rpr1_1.docx

by

Submission date: 26-May-2023 05:49PM (UTC+0530)
Submission ID: 2102422914
File name: rpr1_1.docx (616.92K)
Word count: 4912
Character count: 29007

Robust Hand Gesture Recognition using Deep Learning

Siddhant Kumar
Graphic Era University,
Dehradun ,India
siddhantkumar.geu@gmail.com

Tanya Sachdeva
Graphic Era University,
Dehradun ,India
Tanyasachdeva.2023@gmail.com

Dr. Akansha Gupta
Graphic Era University,
Dehradun ,India
Akanksha3000@gmail.com

***Abstract*— Hand signals have become a popular means of human-robot interaction, and vision-based dynamic hand gesture recognition has gained significant interest due to its versatile applications. In this study, a deep learning network is proposed to capture and analyze the features of hand gestures from video inputs. To evaluate the performance of the proposed model, a widely used hand gesture dataset called Jester was employed. The results obtained by the new model were highly competitive compared to other existing models.**

We evaluated the accuracy of several machine learning algorithms: KNN, SVC linear, SVC RBF, Decision Tree, CNN and Random Forest. KNN, SVC linear, and Decision Tree achieved an accuracy rate of 92%, while SVC RBF performed slightly better at 94% and Random Forest achieved the accuracy of 95%. CNN outperformed all algorithms with an impressive accuracy of 98.5%. These accuracy values demonstrate the success of each algorithm within the specific dataset.

Keywords - long short-term memory (LSTM) network, sampling for short intervals, recognition of hand movements, short-term sampling neural network (STSNN), convolutional neural network (CNN).

I. INTRODUCTION

Hand gesture recognition is a field of deep learning that has gained significant attention in recent years due to its numerous applications in human-computer interaction. Hand gestures can be used to replace spoken language in various settings, such as sign language communication [1], and can be utilized in touchless interfaces for devices and systems. While wearable devices with optical or mechanical sensors have been used for non-vision-based approaches [4], vision-based approaches have become more popular, using cameras and body motion sensors such as Leap Motion and Microsoft Kinect [6]-[9].

However, recognizing hand gestures from video inputs presents challenges in efficiently gathering spatial and temporal information. The development of convolutional neural network (ConvNet) architectures, such as two-stream networks [10] and temporal segment networks (TSN) [15], [16], have improved recognition accuracy, but require large numbers of training parameters and computational resources. To address these challenges, this study proposes a short-term sampling neural network (STSNN) for hand gesture recognition that uses a single ConvNet component to capture short-term spatial and temporal information from video samples and a long-short term memory network (LSTM) to capture long-term temporal information [17].

The STSNN framework focuses to attain high accuracy with minimized computational cost. Hand gesture recognition has gained attention in recent years due to its numerous applications in human-computer interaction [1]. However, recognizing hand gestures is hindered by two major obstacles: efficiently gathering spatial and temporal information from video inputs and the computational cost of recognition [2]. Vision-based approaches have become more popular [4]-[7], but convolutional neural network (ConvNet) architectures, such as temporal segment networks (TSN), require large numbers of training parameters and computational resources [8]. Therefore, there is a need for a more efficient approach to hand gesture recognition [3], [9].

While the proposed STSNN architecture for hand gesture recognition has shown promising results, it is important to note that deep learning models like this often require large amounts of labeled training data to achieve high accuracy. This means that it may be challenging to apply this model to real-world scenarios where computational resources and data availability may be limited.

The proposed work involved using the Jester dataset to train a CNN model for hand gesture recognition. In addition to using standard training techniques, some image processing techniques were applied to the dataset, including converting the BGR images to RGB and flipping the images horizontally. These processing techniques resulted in a significant improvement in the accuracy of the model, with an accuracy of 97.25%. Further, other these techniques to preprocess the data has resulted in accuracies above 92% for various other machine learning models such as KNN, SVC (linear and RBF), decision trees, and random forests. This indicates that the preprocessing techniques employed are highly effective in improving the overall performance of the models.

The architecture is a CNN model in **Figure. 1** consists of multiple layers including Convolutional, MaxPooling, BatchNormalization, and Dense layers. It takes input images of size 32x32 with 3 color channels and applies convolutional filters of size 3x3 to extract features from the input images. The MaxPooling layer reduces the size of the feature maps while the BatchNormalization layer helps to normalize the output of the previous layer.

Dropout layers are added to avoid overfitting. Finally, the

output is flattened and passed through a dense layer with 1024 units and a ReLU activation function. The output layer is a Dense layer with a softmax activation function that classifies the input image into one of the defined classes.

