IoT-based Face Recognition with SNN

Janhavi Adhau, Vaibhav Adsare, Deep Jadhav, Dr. Ketki Kshirsagar, Dr. Jayashree Tamkhade
vaibhav.22010758@viit.ac.in,janhavi.22010552@viit.ac.in, deep.22010818@viit.ac.in,
ketki.kshirsagar@viit.ac.in, jayashree.tamkhade@viit.ac.in
Electronics and Telecommunications, BRACT'S Vishwakarma Institute of Information Technology Pune, India

1. Abstract
This idea presents a face recognition (FER) algorithm using a deep Siamese neural network (SNN) that stores local patterns of images in a similar space. We create networks to show similarity between pairs of ideas by comparing features with design. Additionally, we developed a new image matching (i.e., positive-negative matching) technique to introduce our Siamese model. Our Siamese model has a proven and accepted model for learning in an integrated environment. The validation method reduces the class variance by reducing the distance between features drawn from the same class, whereas a validation procedure raises the class variance by widening the gap between characteristics drawn from different classes. By using four layers of IoT we successfully implemented the multimodal authentication system with the Biometric and Face Recognition System as a mode for the authentication. In this project machine learning with IoT to make fulfill the product requirements and a more secure authentication system.

Keywords: Siamese Neural Network, Face Recognition Algorithm, IoT, Multimodal Authentication System

2. Introduction
Neural networks are pretty much perfect for any task in the era of deep learning, but they require more data to do so. However, you cannot always rely on obtaining additional data for tasks such as facial recognition and authentication. There is Siamese networks are a brand-new kind of architecture for artificial neural network to solve problems like this. In the past, neural networks were trained to predict various classifications. This causes problems when A new set of classes must be added to your data. It is necessary in this case to update and retrain the network using the entire data set. There is a lot of data required for deep neural network training. As opposed to this, SNN recognizes a similarity function. This information may be used to train the system to check whether two images match, which is what we do in this instance. Since the network doesn't need to be retrained, we may categorize new data types. Use fewer pictures for more accurate predictions [1]. Siamese networks have gained more popularity in recent years due to their ability to learn from relatively small amounts of data. The largest benefit, however, is that, in the instance of facial recognition, a new employee has joined the company. Now we just need one image of his face, which will be saved in the database, for the network to recognize his face. The network will determine the similarity for any new instance that is provided to it using this as the reference image. As a result, we may state that the network predicts the score in one go.

In this Project using Siamese networks and OpenCV-Python, we will demonstrate a face validation approach. Which will be used as one of the methods of authentication for our security system. Additionally, this project We used all four IoT layers in this project, we link several aspects of machine learning and the Internet of Things making it safer and better for user experience.

3. Literature Review
The following literature surveys provide a useful resource for researchers and practitioners interested in face recognition with Siamese neural networks. They cover the latest techniques, challenges, and future directions in this field and can serve as a starting point for further research.

Siamese Neural Networks for One-shot Image Recognition- Gregory Koch, et.al, [2] They described a approach for which learning deep convolutional Siamese neural networks first is crucial to verification before doing one-shot classification. They presented fresh findings evaluating the performance of their networks against a current, cutting-edge classifier created for the Omniglot data set. Their networks perform significantly better than all existing baselines and are quite close to the best results obtained by the previous authors. They claimed that the great performance of these networks on this challenge demonstrated the viability of their metric learning approach as well as the need to apply it to one-shot learning tasks in other domains, particularly for image classification. In this study, they solely took into consideration how to process image pairings and train for the verification task. Utilising a global affine transform, their distortions. They have been testing a more advanced algorithm that uses information about each stroke's trajectory to produce final computed distortions.

Human face recognition based on convolutional neural network and augmented dataset - Baoye Song, et.al, [7] learning deep convolutional Siamese neural networks first is crucial to the method's verificationControl Engineering. In order to address the issue of human face recognition on a limited original dataset, a novel strategy has been devised in this study. The original, modest dataset is expanded using the facial image changes, such as flip, shift, scale, and rotation, to produce a big dataset. Face recognition can be accomplished successfully using a clever Based on the noticeably improved face dataset, CNN. The utility of the enhanced dataset is tested in several trials, and the superiority of the new strategy can be shown by comparing it to some of the most often used ones. face recognition techniques. In fact, the suggested approach is an economical way to increase the dataset and may be used in a number of sectors connected to data-based education and training the data augmentation approach to more difficult problems, such as signal processing, picture identification, and image-based defect detection, will be the main goal of our future research.

