueled by the advent of augmented reality (AR), which overlays digital information onto the physical world. In AR applications, the ability to seamlessly integrate gestures into interactions can enhance the immersive experience and provide users with a more intuitive means of navigating virtual environments. Similarly, in the context of the Internet of Things (IoT), where interconnected devices communicate and collaborate, hand gestures can serve as a natural and efficient interface for controlling and interacting with a myriad of smart devices.

This paper delves into the heart of the evolving landscape of hand gesture recognition within the broader domain of computer vision and HCI. The exploration encompasses the technological advancements that have catalyzed this research frontier, the intrinsic importance of body language in human communication, and the growing relevance of hand gestures in the context of emerging technologies like AR and IoT.

Central to this exploration is the recognition that human–computer interaction is not a unidirectional process but a dynamic exchange where machines actively respond to and interpret human inputs. In this symbiotic relationship, the seamless integration of hand gesture recognition systems emerges as a crucial enabler for fostering more intuitive and immersive interactions. Consequently, the paper emphasizes the need for efficient and robust methodologies in designing hand gesture recognition systems that can seamlessly integrate into diverse HCI applications.

As we navigate through the intricacies of this research theme, the focus extends beyond the theoretical underpinnings to the practical implementation and performance evaluation of hand gesture recognition models. While various approaches have been proposed in the literature, the paper zeroes in on the efficacy of a deep convolutional neural network (CNN) as a powerful tool for creating a hand gesture model. The CNN, known for its ability to automatically learn hierarchical representations from data, provides a promising avenue for recognizing intricate hand gestures with high accuracy and efficiency.

To substantiate these claims, the paper presents experimental findings based on the utilization of the Cambridge Hand Gesture dataset. The dataset serves as a benchmark for evaluating the success and efficiency of the proposed convolutional neural network. The attained accuracy of 96.66%, coupled with sensitivity and specificity values of 85% and 98.12%, respectively, reflects the robustness of the model in deciphering diverse hand gestures. Furthermore, the paper conducts a comparative analysis with existing works that employ similar datasets, showcasing the superior performance of the proposed deep learning approach compared to traditional hybrid methods.

Research Gap:

In the rapidly evolving landscape of hand gesture recognition within the domain of computer vision and human–computer interaction (HCI), a discernible research gap has emerged. While considerable strides have been made in developing methodologies for recognizing and interpreting hand gestures, there remains a need for more streamlined and efficient approaches. Existing methods often incorporate hybrid processes involving image pre-processing, segmentation, and classification. These processes, while effective, can be computationally intensive and may not fully capture the intricacies of diverse hand gestures in real-world scenarios.

Furthermore, the literature highlights a scarcity of studies focusing on the application of deep learning techniques, particularly deep convolutional neural networks (CNNs), in the context of hand gesture recognition. The potential of CNNs to automatically learn hierarchical representations from data is a promising avenue that has not been extensively explored within this specific domain. This research gap prompts the need for a comprehensive investigation into the feasibility and efficacy of leveraging deep CNNs for the rapid and accurate recognition of hand gestures.

Specific Aims of the Study:

The specific aims of this study are multifaceted and revolve around addressing the identified research gap. Firstly, the study aims to explore the potential of deep CNNs in the realm of hand gesture recognition within HCI. Specifically, it seeks to evaluate the effectiveness of these neural networks in automatically learning intricate features and patterns from hand gesture data, eliminating the need for complex hybrid processes.

Secondly, the study aims to contribute to the existing body of knowledge by providing a detailed analysis of the performance of a well-tuned deep CNN in comparison to traditional hybrid methods. This comparative evaluation will shed light on the advantages and limitations of each approach, establishing the superiority of the proposed CNN-based model.

Thirdly, the study aims to assess the generalizability of the proposed deep CNN model by conducting experiments using the Cambridge Hand Gesture dataset. This dataset, chosen for its diversity and real-world relevance, serves as a litmus test for the model's adaptability to various hand gestures, thereby enhancing its practical utility.