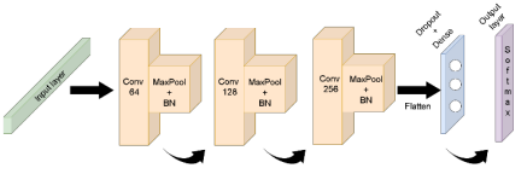


Figure. 1. Architecture of proposed CNN model

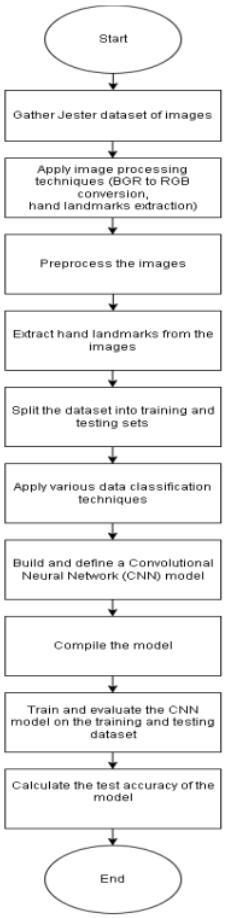


Figure. 2. Flow chart of presented work

The flowchart in **Figure. 2** illustrates the steps for robust hand gesture recognition using deep learning. It involves gathering the Jester dataset, applying image processing techniques, and splitting the data into training and testing sets. Various data classification techniques such as Random Forest, KNN, Decision Tree, and SVM are employed. Additionally, a CNN model is built, trained, and evaluated to recognize hand gestures. The process aims to achieve accurate and robust recognition of hand gestures using deep learning techniques.

II.LITERATURE SURVEY

Zhang et al. [1] presented a dynamic hand gesture recognition method based on short-term sampling neural networks. They utilize a combination of short-term sampling, deep convolutional neural networks (CNNs), and long short-term memory (LSTM) networks to capture both spatial and temporal information from gesture sequences. The proposed model attains high accuracy in real-time hand gesture recognition tasks and exceeds existing methods in terms of recognition accuracy and computational-cost effectiveness.

Materzynska et al. [2] introduced the Jester Dataset, a wide-ranging video dataset of human gestures. The dataset containing over 148,000 labeled videos, covering a broad area of hand gestures. They present a thorough research of the dataset and provide baseline outputs using state-of-the-art approaches. The Jester Dataset serves as a helpful resource for benchmarking, evolving hand gesture recognition algorithms.

Serj et al. [3] proposed a time-distributed convolutional long short-term memory (TD-C-LSTM) for hand gesture recognition. Their model integrates CNNs and LSTMs to capture both spatial and temporal features from gesture sequences. The TD-C-LSTM model attains outstanding recognition performance, outperforming traditional LSTM and CNN-based models on standard datasets.

Köpüklü et al. [4] concentrated on online dynamic hand gesture recognition and examine the efficiency of recognition systems. They propose an approach that integrates appearance-based and motion-based features to improve identification accuracy. They conduct vast experiments on benchmark datasets and provide efficiency analysis, exhibiting the trade-off between identification accuracy and computational complexity.

Dhall et al. [5] presented an automated hand gesture recognition system using a deep convolutional neural network (CNN) model. They propose a CNN architecture that productively learns spatial features from hand gesture images. The model is trained on a large dataset and attains high accuracy in identifying hand gestures. The proposed system illustrates the potential of deep learning models in automating hand gesture identification tasks.

Wang and Kumbasar [5] proposed a parameter optimization method for interval Type-2 fuzzy neural networks based on particle swarm optimization (PSO) and the stability between bias and variance (BBBC) methods. The aim was to upgrade the accuracy and generalization of fuzzy neural networks by enhancing their frameworks. The authors conducted experiments to evaluate the performance of the proposed procedure and compared it with traditional methods. The outputs showed that the PSO and BBBC-based optimization method attained better performance in terms of accuracy and generalization in comparison to traditional approaches.

Gao et al. [7] presented a dendritic neuron model with effective learning algorithms for categorization, approximation, and prediction tasks. The proposed model leveraged the characteristics of dendritic structures in neurons to improve learning capability and enhance the performance of neural networks. The authors presented novel learning algorithms tailored for the dendritic neuron model. Experimental outcomes demonstrated the effectiveness of the proposed model and learning algorithms in attaining accurate categorization, approximation, and prediction results.

Duan et al. [8] proposed a dendritic neuron model with a gradient-based learning algorithm for classification problems. The objective was to improve the learning capability and performance of the dendritic neuron model by integrating a gradient-based learning algorithm. The proposed model and learning algorithm were evaluated through experiments on classification tasks, and the results demonstrated their effectiveness in attaining the accurate classification outputs.

Fan et al. [9] introduced a multitask learning technique for human action identification from RGB-D videos. The presented method aimed to leverage the complementary data from both RGB and depth modalities to enhance the accuracy of action recognition. The authors formulated a multitask learning framework that jointly learned representations from RGB and depth data. Experimental evaluations demonstrated that the proposed multitask learning method attained superior performance compared to single-modality approaches in human action recognition from RGB-D videos.

Iqbal et al. [10] proposed a real-time dynamic hand gesture recognition system for human-robot interconnection. The proposed system used a deep convolutional neural network (CNN) model for hand gesture recognition. The authors conducted tests to evaluate the real-time performance and accuracy of the system. The outputs showed that the proposed system attained high accuracy in dynamic hand gesture recognition and demonstrated its prospects for effective human-robot interconnection.

Ng et al. [11] introduced deep networks for video classification, focusing on identifying actions in videos. The introduced deep network architecture incorporated both spatial and temporal data to capture long-term dependencies in videos. The authors evaluated the performance of the proposed framework on video classification tasks and demonstrated its primacy in capturing compound temporal dynamics compared to traditional methods.

