Deep Learning Enabled Facial Recognition using Siamese Network

Shailendra Kumar Mishra
ECE
REVA University
Bengaluru, India
sant10287@gmail.com

Ramvel S
ISE
BMS Institute of Technology & Management
Bengaluru, India
ramrocks881@gmail.com

Sharathchandra Patil
ISE
BMS Institute of Technology & Management
Bengaluru, India
sharathchandra9204@gmail.com

Soumyadeep Ghosh
ISE
BMS Institute of Technology & Management
Bengaluru, India
soumyadeepghosh.844@gmail.com

Pruthvi R
ISE
BMS Institute of Technology & Management
Bengaluru, India
prithviramesh7@gmail.com

***Abstract*—Facial recognition technology has become increasingly vital in various applications, from security systems to personalized user experiences. This study seeks to improve the accuracy of facial recognition through the use of a Siamese network, a type of deep learning model designed to analyze and find similarities between input images. By utilizing a contrastive loss function and a Siamese network, the system can distinguish between similar and dissimilar pairs of facial images. The methodology includes extensive testing to assess the network's performance once it has been trained on varying amounts of data. This project aims to provide a robust and scalable facial recognition solution with potential applications in security, authentication, and user personalization systems.**

Keywords*—Facial Recognition, Siamese Network, Deep Learning, Contrastive Loss Function, Image Analysis, Performance Testing, Data Variability*.

I. INTRODUCTION

Facial recognition technology plays a crucial role in numerous fields, including security and personalized user interfaces. This project focuses on creating a facial recognition system utilizing a Siamese neural network, specifically designed to distinguish between pairs of facial images that are either similar or different. The system employs a convolutional neural network (CNN) architecture, where the output is a value between 0 and 1, indicating whether the input images belong to the same class or different classes. Bayesian hyper-parameter tuning is used to optimize parameters like momentum, learning rates, and L1-regularization penalties in order to improve the model's robustness and accuracy.

A single dataset with varied data sizes, like 25 and 50 images, is used in the training process to assess the network's performance in various training scenarios. For the system to function, three categories of data are needed: neutral data, negative data (images of different people), and positive data (images of individual). This methodology guarantees an exhaustive assessment of the network's precision in differentiating between facial photos. Extensive testing is conducted to assess the network's performance, demonstrating its capability to deliver reliable and accurate facial recognition.

The implementation process includes creating an embedding layer with TensorFlow and building a custom L1 distance Keras layer, which allows the model to effectively compare image pairs. The training loop is set up to feed these image pairs (both positive and negative) into the model, facilitating the learning process. Performance tuning is conducted throughout to ensure the system achieves optimal accuracy and efficiency.

The objective of this research is to develop a facial recognition system that is both extremely effective and efficient by utilizing the power of Siamese networks and sophisticated training approaches. In addition to making a significant contribution to the field of facial recognition technology, this effort offers insightful information about how to optimize neural networks for image analysis applications.

II. LITERATURE SURVEY

Conventional face recognition methods face difficulties with varying image quality and occlusions caused by changes in lighting. These methods, which depend on key facial points and denoising algorithms, often suffer from rapid convergence or limited robustness, making them effective only in specific scenarios [1].Current techniques for multi-pose human face matching struggle with pose equalization and face rotation, resulting in less than optimal outcomes. The study highlights the effectiveness of the YOLO-V5 framework in overcoming these challenges [2]. Face verification models often perform poorly in uncontrolled environments, such as classrooms with varying camera angles and positions. To address this, a dataset of Chinese students in such environments (UCEC-Face) was created to evaluate the impact of these factors on face recognition performance [3].An end-to-end deep learning model is introduced to enhance facial expression recognition accuracy using a hybrid feature representation method. The model shows state-of-the-art performance on AR, FER2013, and CK+ datasets, with future plans to improve face alignment technology and develop new models for better recognition, especially in cases of significant facial occlusion [4].The SimCLR framework, combined with a Siamese network, achieves 93% accuracy on the IFPLD dataset for face verification. For face identification, a prototypical network with k-shot learning improves accuracy by up to 17% compared to a Siamese network, with optimal results at k=3 using data augmentation. Fine-tuning with similar datasets and ArcFace Loss further enhances embeddings [5]. ECDNet