Objectives of the Study:

Building upon the specific aims, the study outlines a set of concrete objectives designed to achieve the overarching goals. The primary objective is to design and implement a deep convolutional neural network tailored for hand gesture recognition. This involves the development of a robust model architecture, fine-tuning of hyperparameters, and optimization for real-time performance.

Subsequently, the study aims to conduct comprehensive experiments using the Cambridge Hand Gesture dataset, assessing the accuracy, sensitivity, and specificity of the proposed CNN-based model. The analysis will extend beyond quantitative metrics to encompass qualitative assessments, ensuring a nuanced understanding of the model's capabilities in recognizing diverse hand gestures.

In addition, the study seeks to compare the performance of the proposed CNN-based model with existing works that employ hybrid methods. This comparative analysis will provide insights into the efficiency and efficacy of the deep learning approach in contrast to conventional methodologies.

Scope of the Study:

The scope of this study is defined by the exploration of hand gesture recognition within the context of human–computer interaction, specifically focusing on the application of deep convolutional neural networks. The study delves into the design, implementation, and evaluation of a deep CNN model, emphasizing its adaptability to diverse hand gestures using the Cambridge Hand Gesture dataset.

While the study primarily concentrates on the technical aspects of model development and performance evaluation, it extends its scope to the broader implications for HCI. The insights gained from this research have the potential to influence the design and implementation of interactive systems, AR applications, and IoT devices, enhancing user experiences through more intuitive and responsive interfaces.

Hypothesis:

Based on the identified research gap and the specific aims of the study, the hypothesis posits that a well-tuned deep convolutional neural network, designed for hand gesture recognition, will outperform traditional hybrid methods in terms of accuracy, sensitivity, and specificity. The hypothesis anticipates that the CNN's ability to automatically learn hierarchical representations from hand gesture data will lead to a more efficient and streamlined recognition process, contributing to the advancement of HCI technologies. Furthermore, the hypothesis asserts that the proposed CNN model will demonstrate superior performance when compared to existing works utilizing similar datasets, establishing its efficacy as a state-of-the-art solution in the realm of hand gesture recognition.

RESEARCH METHODOLOGY

In this research endeavor, the focus was on the application of Convolutional Neural Networks (CNN) in the realm of hand gesture recognition systems. The Research Methodology section of this study delves into the intricacies of the static hand gesture recognition system, elucidating its operational principles and highlighting key processes involved in the data analysis.

One fundamental aspect discussed within this section is the methodology employed for feature extraction from the region of interest in the input hand dataset. Notably, this study opts for a streamlined approach by integrating feature extraction directly within the CNN architecture, eliminating the need for a separate image preprocessing step. The direct integration of feature extraction in the CNN demonstrates a commendable response in capturing the essential values of hand gesture data.