Support Vector Machines Applied to Face Recognition P. Jonathon Phillips National Institute of Standards and Technology [8]. They unveiled a fresh method for using SVM for facial recognition. They used applications for identification and verification to show the technique. They compared our algorithm's performance to a PCA-based algorithm. To verify, the identical error rate their approach was 7% instead of 13%, which is approximately half that of the PCA algorithm. SVM's identification error was 22-23% versus 46%, which was half that of PCA. This shows that SVM is using the information in face space more effectively than the standard PCA method.

3.4 Face Recognition Based on Haar-Like and Euclidean Distance-To cite this article: Hao Wu et al [9]. This study combines haar-like features and the Euclidean distance method to recognize faces in images. By carefully taking into account how the choice of face features and various lighting conditions affect attendance images, the detection rate of multiple faces is increased. The grey face image in this publication is Face feature values screened using the Haar model, the cascade classifier is created during training, Last but not least, to determine how similar two faces are, the Euclidean distance approach is used. are required toAccording to experimental findings, the Haar feature and Euclidean distance perform better for face image identification under the same data set than a single improved Mahalanobis distance and Euclidean distance technique.

4. Research Methodology

4.1 Dataset

The most important research papers in the field that were published between 2010 and 2020 have been examined. LFW is the dataset used for face recognition. LFW is a database of images of faces created with the purpose of researching the issue of unrestricted face recognition. It is used for image recognition, or the recognition of human faces. More than 13,000 facial photos gathered from the internet are included in the data collection. Names of the people shown have been written on each face. 1680 of the individuals shown had two or more different images included in the data set.[3].

Figure no 4.1.1 - Labelled Face in Wild Home Databases

4.2 Siamese Neural Network Algorithm for Face Recognition

In order to train a machine learning model for face recognition, it is necessary to collect a dataset of positive and anchor images. This can be done using a combination of the Labelled Faces in the Wild dataset and a webcam. collecting positive and anchor images for face recognition training involves untiring and uncompressing the Labelled Faces in the Wild dataset to collect negative images, and using a webcam to capture and generate positive and anchor images with unique file paths. The images can be reviewed and the program should be designed to break gracefully and release the webcam when necessary.

Table 4.2.1 Embedding Model representing the unique features of each individual's face.

Layer (type)	Output Shape	Param
input_img (Input Layer)	[(None, 100, 100, 3)]	0
conv2D-12	(None, 91, 91, 64)	19264
MaxPooling-9	(None, 46, 46, 64)	0
conv2D-13	(None, 40, 40, 128)	401536
MaxPooling-10	(None, 20, 20, 128)	0
conv2D-14	(None, 17, 17, 128)	262272
max_pooling2d_11 (MaxPooling	(None, 9, 9, 128)	0
conv2d_15 (Conv2D)	(None, 6, 6, 256)	524544
flatten_3 (Flatten)	(None, 9216)	0
dense_5 (Dense)	(None, 4096)	37752832
Total params:	38,960,448	
Trainable params	38,960,448	