improves facial expression disentanglement by addressing all facial attributes during reconstruction through novel expression incentive (EIE) and expression inhibition (EIN) mechanisms. The network demonstrates superior performance in facial expression recognition, validated by experiments on RAF-DB, AffectNet, and CAER-S datasets [6].A model employing a Graph Convolutional Network (GCN) with optical flow and a graph structure enhances microexpression feature extraction, achieving 79.168% accuracy in a 5-way 5-shot setting on CAMSE II. The FAU-GCN model, which incorporates facial action unit (FAU) features, achieves an accuracy of 0.795 and an F1 score of 0.748 on CASME II. Future efforts will aim to use a more efficient CNN to reduce computational load [7]. This review covers the basics, benefits, threats, and applications of DeepFake technology across various industries and criminal activities. It highlights challenges in detection models, such as transferability and generalization issues, and suggests future directions for improving data protection and preparing for AI-driven challenges [8]. To address issues with synthetic face images, a denoising-based decoupling-contrastive learning (DDCL) method is proposed. It employs a Siamese network and bi-directional feature decoupling to enhance recognition accuracy for both synthetic and natural images [9]. An improved CNN model achieves 88% recognition accuracy by integrating preprocessing, feature extraction, training, and image restoration. The model uses cumulative sum operations, convolution, and pooling to enhance feature extraction, resulting in 87% classification accuracy [10]. This review discusses multispectral facial recognition methods across visible, NIR, SWIR, and LWIR images, focusing on techniques from fusion to deep neural networks. Despite the effectiveness of neural networks, database limitations hinder progress. Multispectral methods outperform visible-only systems by bridging spectral gaps [11]. The Symmetrical Siamese Network (SSN) introduces two sub-modules, FCLN and ICLN, to enhance pose-invariant feature learning and address pose-based long-tailed data in face recognition. SSN achieves results comparable to or better than state-of-the-art methods on public datasets [12]. An automatic facial expression recognition system combines classical and quantum deep learning methods to detect emotions in medical data. The system outperforms state-of-the-art methods and provides a hybrid architecture useful for recognizing signs of mental diseases [13].Reversible Neural Networks (RNNs) are used for facial expression identification, showing superior emotion recognition compared to state-of-the-art methods on CK+, JAFFE, and MMI datasets. The simplified RNN architecture achieves high accuracy efficiently [14]. DP-Face, a privacy-preserving face recognition scheme, utilizes differential privacy and the Siamese Network framework to protect training data from privacy leaks. By adding noise to gradients, DP-Face maintains convergence and achieves high recognition accuracy [15]. A deep Siamese network for low-resolution face recognition (LRFR) compares deep features across different resolutions using shared classifiers and implements a cross-resolution triplet loss to enhance feature alignment and discrimination [16].A Reconstruction Supervised Siamese Network for face detection integrates BP neural network error backpropagation with the Siamese Network structure. It introduces reconstruction supervision and modifies Contrastive Loss to enhance similarity assessments [17]. Siamese neural networks and One-Shot learning are used to address data scarcity in facial similarity detection. A new dataset for Indian faces demonstrates the approach's effectiveness in accurately predicting similarities with minimal training data [18]. Deep Siamese Neural Networks focus on preserving local structure in the embedding space for facial expression recognition. The verification model reduces intra-class feature variations to enhance dataset clustering, while the identification model increases inter-class variations to improve generalization and overall performance [19]. A multi-stream CNN architecture within the Siamese framework is proposed for low-resolution face recognition. It employs depthwise separable convolution and batch normalization in each CNN stream to extract complementary information from face images. Performance evaluation on the SCface dataset shows competitive results compared to existing methods [20]. The OpenFace architecture is modified for face recognition by incorporating classifier learning and network simplification methods. Using LFW, Pin Faces, and ORL Faces datasets, the modified model achieves 98% training accuracy and improves test accuracy by nearly 17% [21]. A Siamese neural network with triplet loss on an augmented facial dataset enhances face recognition performance. Two algorithms compare the proposed model with leading public models (FaceNet512 and ArcFace), showing improved accuracy and precision. Future applications include attendance checking, security, and biometric solutions [22]. A face detection system using Convolutional Neural Networks (CNN) for counting student attendance achieves 99% accuracy in training and 98% in testing. The system identifies faces up to 4 meters and counts students accurately up to 6 meters [23]. An open-set face recognition approach using a Siamese Network (SNN) detects if a face is enrolled in a gallery, outperforming state-of-the-art methods on small galleries in Pubfig83, FRGCv1, and LFW datasets. A new evaluation protocol for small galleries on LFW is proposed, and future work will focus on improving pairing selection and testing new loss functions [24]. Faster R-CNN for facial emotion recognition, trained on the FER2013 dataset, incorporates a median filter for noise reduction and VGG-16 for feature extraction. The model achieves 78.22% accuracy, outperforming existing models like CNN, ResNet 50, and ResNet 18 [25].