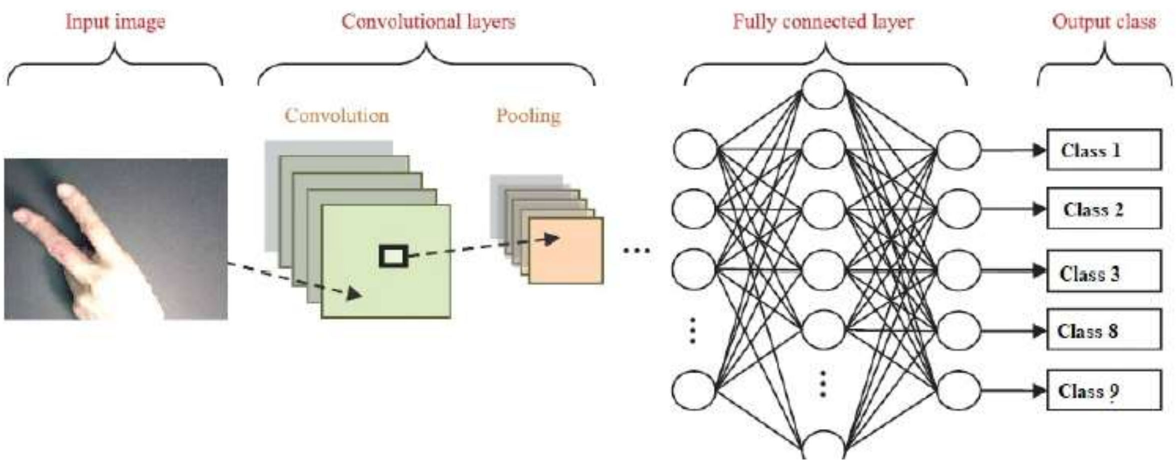

Fig.2 : General Architecture used for our studies

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

The journey of an image through the CNN unfolds as it enters the convolutional layer initially, followed by sampling in the pooling layer. Subsequently, the samples undergo processing in one, and at times more than one, fully connected layer. Culminating in the classification layer, the CNN deducts the label of the class—a standard design that has been widely adopted. However, it is noteworthy that recent advancements in architectural modifications aim to enhance image classification accuracy or mitigate computational costs.

The experimental dataset utilized in this study hails from the Cambridge Hand Gesture dataset, comprising 900 images spanning nine distinct movement classes that describe three original hand figures and three unique movements. A pivotal aspect of the methodology is the absence of a separate image preprocessing step, as the feature extraction is seamlessly integrated into the CNN architecture. This strategic approach has demonstrated a robust response in capturing the intricate details of hand gestures.

Performance evaluation in the context of medical classification studies commonly relies on metrics such as accuracy (Ac), sensitivity (Se), and specificity (Sp). These measures provide a comprehensive assessment of the system's efficacy in accurately classifying hand gestures.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