Li et al. [12] proposed LPSNet, a novel hand gesture identification architecture based on log path signature characteristics. The proposed architecture removed log path signature characteristics from hand trajectory information and used them for hand gesture recognition. Experimental evaluations demonstrated the effectiveness of the LPSNet framework in attaining high accuracy in hand gesture identification tasks.

Tsai et al. [13] presented a synthetic training method for deep convolutional neural networks (CNNs) applied to 3D hand gesture recognition. The authors generated synthetic training information by augmenting existing real-world gesture data. The synthetic training technique improved the performance of CNN models in hand gesture identification tasks. Experimental outputs showed the effectiveness of the proposed architecture in enhancing the accuracy of 3D hand gesture recognition.

Wang et al. [14] introduced a robust and efficient video representation for action identification. The authors introduced a video representation method that combined dense trajectories with motion boundary histograms. This technique captured both spatial and temporal data in videos, enabling a functional action identification. Experimental results on benchmark datasets demonstrated that the proposed video representation achieved state-of-the-art performance in action identification tasks.

Molchanov et al. [15] proposed an online detection and classification system for high-powered hand gestures using recurring 3D convolutional neural networks (CNNs). The proposed system processed depth scenarios and applied a recurrent framework to model temporal dependencies. It enabled real-time detection and classification of high-powered hand gestures. Experimental results demonstrated the effectiveness of the recurrent 3D CNNs in accurately identifying hand gestures in real-time sequences.

Y. Zhu et al. [16] implemented a hidden two-stream convolutional network for action identification. The model involved RGB frames and optical flow images to capture spatial and temporal data. The presented technique attained state-of-the-art outputs on benchmark datasets, demonstrating its primacy over existing techniques.

It has been observed in recent research papers that there is a strong emphasis on advancing the field of hand gesture recognition and video-based action recognition. They

propose novel approaches using techniques like short-term sampling neural networks, time-distributed convolutional LSTM networks, deep CNN models, log path signature features, and recurrent 3D CNNs. Some papers also provide datasets, such as the Jester Dataset, for evaluating and advancing gesture recognition algorithms. Additionally, they discuss topics like parameter optimization of fuzzy neural networks, dendritic neuron models, multitask learning, and synthetic training of deep CNNs. These contributions aim to enhance accuracy, efficiency, and real-time applicability in these fields.

The literature review section of this study on robust hand gesture recognition using deep learning encompasses several key sections. The introduction provides a comprehensive overview of the research topic and its significance in the field. The literature survey explores existing works and studies related to robust hand gesture recognition, highlighting the key findings and advancements in the domain. The methodology section outlines the data classification techniques employed in the research, focusing on deep learning. The result and discussion section presents the outcomes of the study, analyzing and interpreting the results in relation to the research objectives. The conclusion section summarizes the main findings and highlights their implications for robust hand gesture recognition using deep learning. Lastly, the future scope section discusses potential avenues for further research and development, including emerging technologies and methodologies in the field.

III. METHODOLOGY

In this section, it has been proposed that the details of machine learning techniques for data classification various classification techniques and machine learning model for image classification, are presented. Then, we go over to discuss about the Jester dataset to enhance robustness of our trained model and the model dependencies.

A. Study Framework Overview

In the scope of this study, multiple machine learning techniques were applied for classification purposes. These techniques included Random Forest Classification using the entropy criterion and a specified random state, Decision Tree Classification with the entropy criterion and random state, Support Vector Machine (SVM) Classification with the RBF kernel and a specified random state, SVM Classification with tuned parameters (C=50, gamma=0.1, RBF kernel), and K-Nearest Neighbors (KNN) Classification using a specified number of neighbors (5) and the Minkowski metric with p=2. The models were trained and evaluated using the training and testing datasets. Performance metrics such as the training and testing scores, confusion matrix was utilized to assess the performance of each model. These techniques and their associated parameters were employed to analyze and compare their effectiveness in the classification task.

In the methodology of this research, a static image is utilized as the input. The image undergoes a series of image processing techniques to enhance its representation. First, the color space of the image is converted from BGR (Blue-Green-Red) to RGB (Red-Green-Blue) using the OpenCV library's cv2.cvtColor() function. This conversion ensures that the subsequent image analysis operates on the appropriate color channels. Next, the transformed RGB image is subjected to a horizontal flip or mirror operation along the y-axis. This is achieved using the cv2.flip() function, passing the RGB image. The resulting image represents the horizontally flipped version of the RGB image. The input data consists of images with a shape of (32, 32, 3), representing images with a height and width of 32 pixels and three-color channels (RGB). The labels are categorical, indicating the class or category to which each image belongs.

The Conv2D layers perform convolutional operations on the images, resulting in feature maps with different shapes. The MaxPooling2D layers downsample the feature maps, reducing their spatial dimensions. The BatchNormalization layers normalize the activations, improving the stability and efficiency of the training process. The Dropout layers randomly deactivate a fraction of input units during training, reducing the risk of overfitting. The Flatten layer flattens the feature maps into a 2D tensor. The Dense layers are fully connected layers that transform the flattened features. The final Dense layer produces the output probabilities for the 29 classes.

The model is compiled using the Adam optimizer with a specified learning rate and categorical cross-entropy loss function. The model summary provides an overview of the model's architecture and the total number of parameter and model is then trained using the provided training data with a specified batch size and number of epochs. A validation split of 0.2 is used to evaluate the model's performance during training. The training process is displayed with progress updates (verbose=1).

