Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

Design a Contactless Authentication System Using Hand Gestures Technique in COVID -19 Panic Situation

Mohamed YacinSikkandar
Department of Medical Equipment Technology,
College of Applied Medical Sciences,
Majmaah University, Al Majmaah, 11952, Saudi Arabia,
Email: m.sikkandar @mu.edu.sa

Abstract

Spread of Covid 2019 pandemic has shattered the globe, putting everybody in panic stage. In view of its long-term impact on daily life, the need to wear face masks and maintain social distances necessitates. In this current satiation everyone need a contactless biometric communication system for all future authentication schemes. One of the alternatives is the use of Contactless Authentication System biometrics because of its non-contact alike normal biometric finger prints and is capable of recognising even people wearing face masks. In this work, a novel hand gesture-based sign digits scheme is proposed. This work presents the design and implementation of a deep learning system that can be verified without contact by the user using 'authentication code.' The 'authentication code' is a 'n' numeric code, and the digits are hand movements of the sign language digits. We propose a memory-efficient deep learning convolution neural network model to identify and classifying the hand movements of the sign language digits and also to extract the function by combing the two BEMD and SIFT algorithm techniques. The model is deployed in the Raspberry pi 4 Model B edge computing system to act as an edge device for user verification. The model achieves a classification accuracy of 98.47 percent for the publicly accessible sign language digits dataset.

Keywords: *deep learning method, contactless authentication, camera based authentication,* hand *gestures recognition and security, and edge computing.*

1. Introduction

The current COVID-19 pandemic has compromised the world at every point, so when the human race is removed, nobody can predict. Because of its long-term consequences, we humans need to consider and learn how to live with this virus. However, we can mainly rely on "contact less" technology [1] to combat this situation and to comfort our daily lives. Based on the COVID 19 crisis, biometric systems, which are regarded as a cornerstone of the safety system in every region, face many challenges. Nearly all organisations have stopped using biometric communication systems, a major reason for the spread of corona viruses [2]. The secure authentication option is no more regarded as biometric authentication systems that are imported from a small business to the highest safety priority [3].Contactless biometric authentication has become a necessity at this hour as it is perceived to be not only more hygienic but secure and efficient. Biometrics can be divided into two classes, physiological biometrics and behavioral biometrics [4]. Physiological biometrics usually includes fingerprints, facial features, palm print, retinas, ears, and irises. Behavioural biometrics usually consists of keystrokes, signatures, and gaits. Voice biometric can be classified as both because it includes features belonging to both the classes [5].Before the era of Deep Learning, biometric authentication was mainly based on hand-crafted features that were extracted using methods such as scale invariant feature transform (SIFT) [6], wavelet, etc. With the advent of deep learning in this decade, the biometric authentication field is completely transformed. Most contemporary biometric authentication systems use convolution neural networks and different variants of them.

A convolutional neural network (CNN) is a type of deep multilayer artificial neural network. They are widely used in all computer vision tasks because of their convolution process that obtains an effective

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

representation of the input images directly from raw pixels with little to none preprocessing and can easily recognize visual patterns [7]. The representations learned by the CNN models are that of visual features and are effective when compared to the handcrafted features [8]. These networks are being applied in a variety of applications,such as object detection tasks, speech recognition tasks. The deep learning has brought a lot of progress in the field of biometric authentication there are still several challenges to overcome. Some of the challenges are, more challenging datasets have to be developed to train the models, the need for interpretable deep learning models, real-time deployment of the models, memory-efficient models, security and privacy issues, etc.Also in the security space, the most dominant usage of CNNs had been in the intrusion detection [9]. These networks primarily use object detection modules and the recent development of CNNs has been proven to be effective for object detection [10]. Many of these CNNs are memory hungry due to a high number of parameters (hundreds of millions) and have high computational complexity.Thus to tackle the real-time deployment and efficiency challenges of the deep learning models, so we provide an end-to-end contactless authentication system that verifies a user by validating the 'authentication code' using a memory-efficient CNN model. The 'authentication code' which is unique for each user and is an 'n' digit numeric code with each digit having a range of 0–9. The task of automatically classifying an input image into one of the given classes using convolution neural networks is known as image classification task.

Hand gestures are mainly different combinations of fingers producing different shapes of a hand. Thus, the primary focus of gesture recognition methods that use image processing is shape matching or measuring dissimilarity among hand shapes. For instance,Wu et al. [11] represent the pixels of each fingertip as samples of 2D Gaussian distribution in the output tensor of Heatmap-based FCN. By applying a suitable threshold, only the visible fingertips are detected that determines the gesture at the same time. Ren et al. [12] presented a part-based gesture recognition system that uses dissimilarity measure and template matching for an HCI application of arithmetic computation by gesture command. Similarly, Alon et al. [13] use spatiotemporal matching and pruning classifier for gesture learning to recognize the American Sign Language (ASL).