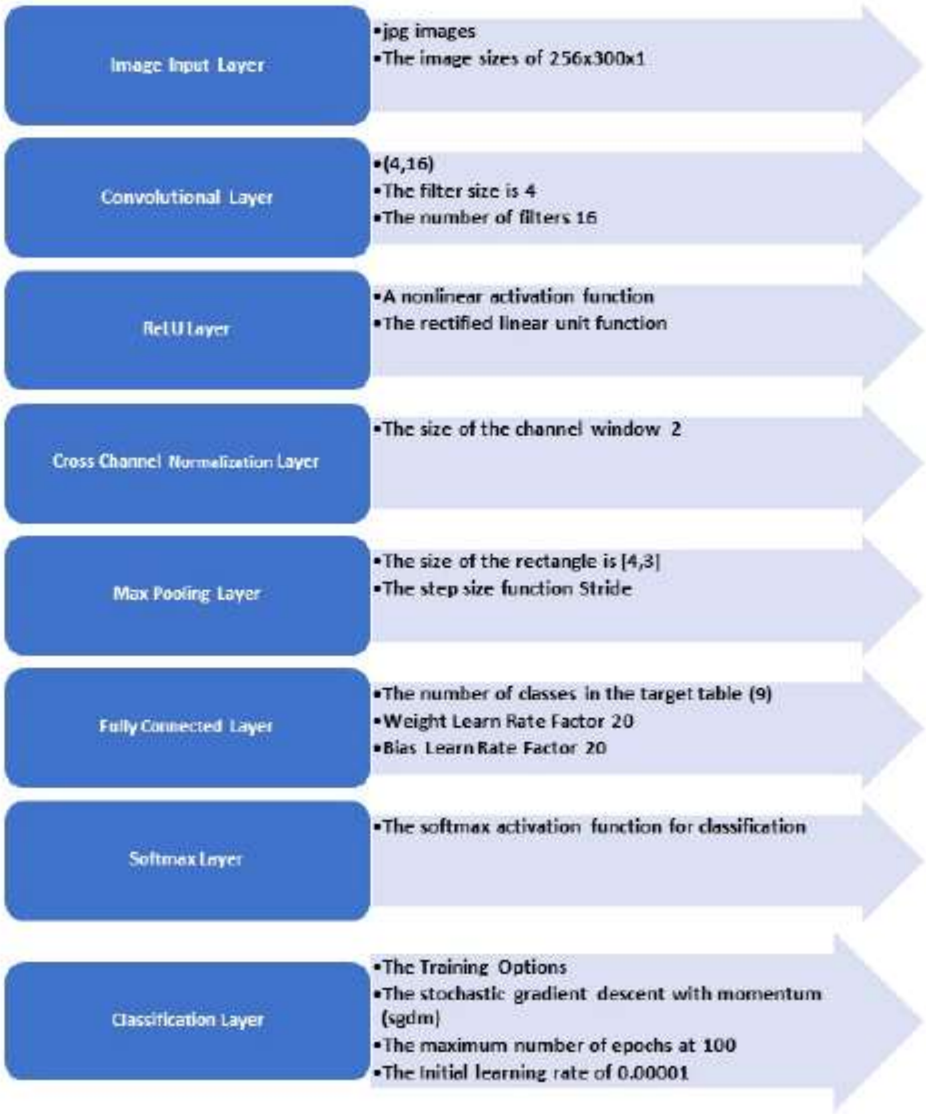

Fig. 3. The structural details of the CNN.

The application of the CNN in hand gesture recognition was rigorously tested using the Cambridge Hand Gesture dataset. Notably, 80% of the dataset was allocated as training data, while the remaining 20% served as test data. It is imperative to underscore that the default maximum epoch in the CNN architecture was set at 1000. This deliberate choice is crucial in ensuring a thorough evaluation of the network's learning capabilities and generalization performance over a significant number of training iterations.

RESULT AND ANALYSIS

Here we have conducted a comprehensive examination of the experimental outcomes, emphasizing scientific interpretation to discern the nuances of individual results. The primary performance measures—accuracy, recall, precision, and F1 score—serve as crucial benchmarks for evaluating the proposed work. The analysis unfolds through a meticulous exploration of various network architectures and comparisons with existing hand gesture recognition works.

Table 1. Performance measures of each layer in CNN.

Measures	CNN 1 (%)	CNN 2 (%)	CNN 3 (%)	CNN 4 (%)
Accuracy	96.00	97.05	97.23	97.66
Recall	96.51	97.21	97.65	97.86
Precision	96.45	97.12	97.87	97.45
F1 score	96.35	97.43	97.89	97.25

The proposed work, as evidenced by Table 1, achieves commendable results with an average success rate of 97%. Notably, CNN1, featuring only two layers of convolution, exhibits an accuracy of 96%, while CNN 2, 3, and 4 consistently maintain accuracy levels above 97%. This indicates the robustness of the proposed work across different layers, with a steady improvement in accuracy as the number of convolutional layers increases.

Table 2 further elucidates the comparative performance measures across different network architectures, including Inception ResNetV2, ResNet 50, Dense Net201, and LeNet. Notably, Inception ResNetV2 outperforms other architectures with an accuracy of 98.95%. The recall, precision, and F1 score metrics also showcase the superiority of Inception ResNetV2 in capturing the intricacies of hand gesture recognition.

An insightful analysis is presented in Table 3, where the performance measures of each layer in CNN are juxtaposed with the Cambridge Hand Gesture dataset and benchmarked against works by Barczak et al. and Moeslund's. The results showcase the adaptability and consistency of the proposed work across different datasets, affirming its robustness and generalization capabilities.

Table 2. Comparison of the performance measures with different network architectures.

Measures	Inception ResNet V2 (%)	ResNet 50 (%)	Dense Net201 (%)	LeNet (%)
Accuracy	98.95	96.99	97.54	92.54
Recall	98.56	96.56	97.77	92.61
Precision	98.76	96.87	97.94	92.53
F1 Score	98.54	96.77	97.63	92.76

Table 3. Performance measures of each layer in CNN.

	CNN 1 (%)	CNN 2 (%)	CNN 3 (%)	CNN 4 (%)
Cambridge Hand gesture data	96.73	96.47	97.23	97.56
Barczak *et al.*[26]	97.65	97.89	97.55	98.96
Moeslund's[27]	98.54	98.66	98.78	98.65

Table 4. Performance measures with different network architectures with respect to hand gesture data.

	Inception ResNetV2 (%)	ResNet 50 (%)	Dense Net201 (%)	LeNet (%)
Cambridge Hand gesture data	98.45	97.69	98.64	88.56
Barczak *et al.*[26]	98.66	97.86	97.57	98.61
Moeslund's[27]	98.46	97.67	98.54	97.53

Table 4 further extends the comparative analysis, evaluating the proposed work against various network architectures concerning different hand gesture datasets. In this context, Inception ResNetV2 consistently emerges as a top-performing architecture, showcasing its versatility and efficacy in diverse scenarios.