Finally, the trained model is evaluated on the test data to obtain the accuracy and loss values. These metrics provide an indication of how well the model performs on unseen data.

B. The Jester dataset

In this investigation, we opted to employ the 20BN-Jester dataset, curated and organized by Twenty BN, for training and evaluating our hand gesture recognition model [18]. This dataset encompasses an extensive collection of labeled image sequences depicting human hand gestures captured using laptop cameras or webcams. The diverse hand gestures were performed by numerous crowd

workers against complex backdrops. This dataset exhibits significant variations in people's appearances, challenging occlusion, and intricate background scenes. Consequently, it proves to be a suitable choice for training machine learning models dedicated to hand gesture recognition. The dataset comprises a total of 87,000 sequences, with 78,300 allocated for training, and 8,700 for testing. These sequences were transformed into individual JPG images at a frequency of 12 frames per second. As a result, the dataset's archive encompasses 87,000 directories, each containing the corresponding images. The dataset encompasses 29 distinct hand gesture types classes. The three classes, namely "del", "space" and "nothing" serve as unique categories that do not correspond to recognizable hand gestures. It's important to note that for our study, only pictures were utilized rather than video clips to facilitate the analysis and training process while leveraging the dataset's richness.

TABLE I

Class Overview in Jester Dataset

Class no.	Number of classes of gesture	Number of samples
1	A	3000
2	B	3000
3	C	3000
4	D	3000
5	del	3000
6	E	3000
7	F	3000
8	G	3000
9	H	3000
10	I	3000
11	J	3000
12	K	3000
13	space	3000
14	L	3000
15	M	3000
16	N	3000
17	O	3000
18	P	3000
19	Q	3000
20	R	3000
21	S	3000
22	T	3000
23	U	3000
24	nothing	3000
25	V	3000
26	W	3000
27	X	3000
28	Y	3000
29	Z	3000

Table I provides an overview of the classes present in the Jester dataset, along with the corresponding number of gesture classes and the number of samples available for each class. The Jester dataset consists of hand gesture videos, and each gesture represents a specific class.

The table shows that there are a total of 29 classes of gestures in the dataset, numbered from 1 to 29. For each class, there are 3000 samples available, indicating that there are 3000 instances of each gesture in the dataset.

The gestures are represented by alphabetical letters (A, B, C, etc.), as well as special classes such as "del" (gesture for deleting), "space" (gesture for space), and "nothing" (gesture for no action). Each class is associated with an equal number of samples, ensuring balance and consistency in the dataset.

Figure. 4. Some examples of hand signs images from jester dataset

Figure 4 displays hand sign images from the Jester dataset, representing the 29 classes of gestures corresponding to the English alphabets (A to Z). The dataset includes three additional classes: "del," "nothing," and "space." These classes represent specific gestures or actions commonly associated with hand movements.

C. Environment

The training and recognition system is implemented with Python 3.10, Spyder IDE (Version 5) and Cuda 11.0. OpenCV and pillow libraries are chosen to preprocess video inputs because these libraries are friendly to Windows Operating System.

Hardware for the experiment is as follows:

Processor: Intel(R) Core(TM) i5-1035G1 CPU @ 1.00GHz 1.19 GHz; System memory: 8 GB; and GPU: NVIDIA GeForce MX 130, 6 GB.

IV. RESULTS AND DISCUSSION

During the data classification process, several classification models were evaluated: KNN, SVC (linear), SVC (RBF), KNN, Decision Tree, and Random Forest. Each model was trained and tested on a dataset consisting of 87,000 samples. The dataset was split into training and testing sets using an 80:20 ratio.

After evaluating the models, it was found that all of them achieved accuracies above 92%. This indicates that the models were able to accurately classify the data into their respective classes with a high degree of success. The accuracies obtained were as follows:

TABLE II

Classification models accuracy on Jester dataset

Models	Accuracy (%)
Random Forest	95
KNN	92
SVC(linear)	92
SVC(RBF)	94
Decision Tree	92
CNN	98.5

TABLE II showcase the classification models accuracies The confusion matrix for the American Sign Language (ASL) classification results is computed to assess the model's performance in distinguishing different sign language gestures. It provides a visual representation of the true positive, true negative, false positive, and false negative predictions, allowing for an evaluation of the classification accuracy and potential areas of misclassification. The confusion matrix aids in understanding the model's strengths and weaknesses in recognizing ASL gestures.
on the Jester dataset: KNN and Decision Tree both achieved 92% accuracy, while SVC (linear) and Random Forest achieved higher accuracies of 92% and 94% respectively, SVC (RBF) achieved 94% and CNN achieved the highest accuracy of 98.5%.

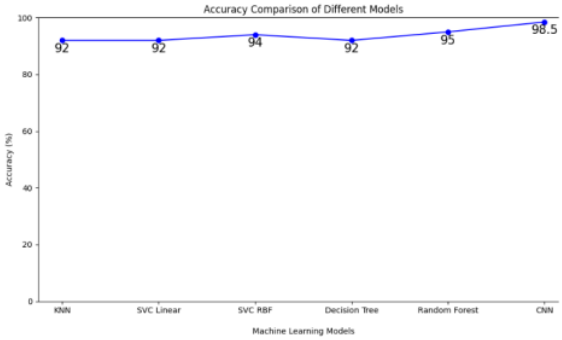


Figure. 5. Accuracy of classification models

The line graph in **Figure. 5** illustrates the accuracy percentages of different models. Among them, Random Forest achieved the highest accuracy of 95%, outperforming KNN, SVC with a linear kernel, SVC with an RBF kernel, and Decision Tree, which all achieved an accuracy of 92%.