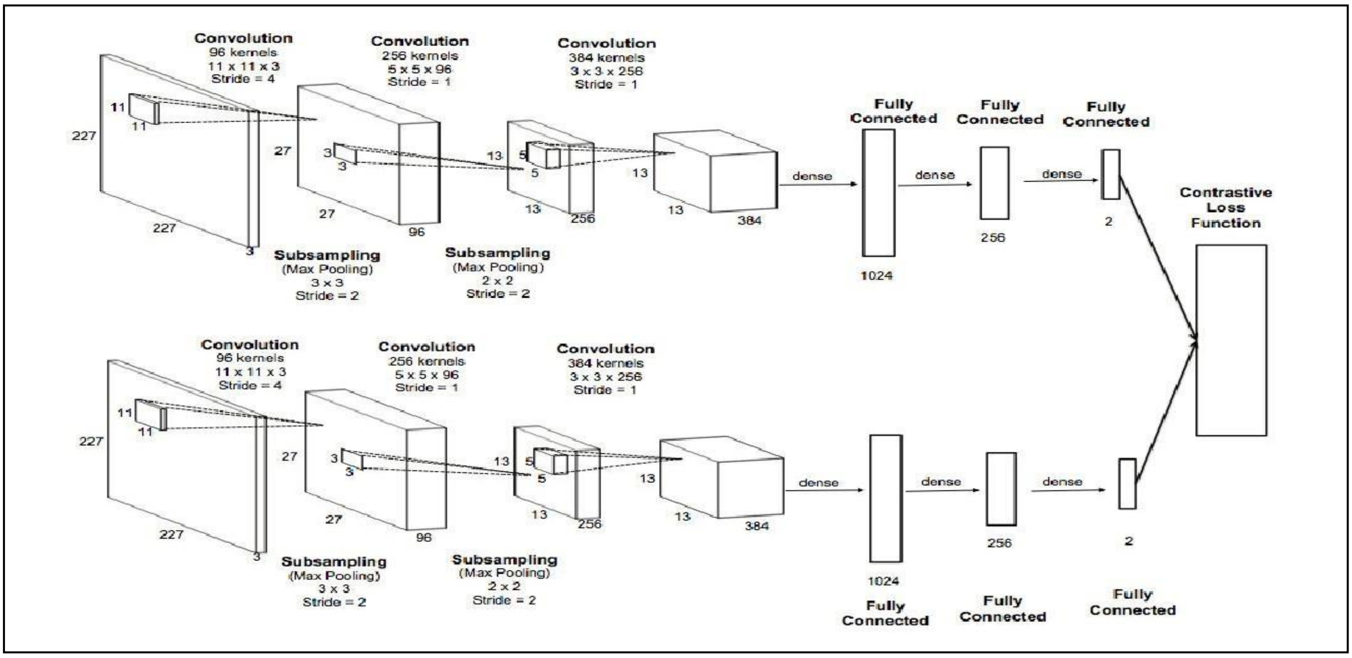

Figure no 4.2.2 - The best convolutional architecture was chosen for the purpose of verification. The Siamese twin links The L1 component-wise distance between vectors is computed in the not shown 4096 unit completely-connected layer.

Scaling and resizing are important pre-processing steps as they ensure that the images have consistent features, which improves the accuracy of the model. The labelled dataset helps the model differentiate between the faces of different individuals after which it was separated into a training and testing portion. The dataset is divided into two divisions utilising a data loader pipeline to do this. The training partition is used to train the model.and its effectiveness is assessed using the testing partition. An embedding layer is built, which is responsible for generating embeddings of the input images. The embeddings represent the unique features of each individual's face and are used to recognize the individuals in the images. [13]

The Siamese L1 Distance class uses a similarity calculation method to compare the embeddings of the anchor image and the validation image. The similarity calculation method computes the L1 distance between the embeddings, which is used to determine the similarity between the two images. The Siamese model combines the Siamese distance components by using the Siamese L1 Distance class to calculate the similarity between the anchor and validation images.

Table 4.2.2 Siamese network Model representing the unique features of each individual's face.

Layers	Output	Parameters
input_img (Input Layer)	(None, 100, 100, 3)	0
validation_img (Input Layer)	(None, 100, 100, 3)	0
embedding (Functional)	(None, 4096)	38960448
distance (L1Dist)	(None, 4096)	0
Dense	(None, 1)	4097
Total parameter	38,960,448	
Trainable param.	38,960,448	

The loss function for binary cross entropy function is implemented in the class Binary Cross entropy, which is a subclass of the Loss Function Wrapper. This class determines the cross-entropy loss between the actual labels and the anticipated labels. It assumes that there are only two label classes, 0 and 1 and that there is a single floating-point value per prediction. [2]