To contextualize the findings, Table 5 provides a comprehensive comparison with previous hand gesture recognition works. The proposed CNN achieves an overall accuracy of 96.66%, showcasing competitive performance in comparison to alternative methods. Notably, it exhibits higher sensitivity and specificity than several existing methods, underlining its potential for real-world applications.

The scientific interpretation of individual results is paramount in discerning the strengths and limitations of the proposed work. The consistent improvement in accuracy with the addition of convolutional layers suggests that the depth of the network contributes significantly to enhanced performance. Furthermore, the benchmarking against different network architectures and datasets highlights the adaptability and generalization capabilities of the proposed CNN.

Conclusion:

In conclusion, this study has delved into the realm of hand gesture recognition systems, specifically exploring the application of Convolutional Neural Networks (CNNs). The results obtained from the experimental work underscore the efficacy of the proposed CNN architecture, with an impressive average success rate of 97%. Notably, the incremental increase in accuracy with additional convolutional layers signifies the importance of depth in enhancing the performance of hand gesture recognition systems.

The comparative analysis with various network architectures, such as Inception ResNetV2, ResNet 50, Dense Net201, and LeNet, further solidifies the standing of the proposed work. Inception ResNetV2 emerges as a top-performing architecture, showcasing its adaptability and superior performance in capturing the intricacies of hand gestures. The study positions the proposed CNN as a robust solution, offering competitive accuracy and outperforming several existing methods in terms of sensitivity and specificity.

Limitation of the Study:

While the study demonstrates promising outcomes, it is crucial to acknowledge its limitations. One notable limitation lies in the reliance on specific datasets, such as the Cambridge Hand Gesture dataset. The generalizability of the proposed CNN to diverse datasets and real-world scenarios warrants further exploration. Additionally, the study primarily focuses on static hand gestures, and future research could extend the scope to dynamic gestures for a more comprehensive understanding.

Another limitation pertains to the architectural choices made in this study. While the proposed CNN configuration yields favorable results, the exploration of alternative architectures and hyperparameter tuning could provide insights into further optimization possibilities. The study also assumes static hand gestures without incorporating variations in lighting conditions, which may impact the model's performance in practical, dynamic environments.

Implication of the Study:

The implications of this study extend beyond the confines of academic research, offering practical applications in diverse domains. The robust performance of the proposed CNN architecture in hand gesture recognition has potential implications for human-computer interaction, sign language recognition, and assistive technologies. The ability to accurately interpret hand gestures opens avenues for intuitive and natural communication between humans and machines, fostering advancements in user interface design and accessibility.

Moreover, the study's findings can inform the development of gesture-based interfaces in virtual and augmented reality environments, gaming, and smart home applications. The reliable recognition of hand gestures can enhance user experience and interaction in these contexts, contributing to the evolution of technology in various industries.

Future Recommendations:

To advance the field of hand gesture recognition, future research endeavors should address the identified limitations and explore new frontiers. Firstly, investigations into the generalizability of the proposed CNN to diverse datasets, including dynamic gestures and varying environmental conditions, will enhance the model's applicability in real-world scenarios.

Exploring alternative architectures and conducting thorough hyperparameter tuning could uncover optimization strategies to further boost performance. Future studies may also delve into the interpretability of the CNN, elucidating the features driving its decisions, and facilitating a deeper understanding of the recognition process.

Additionally, the integration of real-time capabilities and considerations for edge computing could be pivotal for deploying hand gesture recognition systems in resource-constrained environments. Collaboration with experts in fields such as human-computer interaction, robotics, and healthcare can provide valuable insights and foster interdisciplinary advancements.