Koller et al. [14] embedded a CNN within an iterative expectation-maximization (EM) algorithm for the classification of hand shapes particularly in the case of continuous and weakly labeled data. Lin et al. [15] proposed that the background of a hand can be segmented first by using the Gaussian mixture model (GMM) and then the binarized image can be feed to a CNN classifier for learning instead of directly using the captured RGB image for hand gesture recognition.. Nunez et al. [16] reported a method that combines the CNN and the long short term memory (LSTM) network for skeleton-based temporal 3D hand gesture recognition.

Huang et al. [17] use two-stage CNN to detect fingertips from a hand image for an application of air writing wherein a fingertip acts like a pen. Jain et al. [18] report the detection of only the index fingertip using a direct regression approach for an MR application in which the fingertip functions as a gestural interface for smart-phones or head-mounted devices. Wetzler et al. [19] mainly focus on CNN-based fingertip detection using a Kinect camera. This method uses a computationally extensive global orientation regression approach and an in-plane derotation scheme of depth images to predict the coordinate of fingertips.

Lee et al. [20] estimates the scale-invariant angle between the fingers to determine the different number of visible fingertips. Afterward, fingertip gestures are recognized using a contour analysis of the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

fingers. Nguyen et al. [21] use a deep learning-based approach where a modified multi-task segmentation network is employed for both segmentation of hand and detection of a variable number of fingertips.

2. Materials and Methods

Dataset

We have utilized 'Sign Language Digits Dataset' [22] for the training of the proposed CNN and MobilenetV2. The dataset consists of a total of 2062 samples. There are 10 classes from digit 0 to digit 9. Samples from each class are shown in Figure 1. Each sample is of thesize [100 × 100] pixels. Table 1 shows dataset details with the total sample count per class. The entire dataset is split into three parts i.e. training data, validation data, and test data. Since there are 10 classes, the test data is split in such a way that there are equal numbers of samples in each class. We randomly selected 41 samples from each class leading to a total of 410 samples for the test set, which is 19.88% of the entire dataset. The rest of the dataset is split into training and validation sets with 20.07% and 60.03% samples of the dataset respectively.

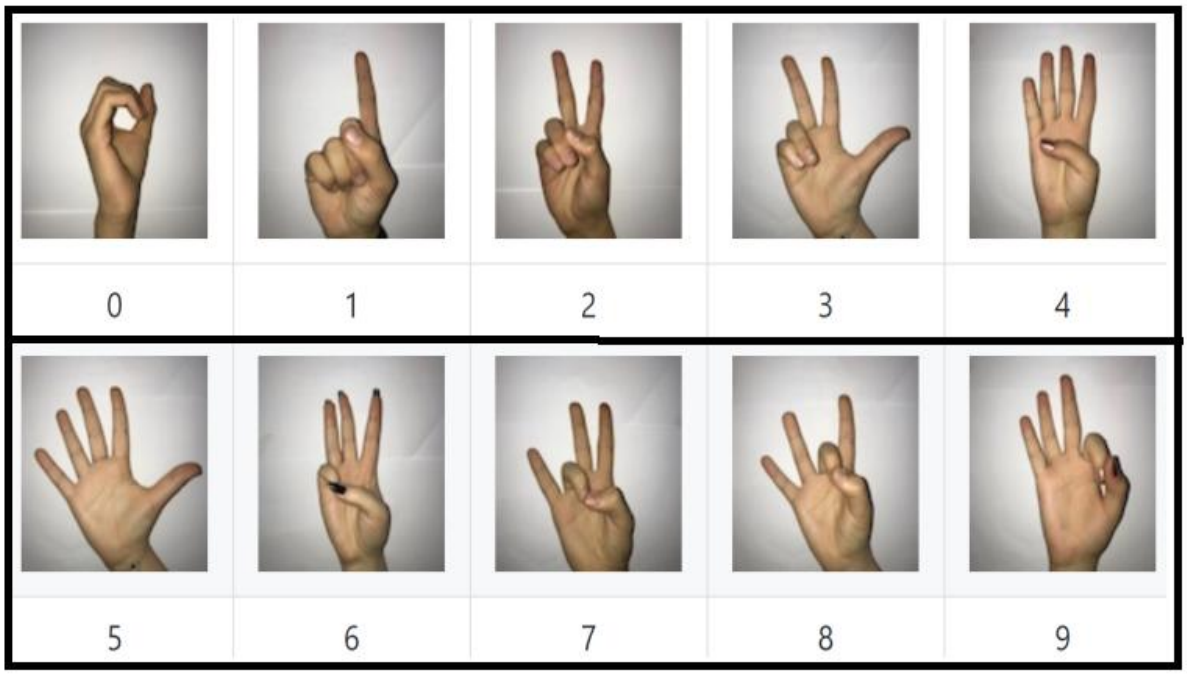

Figure 1. *Sample images of the dataset*

As mentioned above, the image size of each sample in the dataset is [100 × 100] pixels. These image samples are resized (upscale) to [256 × 256] pixel size. And also table 1 signifies the dataset training and testing values for this technique.