To build the training loop, we first need to record all the operations that will be performed during training. This includes getting the anchor and positive/negative image, getting the label, performing a forward pass through the model, calculating the loss, computing the gradients, updating the weights, and returning the loss. Once we have imported the metrics, we can start making predictions on the test data. This involves getting a batch of test data and passing it through the trained model to provide the anticipated results. We then post-process the results to convert the predictions into a format that can be compared against the ground truth labels. By comparing the predicted output to the ground truth labels, we can calculate the model's accuracy and other relevant metrics, including accuracy, recall, and F1 score. These metrics will give us a better understanding of how well the model is performing and help us identify any areas for improvement. Overall, the evaluation process is a critical step in any machine learning project, as it helps us understand how well the model is performing and identify areas for improvement. By carefully analysing the model's performance and making necessary adjustments, we can ensure that our model is accurate and efficient in completing the current assignment Following a review of the model's functionality, we can visualize the results to get a better understanding of how well the model is performing. This involves creating visualizations of the model's output, such as confusion matrices or ROC curves, to help us interpret the results the Siamese model predicts. [12]

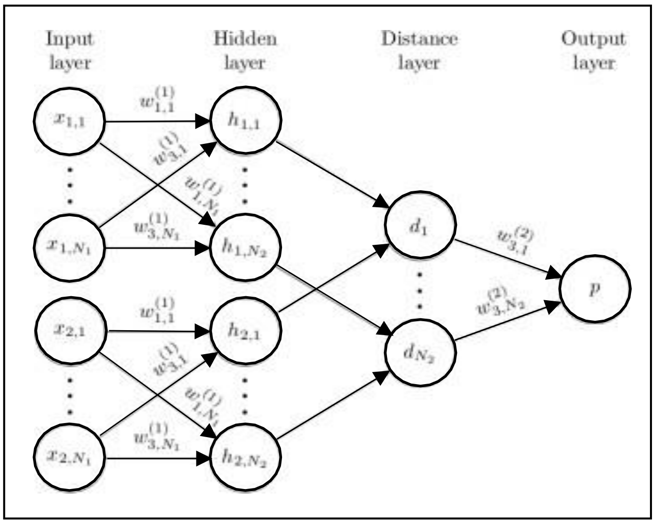

Figure no – 4.2.3 a simple Siamese network with two hidden layers for binary classification and logistic prediction. Twin networks are created by duplicating the network's structure between the top and bottom parts. These twin networks share weight matrices at every layer.

First, we build a results array that will hold the predictions made on the input image. We then make predictions on the input image using the trained model's weights.

Table 4.2.3 predictions with reloaded model

Siamese model. predict ([test_input, test_val])	array
	[3.9103380e-03]
	[3.8859820e-05]
	[4.5842552e-01]
	[3.5943764e-09]
	[9.8302263e-01]
	[9.4812006e-01]
	[1.1292892e-06]
	[2.0155067e-02]
	[2.8998731e-02]
	[9.3272382e-01]
	[9.7638303e-01]
	[1.8349889e-01]
	[2.5410500e-06]
	[2.1068140e-06]
	[4.2521237e-06]

Table 4.2.4 Siamese Model representing the unique features of each individual's face.

Layer (type)	Output Shape	Param
input_img (Input_Layer)	[(None, 100, 100, 3)	0
validation_img (Input_Layer)	[(None, 100, 100, 3)	0
embedding (Functional)	(None, 4096)	38960448
l1_dist_6 (L1Dist)	(None, 4096)	0
dense_7 (Dense)	(None, 1)	4097
Total params:	38,964,545	
Trainable params	38,964,545	

We set a detection threshold, which is the metric above which a prediction is considered positive. We also set a verification threshold, which is the proportion divides the all result of positive forecasts by the result of positive predictions.
Next, we use OpenCV to perform real-time verification of the model's performance. We create a verification trigger that listens for input images and passes them to the verification function for processing. We also save the input image to a designated folder for later reference. [10]

With real-time verification in place, we can continuously monitor the model's performance and ensure that it is accurately identifying relevant images. By setting appropriate thresholds and carefully analysing the results, we can ensure that the model is meeting the desired level of accuracy and effectiveness in real-world scenarios. The verification function is used to verify a person, with the detection threshold set to 0.95 and the verification threshold set to 0.6. The input image is captured from the webcam and added to the results array. Predictions are then made using the model, with a detection threshold A verification threshold is determined by dividing the percentage of accurate forecasts by the total number of accurate samples, over which a prediction is deemed positive. Based on the outcomes, the verification text is chosen, and the details are logged out.