REFERENCES

1. H. Francke, J. Ruiz-del-Solar, and R. Verschae, "Real-time hand gesture detection and recognition using boosted classifiers and active learning," in Pacific-Rim Symp. Image and Video Technology (Springer, Berlin, Heidelberg, 2007), pp. 533–547.
2. E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," Eng. Appl. Artif. Intell. 22(8), 1141–1158 (2009).
3. P. Barros, N. T. Maciel-Junior, B. J. Fernandes, B. L. Bezerra, and S. M. Fernandes, "A dynamic gesture recognition and prediction system using the convexity approach," Comput. Vis. Image Underst. 155, 139–149 (2017).
4. P. K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," Comput. Vision Image Underst. 141, 152–165 (2015).
5. K. Martin Sagayam and D. Jude Hemanth, "Hand posture and gesture recognition techniques for virtual reality applications: A survey," Virtual Reality 21(2), 91–107 (2017).
6. C. Wang, Z. Liu, M. Zhu, J. Zhao, and S. C. Chan, "A hand gesture recognition system based on canonical superpixel-graph," Signal Process.: Image Commun. 58, 87–98 (2017).
7. A. Hariaa, A. Subramaniana, N. Asokkumara, S. Poddar, and J. S. Nayak, "Hand gesture recognition for human-computer interaction," Procedia Comput. Sci. 115, 367–374 (2017).
8. N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," IEEE Trans. Instrum. Meas. 60(11), 3592–3607 (2011).

9. W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," Neural Comput. 29(9), 2352–2449 (2017).

10. S. Shin and W. Sung, "Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks," 2016 IEEE Int. Symp. Circuits and Systems (ISCAS) (IEEE, 2016), pp. 2274–2277.

11. A. Dadashzadeh, A. T. Targhi, and M. Tahmasbi, "HGR-net: A two-stage convolutional neural network for hand gesture segmentation and recognition," arXiv:1806.05653.

12. P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (IEEE, 2015), pp. 1–7.

13. V. John, A. Boyali, S. Mita, M. Imanishi, and N. Sanma, "Deep learning-based fast hand gesture recognition using representative frames," in 2016 Int. Conf. Digital Image Computing: Techniques and Applications (DICTA) (IEEE, 2016), pp. 1–8.

14. A. Tang, K. Lu, Y. Wang, J. Huang, and H. Li, "A real-time hand posture recognition system using deep neural networks," ACM Trans. Intell Syst. and Technol. (TIST), 6(2), 21 (2015).

15. A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," IEEE Access 6, 13949–13957 (13957).

16. S. Swarup Rautaray and A. Agrawal, "Hand gesture recognition towards vocabulary and application independency," Int. J. Image Graphics 13(2), 1340001 (2013), https://doi.org/10.1142/S0219467813400019.

17. O. Deperlioglu, "Classification of phonocardiograms with convolutional neural networks," Broad Res. Artif. Intell. Neurosci.: 9(2), 22–33 (2018).

18. O. Deperlioglu, "Classification of segmented phonocardiograms by convolutional neural

networks," Broad Res. Artif. Intell. Neurosci. 10(1), 5–12 (2019).

19. T. K. Kim, S. F. Wong, and R. Cipolla, "Tensor canonical correlation analysis for action classification," in CVPR'07 IEEE Conf. Computer Vision and Pattern Recognition, IEEE, 2007. pp. 1–8.

20. A. Gharehbaghia, M. Borgaa, B. J. Sj€obergc, and P. Ask, "A novel method for discrimination between innocent and pathological heart murmurs," Med. Eng. Phys. 37, 674–682 (2017).

21. H. Kahramanli and N. Allahverdi, "Design of a hybrid system for diabetes and heart diseases," Expert Syst. Appl. 35, 82–89 (2008).

22. K. Martin Sagayam and D. Jude Hemanth, "ABC algorithm-based optimization of 1-D hidden Markov model for hand gesture recognition applications," Comput. Industry 99, 313–323 (2018).

23. W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," Neural comput. 29(9), 2352–2449 (2017).

24. T. Aora and R. Dhir, "Geometric feature-based classification of segmented human chromosomes," Int. J. Image Graphics, 20(1), 2050006, (2020), https://doi.org/10.1142/S0219467820500060.

25. A. L. C. Barczak, N. H. Reyes, M. Abastillas, A. Piccio, and T. Susnjak, "A new 2D static hand gesture color image dataset for ASL gestures," Res. Lett. Inf. Math. Sci. 15(4356), 12–20 (2011).