Table 1: *Dataset description*

Class	Number of Samples	Number of Training	Number of Validation	Number of Testing
1	*205*	*123*	*41*	*41*
2	*206*	*124*	*41*	*41*
3	206	*124*	*41*	*41*
4	207	*124*	*41*	*41*
5	207	*124*	*42*	*41*
6	207	*124*	*42*	*41*
7	206	*124*	*42*	*41*

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

8	*208*	*125*	*42*	*41*
9	*204*	*122*	*41*	*41*
Total	*2062*	*1238*	*414*	*410*

3. Proposed End to End System for Contactless Authentication

In this work we use CNNs for providing contactless authentication code from input images of hand gestures of sign language digits. The whole system is composed of two steps namely capture of images and classification (digit recognition) of the captured images. The entire classification task used in the system is shown in Figure 2. As shown in Figure 2 the input image to the system is first resized to [256 × 256] pixel size and is then fed into the deep learning model for digit recognition

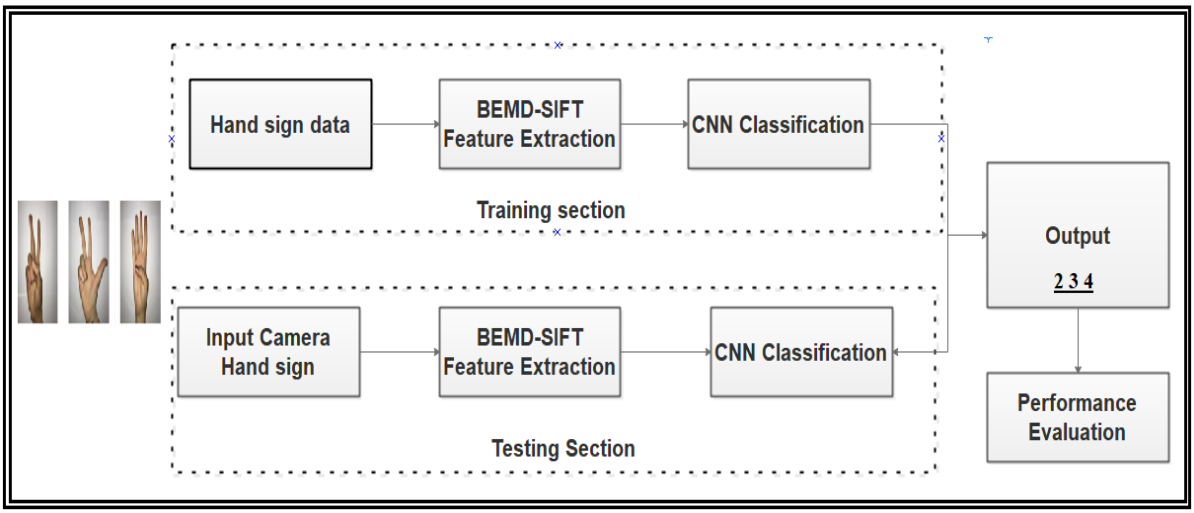

Figure 2.*Block diagram ofproposed methodology*

3.1 FEATURE EXTRACTION

The most important part of the object recognition system isthe method used to transform the image to its feature vector.This process is known as the feature extraction. In this workBEMD- SIFT features are used which are discussedin the sections following.Bi-dimensional Empirical Mode DecompositionEdge is one of the important features in an image since itcarries important information about the objects that arepresent in an image. In this work the edge detection is basedon the BEMD technique. EMD is basically a decomposition technique whichdecomposes the nonlinear and non-stationary data to form aset of Intrinsic Mode Functions (IMF) through the siftingprocess. EMD deals with one dimensional data, BEMD wasproposed to cope with two dimensional data. An image isconsidered to be a two dimensional data $f(x, y)$ where $x = 1, \ldots \ldots \ldots, M$, and $y = 1, \ldots \ldots \ldots, N$, and the BEMD is appliedon it.

3.2 SIFT Feature

In SIFT is invariant in rotation, scaling, minor noise, lighting changes and changes in point of view. The four key computational stages for image generation are:

1. Scale-space extrema – In the initial stage ofcomputation the image is searched over all possiblescales and locations. This is achieved by using the Gaussian function in an efficient way to classify invariant points of interest in scale and orientation.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

2. Keypoint localization – The first stage detects manyinterest points and in the second stage the interestpoints with low contrast and the interest points thathave poor localization along the edges are removed.Thus, based on the stability, the strong interest pointsare selected as keypoints.

3. Orientation assignment – According to the local image gradients, one or more directions are allocated to all keypoints. The operations of these images in the future shall be carried out on the assigned scale, position and orientation, thus making the images invariant to alter.

4. Keypoint descriptor – The area around the key point is considered for any key point for calculating local gradients, which can be converted to a representation, with substantial distortions of the local shape and changes in lighting..