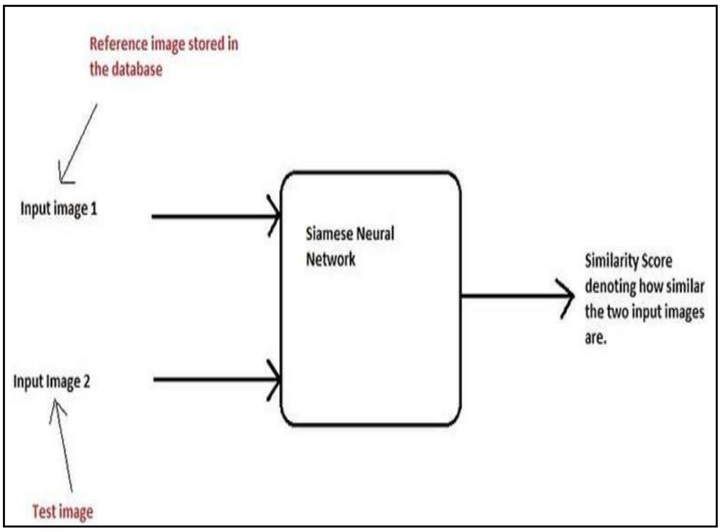

Figure no 4.2.4 - Schematic representation of Siamese neural network.

This network uses an additional reference image of the subject as input and generates a similarity score that indicates the likelihood that the two input photographs are of the same subject. A sigmoid function with a range between 0 and 1, where 1 denotes total similarity and 0 denotes no similarity, is frequently used to compress the similarity score. This is true for any number between 0 and 1 [4].

5. Approach of Study

Using a mobile application interface for user profile registration, we will implement a model for multimodal authentication in this paper. To do this experiment, we will an ESP 32 cam and an RS307 as the input hardware, which will process and save the input that was received from the hardware with the user's unique identifier. After setting up all user registration information, this will be stored inside the data storage cloud with the assistance of the ESP 32 camera