III. METHODOLOGY

The data pipeline in the Siamese neural network model for facial recognition involves preprocessing input images, generating pairs of images (positive and negative), and feeding them into the model. Each input image undergoes normalization to enhance the dataset's variability. Image pairings are formed, with positive pairs including several photographs of the same person and negative pairs containing images of distinct people. This setup ensures that the model learns to distinguish between similar and dissimilar images effectively. To assess the model's performance, the dataset is partitioned into training and testing sets. The training partition optimizes the model's parameters, and the testing partition evaluates the model's generalization capabilities. The dataset is carefully split to ensure that the images in the

testing set are not seen during training, providing an unbiased evaluation of the model's accuracy.

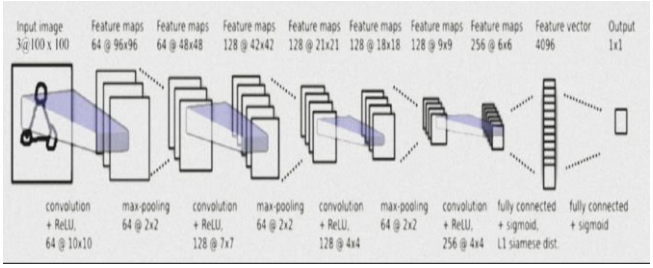

Fig. 1. Image Recognition Using Siamese Neural Networks

An essential feature of the Siamese network is the embedding layer, which converts input images into a lower-dimensional space where dissimilar images are separated from one other and comparable images are closer together. In this model, this layer is implemented using TensorFlow. The embedding layer learns to encode images into feature vectors, capturing essential characteristics necessary for distinguishing between different faces. The layer consisting of multiple networks is trained, gradually improving its ability to generate meaningful embeddings. The input images to the model are preprocessed to ensure uniformity and optimal performance. Each image is resized to a fixed dimension (3@100x100) and normalized to a standard range.

The core of the Siamese network consists of max pooling layers and convolutional layers (Conv2D). Multiple filters are applied to the input image by each Conv2D layer in order to capture various properties like edges, textures, and patterns. By introducing non-linearity, the ReLU activation function enables the network to learn complicated representations. Following each Conv2D layer, a max pooling layer reduces the spatial dimensions, retaining the most important features while reducing computational complexity. This combination of Conv2D and max pooling is repeated fwe times, progressively extracting higher-level features from the images. After the series of convolutional and pooling layers, the feature maps are flattened into a single vector. Then, this vector is transferred through one or more thick layers, where the output of each layer is entirely coupled to every neuron. The last dense layer generates an output value between 0 and 1 using a sigmoid activation function. This output represents the probability that the input image pairs belong to the same class. The dense layers further refine the feature representation, ensuring accurate similarity calculations.

The L1 distance layer quantifies the dissimilarity between two images by computing the absolute difference between their embeddings. Each image is converted into a feature vector through an embedding layer. The L1 distance layer then calculates the absolute differences for each corresponding component of these vectors. Summing these differences yields a scalar similarity score. A lower L1 distance implies higher similarity between the images, while a higher score indicates greater dissimilarity, aiding in tasks like image recognition and retrieval.