3.3 Construction of SIFT Descriptor

In below Figure 3 depicts the construction of the SIFT descriptors, basedon the sample of the image gradients magnitudes andorientations that are sampled around a keypoint location. Thekeypoint scale is used to choose the level of Gaussian blur.According to the keypoint orientation the descriptorcoordinates are rotated to make an image orientation invariant.This concept is illustrated in the Fig. 3. The 4x4 sampleregions are considered for the creation of orientationhistograms that allows for the significant gradient positionshifting. For every orientation histograms, Fig. 3 shows the 8direction orientation, where the length of the arrow representsthe histograms magnitude. Each sample is composed of 4x4array with 8 bin orientation leading to 128 keypointdescriptors.

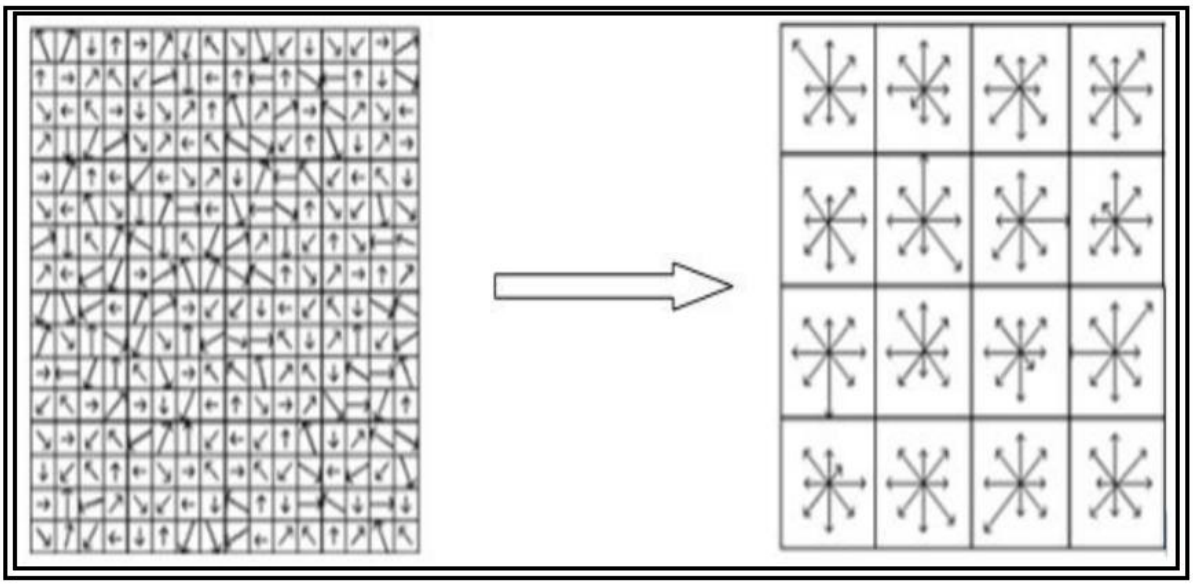

Figure.3 *SIFT descriptors Generation*

3.4 Image feature extraction using proposed BEMD–SIFT

We detail the BEMD–SIFT process for image extraction in this section. The algorithm's key concept is to decompose BEMD and extract SIFT functionality. Features must be extracted from BEMD images to obtain several BIMF images and the general pattern during decomposition. In addition, the SIFT algorithm break down the BIMF and the global pattern for image extraction in the fusion process separately. In order to obtain the original image feature information, the information for synthesis and accumulation is combined. The extraction feature is accelerated from the original image. In addition, an objective and effective extraction of the SIFT algorithm is solved. Provided that the breakdown of the BIMF and trends

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

well reflect the nature of the original picture information, the problem of insufficient details can also be resolved. Fig 4demonstrates the fundamental idea..