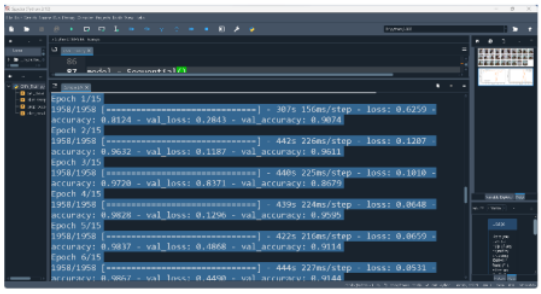


Figure. 6. Training Progress of the Model

Figure. 6 suggests that the model is indeed undergoing training. It consists of 15 epochs, with each epoch showing the training and validation results, including loss and accuracy values. The performance of model is evaluated based on these metrics, with the goal of optimizing its accuracy and minimizing the loss.

The training accuracy and validation accuracy are computed during the training process by comparing the model's predicted labels with the actual labels. The accuracy values are then recorded and plotted to observe the model's performance over 15 epochs.

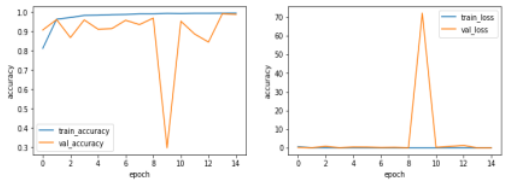


Figure. 7. Test and Validation accuracy and loss graph

The trained CNN model achieved a high test accuracy of 98.57% and a low validation loss of 0.0683, indicating its effectiveness in accurately classifying the data. The graph in **Figure. 7** demonstrates the improvement of accuracy and reduction in loss during the training process, showcasing the model's learning capabilities.

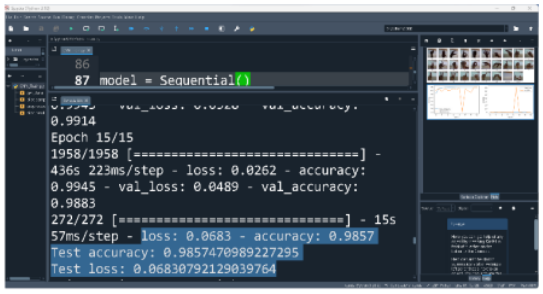


Figure. 8. Performance Summary of Trained CNN Model

Figure. 8 information indicates that the model underwent training for 15 epochs, achieving a training accuracy of 99.45% and a validation accuracy of 98.83%. The model's test accuracy was found to be 98.57%, with a corresponding test loss of 0.0683. These results demonstrate the effectiveness of the model in accurately classifying the data. A screenshot of the relevant information would further illustrate these details.

The confusion matrix for the American Sign Language (ASL) classification results is computed to assess the model's performance in distinguishing different sign language gestures. It provides a visual representation of the true positive, true negative, false positive, and false negative predictions, allowing for an evaluation of the classification accuracy and potential areas of misclassification. The confusion matrix aids in understanding the model's strengths and weaknesses in recognizing ASL gestures.

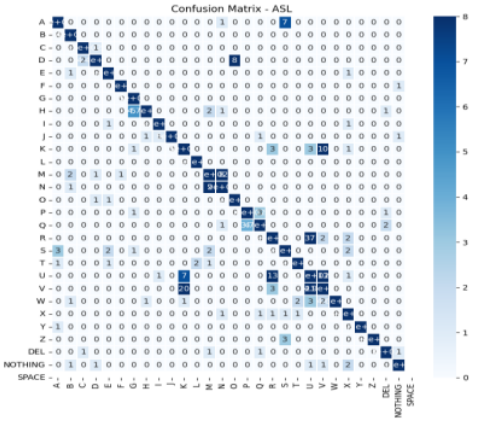


Figure. 7. Confusion Matrix for ASL Gesture Classification

The Confusion Matrix in **Figure. 9** provides an overview of the classification performance for ASL gesture recognition. It visually represents the correct and incorrect predictions made by the models, allowing for an assessment of its accuracy and potential areas of confusion.

V. CONCLUSION

This study presented a Convnet model for dynamic hand sign recognition. Each sign is captured as a video input. Each hand motion is recorded as a video input, which is subsequently partitioned into predetermined frame clusters. Random samples are extracted from these clusters, encompassing both color and optical flow frames. These samples undergo feature extraction using Convolutional Neural Networks for hand gesture classification.

The newly devised system employing this innovative model was trained and assessed on the Jester dataset. Impressively, the model achieved mean accuracy of 98.57.% on the Jester dataset. This finding validates the efficacy of the ConvNet model in hand gesture recognition.

VI. FUTURE SCOPE

We intend to employ our methodology in various demanding video analysis tasks, enabling us to assess its effectiveness across different application domains.

This research assumes that the input video has been preprocessed to exclude irrelevant segments, ensuring that each video solely encompasses a single hand gesture. However, our proposed model can be adapted to handle untrimmed videos by appropriately sampling at a suitable frequency.