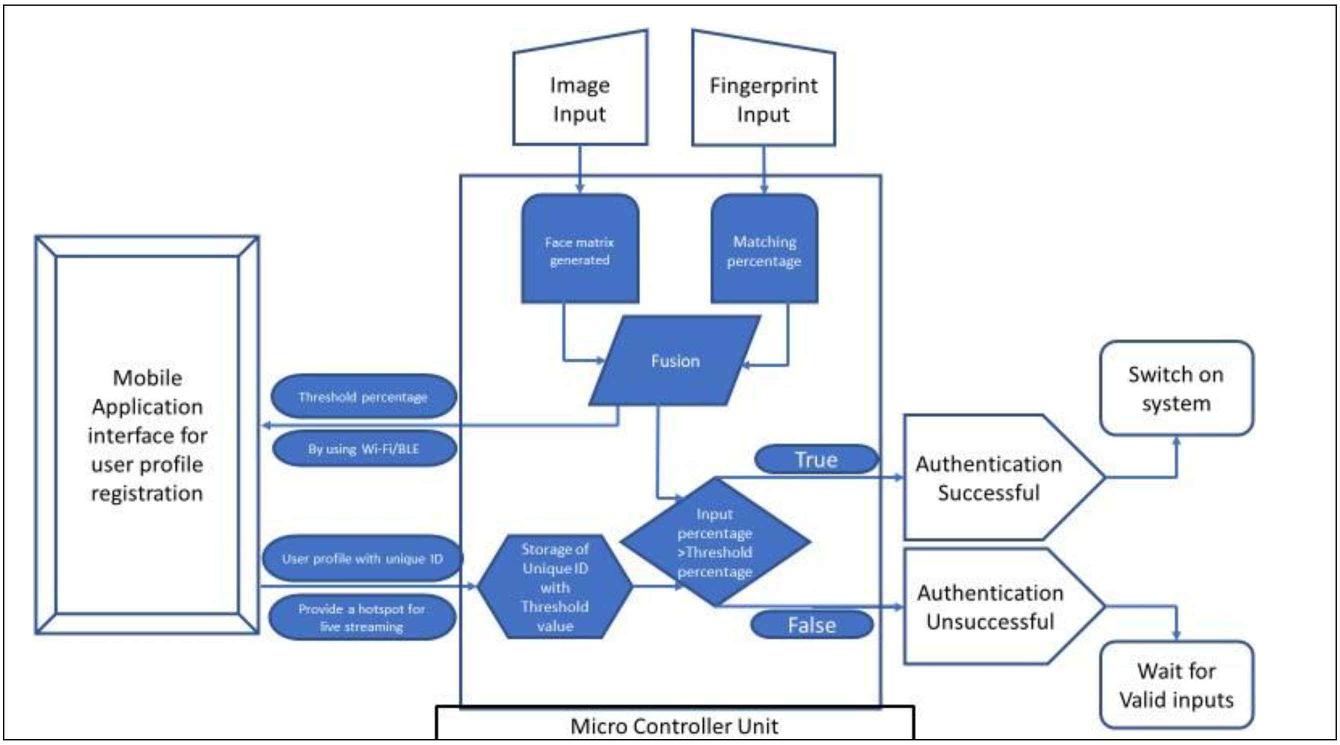

Figure no 5.1- Block Diagram representing the flow of the program

for image input and the RS 307 fingerprint module to obtain the fingerprint transmission of data can be done with the help of the mobile application by using the second IOT data link layer protocol. In the course of the authentication procedure, when stored data is compared to threshold values, our model will produce the appropriate outputs based on the features from the inputs that it has processed. The data signal is then delivered back to the controller so that it can respond to the inputs that it has received. If the authentication validation is completed successfully, a mechanical switch known as a relay will turn on the circuit. If the authentication is unsuccessful, however, the relay will continue to turn off the circuit while it waits for another affirmative input. [11]

In the various application areas, such as automotive, home security, and other security and privacy-related applications, we are able to make things safer and more reliable with a simple user interface by adopting this experimental technique.

6. Justification for components used

A low-cost research board called the ESP32-CAM has an ESP32-S chip with an OV2640 camera module that can broadcast video over Wi-Fi. The ESP32-CAM has the following streaming capabilities: [5] Video streaming: The ESP32-CAM can broadcast video at up to 320x240 pixels at 30 frames per second (fps) or 1600x1200 pixels at 15 fps. [5] HTTP server: The ESP32-CAM may serve as an HTTP server, enabling you to use a web browser to view the video stream and other capabilities. Access point mode: The board also functions as an access point, enabling you to stream video without a third-party network by connecting directly to it. Remote control: You may operate the ESP32-CAM remotely using the ESP-NOW protocol or using MQTT, allowing you to stream footage and control the camera from a distance. Motion detection: Because the board has built-in motion detection features, you can instruct it to start streaming video as soon as motion is detected.

A compact, inexpensive fingerprint sensor module called the "R307 Fingerprint Module" is available for use in biometric authentication and security applications. The R307 Fingerprint Module has the following characteristics: [6] High accuracy: The module contains an optical sensor with high resolution that can accurately capture fingerprints. The module is simple to use and straightforward to connect with microcontrollers and other devices thanks to its simple UART interface. Quick identification: The module can identify and match fingerprints in less than a second. The module has a large storage capacity and can store up to 1000 fingerprint templates in its onboard memory. Built-in algorithms: The module includes built-in algorithms for matching, extracting, and processing fingerprint image data.

Broad compatibility: The module works with many different platforms and development boards, including Arduino, Raspberry Pi, and others. The module has a very low power consumption architecture that makes it appropriate for battery-powered applications.

A way to store data on distant servers that can be accessed online is through cloud storage. These are some of the main attributes and advantages of cloud storage: Accessibility: Data stored in the cloud is accessible from any location with an internet connection, making it simple to share and collaborate on documents with co-workers, business partners, and clients. Scalability: Cloud storage can be readily scaled up or down to meet a company's storage requirements, allowing them to pay only for what they really use. Cost-effectiveness: Since cloud storage does not require expensive hardware or ongoing maintenance, it is frequently less expensive than traditional storage solutions.

Security: To prevent unauthorized access to data, cloud storage providers use a variety of security methods, such as data encryption, user authentication, and network security. Disaster recovery: To guarantee that data is protected in the event of natural catastrophes, power outages, or other unforeseen events, cloud storage can offer backup and disaster recovery solutions. Integration: To offer a comprehensive cloud-based solution, cloud storage can be connected with other cloud-based services like cloud computing and software-as-a-service (SaaS). Collaboration: Because cloud storage enables numerous users to access and update files in real-time, it is simple to collaborate on documents with clients and co-workers.