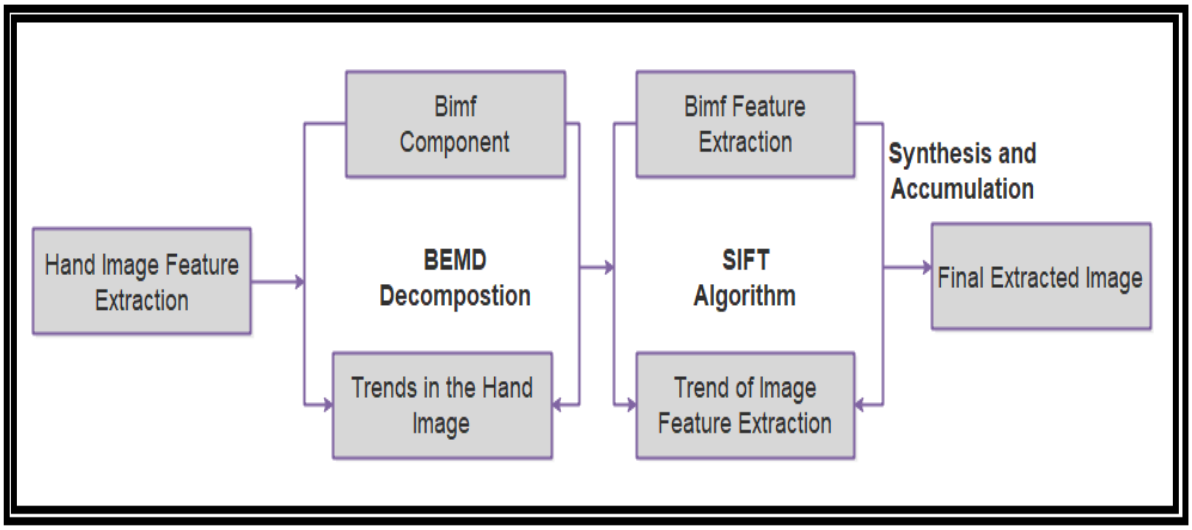

Figure 4.*BEMD–SIFT algorithmprincipleDiagram*

The basic steps of the BEMD-SIFT-based image extraction algorithm are as follows:

(1) As a f(x,y) BEMD operation, the information on decomposed pictures should be extracted and several images and trends of the BIMF component identified.

$$f(m,n) = \sum_{k=1}^{k} bimf_k(m,n) + r_k(m,n) \qquad (1)$$

(2) Step 1, with the SIFT algorithms mentioned in this analysis, the decomposed BIMF portion and the image trend are determined in terms of feature information extraction. This step is important to promote rotation invariance since an orientation keypoint descriptor is represented. The gradient magnitude M(x, y) and the guidance ̈For each neighbouring pixel is determined as follows in the light of a keypoint located in Gaussian imageL(x, y, ̈):

$$m(x,y) = \sqrt{L(x+1,y) - L(x-1,y))^2 + L(x,y+1) - L(x,y-1))^2} \qquad (2)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right) \qquad (3)$$

(3) Many components of BIMF are generated in Phase (2). In order to obtain final feature of the original image and detailed details, the trend of extracting synthetically accumulated feature information is created.

3.5 CNN Design

For gesture recognition and fingertip detection, the relevant portion of the hand is cropped from the input image using a bounding box and resized to (128 × 128). The resized image is used as the input to the proposed network for learning.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

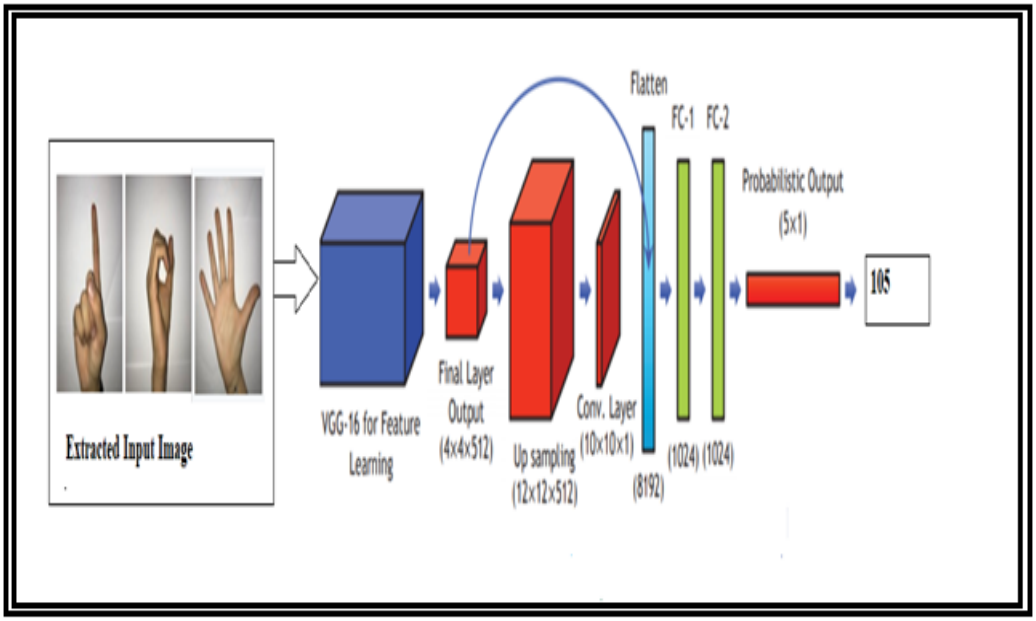

Figure 5: *A block diagram of the CNN architecture with input and output.*

During detection, the real-time object detection algorithm is used for hand recognition in the first stage. Later, that hand portion can be cropped and resized to feed to the proposed framework. For feature learning, 16-layers visual geometry group (VGG) configuration given in is employed. This output is utilized to generate both the probabilistic output and positional output. First, the output of the feature learning stage is flattened and two FC layer is used back-to-back for better classification. Each of the FC layers is followed by a rectified linear unit (ReLU) activation function and a dropout layer. Finally, an FC layer is appended at the end to reduce the feature vector size to the same as that of the desired probabilistic output P of length N given by