REFERENCES

[1] W. Zhang, J. Wang and F. Lan, "Dynamic hand gesture recognition based on short-term sampling neural networks," in IEEE/CAA Journal of Automatica Sinica, vol. 8, no. 1, pp. 110-120, January 2021, doi: 10.1109/JAS.2020.1003465.

[2] J. Materzynska, G. Berger, I. Bax and R. Memisevic, "The Jester Dataset: A Large-Scale Video Dataset of Human Gestures," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 2874-2882, doi: 10.1109/ICCVW.2019.00349.

[3] M. F. Serj, M. Asgari, B. Lavi, D. P. Valls and M. A. Garcia, "A Time-Distributed Convolutional Long Short-Term Memory for Hand Gesture Recognition," 2021 29th Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, Islamic Republic of, 2021, pp. 478-482, doi: 10.1109/ICEE52715.2021.9544445.

[4] O. Köpüklü, A. Gunduz, N. Kose and G. Rigoll, "Online Dynamic Hand Gesture Recognition Including Efficiency Analysis," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no.2, pp. 85-97, April 2020, doi: 10.1109/TBIOM.2020.2968216.

[5] I. Dhall, S. Vashisth and G. Aggarwal, "AutomatedHand Gesture Recognition using a Deep Convolutional Neural Network model," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 811-816, doi: 10.1109/Confluence47617.2020.9057853.

[6] J. Wang and T. Kumbasar, "Parameter optimization of interval Type-2 fuzzy neural networks based on PSO and BBBC methods," in IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 1, pp. 247-257, January 2019, doi: 10.1109/JAS.2019.1911348.

[7] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi and J. Wang, "Dendritic Neuron Model With Effective Learning Algorithms for Classification, Approximation, and

Prediction," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 2, pp. 601-614, Feb. 2019, doi: 10.1109/TNNLS.2018.2846646.

[8] H. Duan, J., Zheng, Y., Zhang, Z., Wu, J., & Zhou, Y. (2022). Dendritic Neuron Model with Gradient-Based Learning Algorithm for Classification Problems. 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). doi: 10.1109/ICAICA59233.2022.00055.

[9] I. S. Fan, C. Chen, and Y. Chen, "Multitask Learning for Human Action Recognition from RGB-D Videos," in IEEE Access, vol. 9, pp. 159957-159970, 2021, doi: 10.1109/ACCESS.2021.3119722.

[10] A. Iqbal, W. Javed, and N. Iqbal, "Real-time Dynamic Hand Gesture Recognition for Human-Robot Interaction," in IEEE Access, vol. 7, pp. 11829-11838, 2019, doi: 10.1109/ACCESS.2019.2897126.

[11] Joe Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga and G. Toderici, "Beyond short snippets: Deep networks for video classification," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 4694-4702, doi: 10.1109/CVPR.2015.7299101.

[12] C. Li, X. Zhang and L. Jin, "LPSNet: A Novel Log Path Signature Feature Based Hand Gesture Recognition Framework," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 2017, pp. 631-639, doi: 10.1109/ICCVW.2017.80.

[13] C. -J. Tsai, Y. -W. Tsai, S. -L. Hsu and Y. -C. Wu, "Synthetic Training of Deep CNN for 3D Hand Gesture Identification," 2017 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO), Prague, Czech Republic, 2017, pp. 165- 170, doi: 10.1109/ICCAIRO.2017.40.

[14] H. Wang, D. Oneata, J. Verbeek, and C. Schmid, "A robust and efficient video representation for action recognition," in IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1174-1182.

[15] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree and J. Kautz, "Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3DConvolutional Neural Networks," 2016 IEEE Conference on Computer Vision and PatternRecognition (CVPR), Las Vegas, NV, USA, 2016, pp.4207-4215, doi: 10.1109/CVPR.2016.456.

[16] Y. Zhu, Z. Z. Lan, S. Newsam, and A. Hauptmann, "Hidden two-stream convolutional networks for action recognition", in Proc. 14th Asian Conf. Computer Vision, Perth, Australia, 2018.

[17] Twentybn, Twenty Billion Neurons Inc. Toronto, Canada. (2017) [Online]. Available:https://20bn.com/datasets/jester.

[18] J. Materzynska, G. Berger, I. Bax, and R. Memisevic, "The jester dataset: A large-scale video dataset of human gestures," in Proc. IEEE/CVF Int. Conf. Computer Vision Workshop, Seoul, Korea (South), 2019, pp. 2874–2882.

[19] R. Haridy. (2017, Aug. 22). Microsoft's speech recognition system is no. as good as a human. Microsoft, Redmond, Washington. [Online] Available: https://newatlas.com/microsoft-speech-recognition-equals humans/50999/

[20] X. H. Yuan, L. B. Kong, D. C. Feng, and Z. C. Wei, "Automatic feature point detection and tracking of human actions in time-of-flight videos," IEEE/CAA J. Autom. Sinica, vol.4, no.4, pp.677–685, Sept. 2017.