In conclusion, cloud storage is an adaptable, scalable, and affordable option for storing data that provides a number of features and advantages over conventional storage

7. Conclusion

The Siamese model and face recognition and identification on LFW datasets utilizes a CPU unit only are discussed in the paper along with several features of deep learning. We have described a method for verifying deep convolutional Siamese neural networks before conducting face recognition. We presented fresh findings by contrasting the performance of our networks with that of an existing cutting-edge classifier created for the LFW data set.

Our model performs far better than the majority of other models that are already in use and comes close to matching the best results obtained by the prior authors. On the LFW dataset, the computed accuracy is very nearly 80% after 50 epochs. By including more hidden layers, the accuracy of the training set may also be enhanced further. Additionally, this technology can be used for face recognition.

8. References

[1] Sounak Deya, Anjan Duttaa, J. Ignacio Toledoa, Suman K.Ghosha, Josep Llados´, "Convolutional Siamese Network for Writer Independent Offline Signature Verification", Umapada Palb aComputer Vision Center, Computer Science Dept., Universitat Aut`onoma de Barcelona, Edifici O, Campus UAB, 08193 Bellaterra, Spain b Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata-700108, India

[2] Gregory, Koch Richard, Zemel Ruslan, Salakhutdinov, "Siamese Neural Networks for One-shot Image Recognition" Department of Computer Science, University of Toronto. Toronto, Ontario, Canada.

[3] Emrah Basaran, Muhittin Gökmen, and Mustafa E. Kamasak, "An Efficient Multiscale Scheme Using Local Zernike Moments for Face Recognition", Published: 21 May 2018

[4] Chae Lin Seok 1, Young Do Song 1, Byeong Seon An 1 and Eui Chul Lee 2, "Photoplethysmogram Biometric Authentication Using a 1D Siamese Network", 1 Department of AI & Informatics, Graduate School, Sangmyung University, Hongjimun 2-Gil 20, Jongno-Gu, Seoul 03016, Republic of Korea; C.L.S.); Department of Human-Centered Artificial Intelligence, Sangmyung University, Hongjimun 2-Gil 20, Jongno-Gu, Seoul 03016, Republic of Korea

[5] Henry Dietz, Dillon Abney, Paul Eberhart, Nick Santini, William Davis, Elisabeth Wilson, and Michael McKenzie "ESP32-CAM as a programmable camera research platform", University of Kentucky; Lexington, Kentucky

[6] Deepak Kumar, Prof. (Dr.) Gurpreet Singh, Er. Ramandeep Kaur, "Implementation of Enhanced IOT Based Biometrics Attendance System using R307 Fingerprint Sensor with Arduino UNO and Real-Time Database to Improve Accuracy"

[7] Peng Lu, Baoye Song & Lin Xu, "Human face recognition based on convolutional neural network and augmented dataset" To cite this article (2021) pp-29-37, DOI: 10.1080/21642583.2020.1836526.

[8] P. Jonathon Phillips, "Support Vector Machines Applied to Face Recognition", National Institute of Standards and Technology Bldg 225/ Rm A216 Gaithersburg. MD 20899 Tel 301.975.5348; Fax 301.975.5287

[9] Hao Wu et a, "Face Recognition Based on Haar Like and Distance"-To cite this article: l 2021 J. Phys.: Conf. Ser. 1813 012036

[10] U. S. Rokade, D. Doye and M. Kokare, "Hand Gesture Recognition Using Object Based Key Frame Selection," 2009 International Conference on Digital Image Processing, Bangkok, Thailand, 2009, pp. 288-291, doi: 10.1109/ICDIP.2009.74.

[11] Rajeshree Rokade, Ketki Kshirsagar, Jayashree Sonawane, Sunita Munde, "Analysis of Human-Machine Interaction Through Facial Expression and Hand-Gesture Recognition", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-9S3, July 2019

[12] Ketki P. Kshirsagar, Rajeshree Rokade, Dhanapal Kamble, "IOT Based Gesture Recognition for Smart Controlling", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019.

[13] K. P. Kshirsagar and R. A. Shinde, "Comparing techniques of segmenting hand region," 2017 2nd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2017, pp. 721-724, doi: 10.1109/CESYS.2017.8321175.