$$P = \left[p_t p_i p_m\, p_r\, p_p\right]^T \tag{3}$$

where from$p_t$to p_p are the probability of thumb (t), index (i), middle (m), ring (r), and pinky (p) finger, respectively. A sigmoid activation function is applied to the output of the final FC layer to normalize the probabilistic output. Moreover, the output of the feature learning stage is up-sampled followed by a ReLU activation function. Next, a convolution operation with a single filter is performed to further reduce the size of the feature vector to the same as that of the desired ensemble of positional output X of size 2N × 2N given by

$$x = \begin{bmatrix} x_t\, y_t\, x_i\, y_i\, x_m y_m\, x_r\, y_r\, x_p\, y_p \\ x_t\, y_t\, x_i\, y_i\, x_m y_m\, x_r\, y_r\, x_p\, y_p \end{bmatrix} \tag{4}$$

Where,$xf\ and\ yf\ (f \in t,i,m,r,p)$ stand for the coordinate position of the fingertips from thumb to pinky finger successively. In the final convolution operation, a linear activation function is applied. Finally, the column-wise ensemble average is taken as the final output of the fingertip positions. The overall system with CNN architecture is presented in Fig. 5. The activation functions and dropout layers are not shown in the figure for brevity.In the proposed framework, the probabilistic output and the positional output need to be optimized independently at the same time and thus two loss functions are defined. The probabilistic output predicts the binary sequence of '1' and '0' considering the visibility of the finger,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

andtherefore, the following binary cross-entropy loss function is proposed to optimize the probabilistic output.

4. Experiments and Results

This work is implemented on MATLAB. The system specification used in this work is Intel Core i5, 2.2 GHz, 8 GB DDR4 RAM system. The model were conducted on Sign Language Digits Dataset. Experiments are performed based on the proposed method to validate the unified gesture recognition and fingertip detection algorithm. This characteristics of the dataset on which experiments are carried out and a short description of data augmentation which is applied during the training period of the network. Afterward, the training and detection procedure of the gesture recognition are explained. All the training and testing code concerning the experimentations and results along with the pre-trained weights of the model are publicly available to download.

4.1. Deployment on Edge Computing Device

A low-cost edge computing hardware, the Raspberry pi 4 model B microprocessor is utilized for implementing the given task. Raspberry pi 4 model B is the latest product from the raspberry pi collection of single-board edge computing devices. This new version of raspberry pi provides increased processor speed, connectivity, and memory capacity compared to the raspberry pi 3 model B+. The cost of this edge computing device is around $35. We attached a Raspberry pi camera V2.1 camera module to the raspberry pi 4 micro-controller to capture images in real-time. The camera module is capable of capturing images at 3280 × 2464 pixel resolution. It can also take videos at 1080p30, 720p60 and 640 × 480 p90 quality. The camera module costs around $27. The trained model is deployed on the raspberry pi 4 to predict the hand gestures. The complete setup of the hardware has been shown in Figure 6.

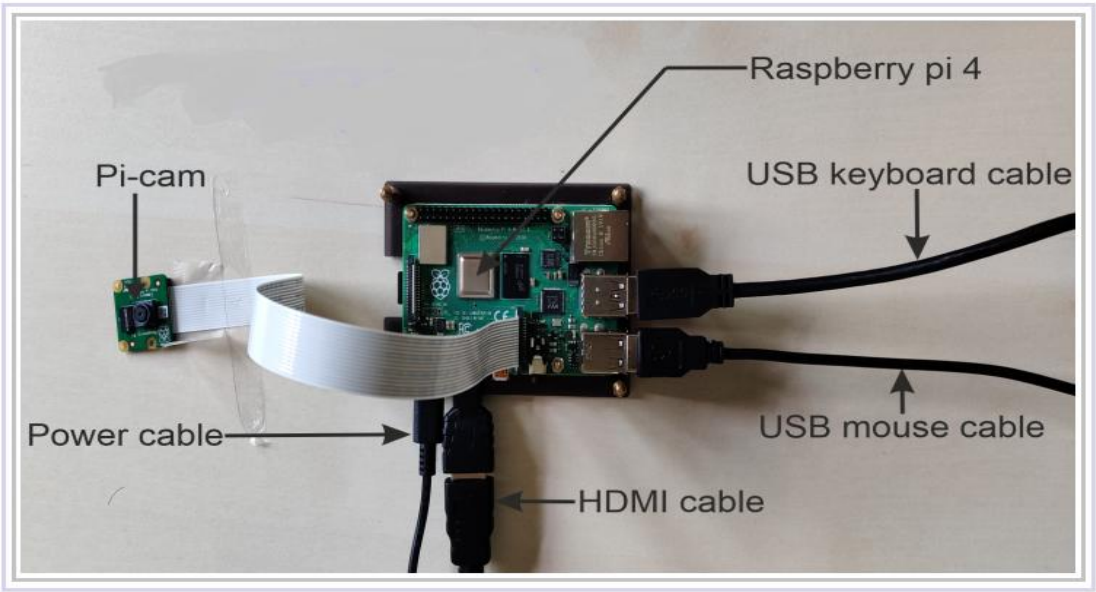

Figure 6. *Hardware setup*

4.2 Performance Metrics

The performance of the classification of hand gestures and that of estimation of the fingertips position are evaluated separately. The performance of the classification is assessed in terms of four measures, namely, accuracy, precision, recall, and F1 score. The higher the value of accuracy or F1 score,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

and the closer the value of precision or recall to unity, the better is the performance of the classification algorithm.