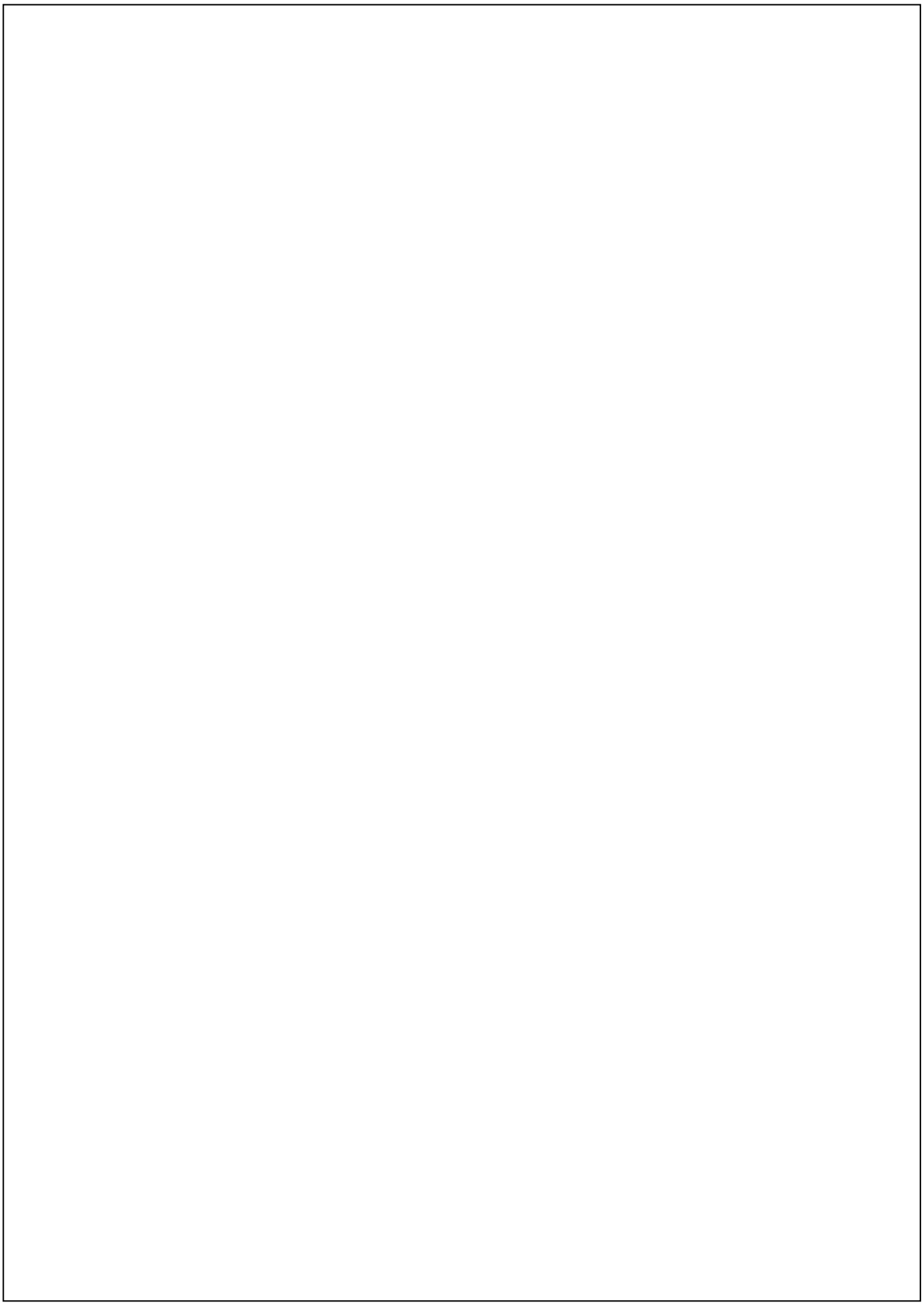
[21] R. Girdhar, D. Ramanan, A. Gupta, J. Sivic, and B. Russell, "ActionVLAD: Learning spatio-temporal aggregation for action classification," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Honolulu, USA, 2017, pp. 3165–3174.