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (6)$$

$$F - measure = \frac{2TP}{2TP+FP+FN} \times 100 \quad (7)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (8)$$

Where, FP is signified as false positive, TN is indicated as true negative, TP is specified as true positive, and FN is definite as false negative.

Table 2. Comparison analysis of feature extraction method with CNN classifier

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F - measure(%)
BEMD with CNN	96.67	99	96.40	75.69	100	85.71
SIFT with CNN	93.20	100	88.49	75.58	99	66.17
BEMD-SIFT with CNN	98.47	100	98.61	84.15	100	99.66

In table 2 and figure 7 signifies the performance analysis of proposed model, which model analysis done by three different scenarios as feature extraction method combined and separated with CNN (BEMD with CNN), (SIFT with CNN) and (BEMD+SIFT with CNN). By this comparison study, feature extraction technique of BEMD+SIFT technique with CNN classifier combination achieved better performance than separate of that two feature extraction techniques.

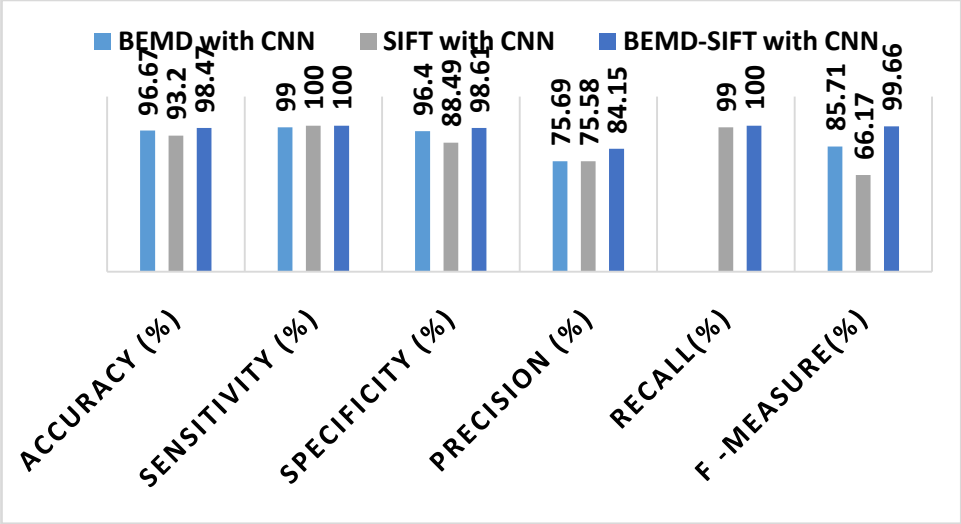

Figure 7. *Graphical representation of experimental setup performance*

4. Conclusion

In this paper, a new CNN-based method is proposed that unifies the gesture recognition and prediction of fingertip position in a single step process.In order to cope up with the COVID-19 pandemic,

this research proposes a robust particular region based biometric authentication system. In particular, the proposed method regressed the ensemble of the position of finger using a fully convolutional network instead of directly regressing the positions of finger using the fully connected layer. The experiments have been carried out by employing a commonly referred Sign Language Digits Dataset. The accuracy of the automatic gesture recognition has been found to be at least 98.47%, and the minimum F1 score among the classes have been found to be at least 99.66. Moreover, the proposed method has achieved lower false positive and false negative rates in classification and made less localization error in regression. The performance of the proposed method is also ensured by experimentation using the hand gesture images available in the wild. In conclusion, with the speed of the detection, and accuracy in performance, the proposed algorithm can play a significant role in several real time applications.

5. Reference

[1]. Givi B, Schiff BA, Chinn SB, Clayburgh D, Iyer NG, Jalisi S, Moore MG, Nathan CA, Orloff LA, O'Neill JP, Parker N. Safety recommendations for evaluation and surgery of the head and neck during the COVID-19 pandemic. JAMA otolaryngology–head & neck surgery. 2020 Jun 1;146(6):579-84.

[2]. M.Kavitha, T.Jayasankar, P.Maheswara venkatesh, G.Mani, C.Bharatiraja, and Bhekisipho Twala, "COVID-19 Disease Diagnosis using Smart Deep Learning Techniques", Journal of Applied Science and Engineering (2021), vol.24,No.3. http://dx.doi.org/10.6180/jase.202106_24(3).0001

[3]. Aman AH, Hassan WH, Sameen S, Attarbashi ZS, Alizadeh M, Latiff LA. IoMT amid COVID-19 pandemic: Application, architecture, technology, and security. Journal of Network and Computer Applications. 2020 Nov 2:102886.

[4]. Lakshmi R. Nair, KamalrajSubramaniam. G. K. D. PrasannaVenkatesan, P. S. Baskar ,T. Jayasankar, "Essentiality for bridging the gap between low and semantic level features in image retrieval systems: an overview", Journal of Ambient Intelligence and Humanized Computing (2020),ISSN: 1868-5137 (Print) 1868-5145 (Online) IF.4.594.

[5]. Iannizzotto, G.; Rosa, F. A SIFT-Based Fingerprint Verification System Using Cellular Neural Networks. In Pattern Recognition Techniques, Technology and Applications; I-Tech: Vienna, Asutria, 2008.

[6]. P.Shanmugapriya, V.Mohan, T.Jayasankar, Y.Venkataramani, "Deep Neural Network based Speaker Verification System using Features from Glottal Activity Regions", Appl. Math. Inf. Sci. vol.12, no.6, Nov 2018, pp.1147–1155.
DOI: http://dx.doi.org/10.18576/amis/120609

[7]. A. Sheryl Oliver, Kavithaa Ganesan, S. A. Yuvaraj, T. Jayasankar, Mohamed Yacin Sikkandar &N. B. Prakash, Accurate prediction of heart disease based on bio system using regressive learning based neural network classifier, Journal of Ambient Intelligence and Humanized Computing (2020), https://doi.org/10.1007/s12652-020-02786-2

[8]. Aizat, K.; Mohamed, O.; Orken, M.; Ainur, A.; Zhumazhanov, B. Identification and authentication of user voice using DNN features and i-vector. Cogent Eng. 2020, 7, 1751557.

[9]. S. Ramesh, C. Yaashuwanth, K. Prathibanandhi, Adam Raja Basha, T. Jayasankar, "An optimized deep neural network based DoS attack detection in wireless video sensor network", Journal of Ambient Intelligence and Humanized Computing (2020),https://doi.org/10.1007/s12652-020-02763-9

[10]. Kim, K.H.; Hong, S.; Roh, B.; Cheon, Y.; Park, M. Pvanet: Deep but lightweight neural networks for real-time object detection. arXiv 2016, arXiv:1608.08021.

[11]. W. Wu, C. Li, Z. Cheng, X. Zhang, L. Jin, Yolse: Egocentric fingertip detection from single rgb images, in: Proceedings of the IEEE Int. Conf. on Computer Vision, Venice, Italy, 2017, pp. 623–630.

[12]. Z. Ren, J. Yuan, J. Meng, Z. Zhang, Robust part-based hand gesture recognition using kinect sensor, IEEE transactions on multimedia 15 (5) (2013) 1110–1120.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 2149 - 2159
Received 15 December 2020; Accepted 05 January 2021.

[13]. J. Alon, V. Athitsos, Q. Yuan, S. Sclaroff, A unified framework for gesture recognition and spatiotemporal gesture segmentation, IEEE transactions on pattern analysis and machine intelligence 31 (9) (2008) 1685–1699.

[14]. O. Koller, H. Ney, R. Bowden, Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 3793–3802.

[15]. H.-I. Lin, M.-H. Hsu, W.-K. Chen, Human hand gesture recognition using a convolution neural network, in: Proc. IEEE Int. Conf. on Automation Science and Engineering (CASE), IEEE, Taipei, Taiwan, 2014, pp. 1038– 1043.

[16]. J. C. Nunez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, J. F. Velez, Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition, Pattern Recognition 76 (2018) 80–94.

[17]. Y. Huang, X. Liu, X. Zhang, L. Jin, A pointing gesture based egocentric interaction system: Dataset, approach and application, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, Las Vegas, NV, USA, 2016, pp. 16–23.

[18]. V. Jain, R. Hebbalaguppe, Airpen: A touchless fingertip based gestural interface for smartphones and headmounted devices, arXiv preprint arXiv:1904.06122.

[19]. A. Wetzler, R. Slossberg, R. Kimmel, Rule of thumb: Deep derotation for improved fingertip detection, arXiv preprint arXiv:1507.05726.

[20]. D. Lee, S. Lee, Vision-based finger action recognition by angle detection and contour analysis, ETRI journal 33 (3) (2011) 415–422.

[21]. H.-D. Nguyen, S.-H. Kim, Hand segmentation and fingertip tracking from depth camera images using deep convolutional neural network and multi-task segnet, arXiv preprint arXiv:1901.03465.

[22]. ZeynepDikle, A.M.; Students, T.A.A.A.H.S. Sign Language Digits Dataset. 2017. Available online: https://github.com/ardamavi/Sign-Language-Digits-Dataset (accessed on 2 June 2020).