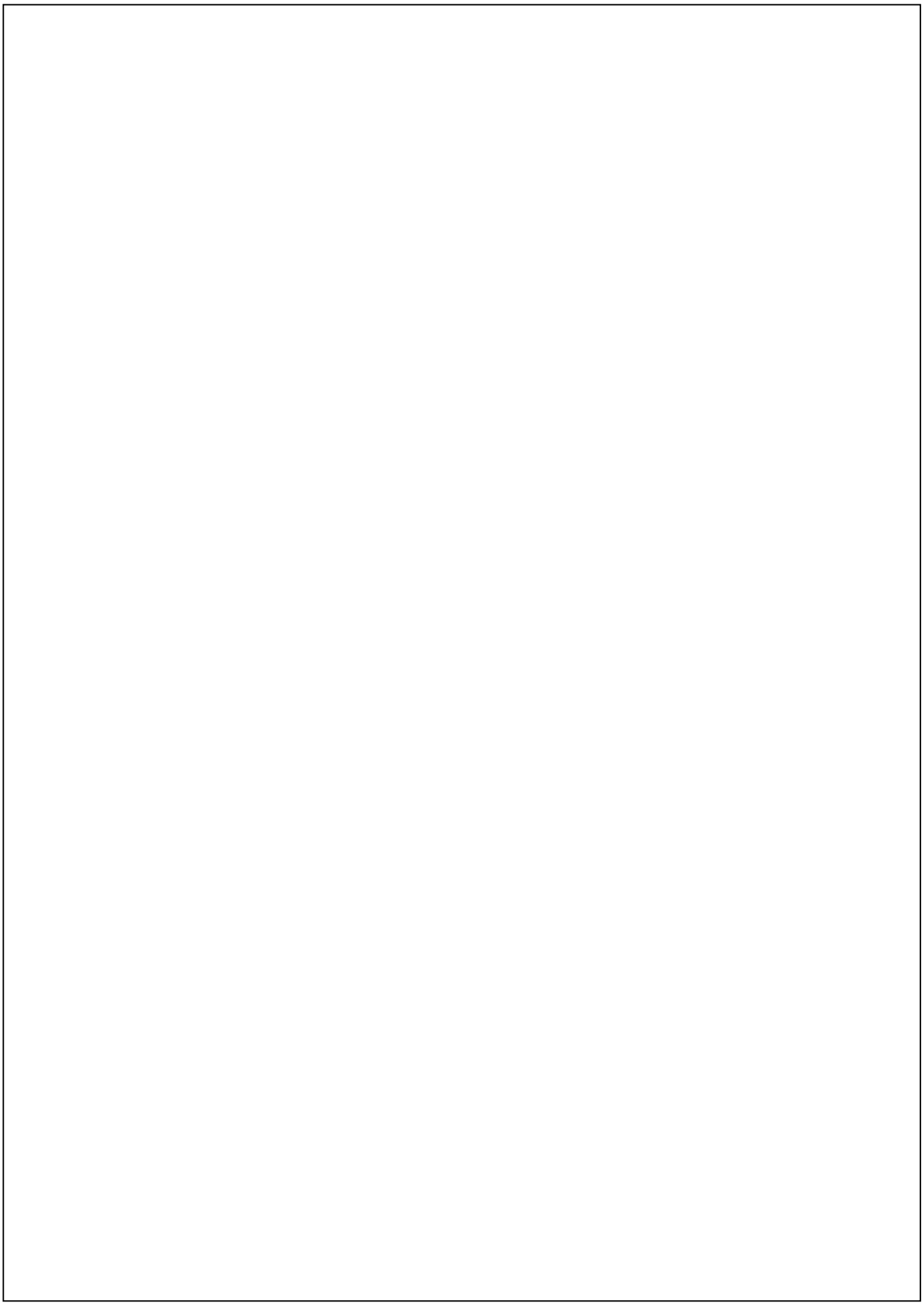
[22] L. M. Wang, Y. J. Xiong, Z. Wang, Y. Qiao, D. H. Lin, X. O. Tang, and L. Van Gool, "Temporal segment networks: Towards good practices for deep action recognition," in Proc. 14th European Conf. Computer Vision, Amsterdam, The Netherlands, 2016.

[23] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human computer interaction," inIET Comput. Vis., vol.12, no. 1, pp. 3–15, Feb.2018

[24] P. Mishra, K. Khurana, S. Gupta and M. K. Sharma, "VMAnalyzer: Malware Semantic Analysis using Integrated CNN and Bi-Directional LSTM for Detecting VM-level Attacks in Cloud," 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 2019, pp. 1-6, doi: 10.1109/IC3.2019.8844877.

[25] C. Zhu, J. Y. Yang, Z. P. Shao, and C. P. Liu, "Vision based hand gesture recognition using 3D shape context," IEEE/CAA J. Autom. Sinica, DOI: 3).1109/JAS.2019.1911534.
Conf. Computer Vision, Amsterdam, The Netherlands, 2016.





rpr1_1.docx

ORIGINALITY REPORT

12%	8%	13%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.fx361.cc Internet Source	4%
2	"Computer Vision – ACCV 2018", Springer Science and Business Media LLC, 2019 Publication	2%
3	Wenjin Zhang, Jiacun Wang, Fangping Lan. "Dynamic hand gesture recognition based on short-term sampling neural networks", IEEE/CAA Journal of Automatica Sinica, 2021 Publication	1%
4	ruor.uottawa.ca Internet Source	1%
5	www.ieee-jas.net Internet Source	1%
6	Qing Pan, Jintao Zhu, Gangmin Ning, Lingwei Zhang, Luping Fang. "ST-GCN AltFormer: GestureRecognition with Spatial temporal Alternating Transformer", Institute of Electrical and Electronics Engineers (IEEE), 2023 Publication	1%

7	pdfs.semanticscholar.org Internet Source	1 %
8	"Computer Vision and Image Processing", Springer Science and Business Media LLC, 2023 Publication	1 %
9	Wenjin Zhang, Jiacun Wang. "Dynamic Hand Gesture Recognition Based on 3D Convolutional Neural Network Models", 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), 2019 Publication	<1 %
10	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
11	Submitted to Eastern University Student Paper	<1 %
12	A Spoorthi Alva, R Nayana, Noorain Raza, Gambhire Swati Sampatrao, Koduru Bharath Subba Reddy. "Object Detection and Video Analyser for the Visually Impaired", 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2023 Publication	<1 %
13	Submitted to University of Salford Student Paper	<1 %

Exclude quotes	Off	Exclude matches	< 14 words
Exclude bibliography	On		