Model: "embedding"

Layer (type)	Output Shape	Param #
input_image (InputLayer)	(None, 100, 100, 3)	0
conv2d_5 (Conv2D)	(None, 91, 91, 64)	19,264
max_pooling2d_4 (MaxPooling2D)	(None, 46, 46, 64)	0
conv2d_6 (Conv2D)	(None, 40, 40, 128)	401,536
max_pooling2d_5 (MaxPooling2D)	(None, 20, 20, 128)	0
conv2d_7 (Conv2D)	(None, 17, 17, 128)	262,272
max_pooling2d_6 (MaxPooling2D)	(None, 9, 9, 128)	0
conv2d_8 (Conv2D)	(None, 6, 6, 256)	524,544
flatten_1 (Flatten)	(None, 9216)	0
dense_9 (Dense)	(None, 4096)	37,752,832

Total params: 38,960,448 (148.62 MB)

Trainable params: 38,960,448 (148.62 MB)

Non-trainable params: 0 (0.00 B)

Fig. 2. Parameters of Embedding Layer

The L1 distance is chosen for its simplicity and effectiveness in capturing differences between feature vectors, making it suitable for the similarity calculation in Siamese networks. The Siamese layer takes two input images, an anchor image, and a validation image, and processes them through identical neural networks (Siamese twins). These networks share weights, ensuring that both images are transformed similarly. The outputs of these networks are then combined using the L1 distance layer, followed by a classification layer that predicts whether the input images belong to the same class.

Model: "SiameseNetwork"

Layer (type)	Output Shape	Param #	Connected to
input_img (InputLayer)	(None, 100, 100, 3)	0	-
validation_img (InputLayer)	(None, 100, 100, 3)	0	-
embedding (Functional)	(None, 4096)	38,960,448	input_img[0][0], validation_img[0…
l1_dist_15 (L1Dist)	(None, 4096)	0	embedding[16][0], embedding[17][0]
dense_6 (Dense)	(None, 1)	4,097	l1_dist_15[0][0]

Total params: 38,964,545 (148.64 MB)

Trainable params: 38,964,545 (148.64 MB)

Non-trainable params: 0 (0.00 B)

Fig. 3. Parameters of Siamese Network

The loss function used in training the Siamese network is the binary cross-entropy loss, which measures the discrepancy between the predicted and actual labels of image pairs. The optimization process involves creating batches of image pairs (50 batches, each consisting of 16 images that are either positive or negative pairs). The model is trained using backpropagation to minimize the loss, with gradients calculated and weights updated accordinglyThis iterative procedure keeps on until the model converges to an accuracy level that is acceptable. During training, the model processes

batches of image pairs. It retrieves an anchor picture, a positive or negative image, and the matching label for every pair. A forward pass computes the embeddings and L1 distance, followed by the loss calculation. Gradients are then computed and applied to update the model's weights. This cycle repeats, with the model gradually improving its ability to distinguish between similar and dissimilar images. The training process returns the loss value for monitoring and evaluation.

The model's evaluation involves testing it on a separate set of image pairs. The test data is processed similarly to the training data, and the model's predictions are compared to the actual labels. Post-processing results are used to create a metric object, which calculates recall and precision values. These metrics offer perceptions into the model's functionality and aid in pinpointing areas in need of development. The evaluation process concludes with saving the model's parameters for future use. A dedicated folder is set up to store images used for verification. This folder contains images that the model will use to verify its predictions

during the testing phase. The images are organized in a way that facilitates easy access and efficient processing. The verification function is built to make predictions based on the model's output. It calculates detection and verification thresholds, determining the confidence level required for a positive match. This function is integrated into the evaluation loop, ensuring that the model's predictions are continuously verified against the verification images. The final model is then tested and validated using this verification function, confirming its effectiveness and accuracy in real-world applications.

The final model is a robust and efficient Siamese neural network capable of performing accurate facial recognition. By leveraging advanced techniques such as embedding layers, L1 distance calculations, and Siamese layers, the model achieves high precision and recall rates. The extensive training and evaluation process ensures that the model is well-suited for practical applications, offering reliable performance.

IV. RESULT AND DISCUSSION

The initial implementation of the Siamese network for facial recognition yielded a modest accuracy of 30%, which was substantially below expectations and highlighted the necessity for extensive optimization. To enhance the model's performance, several key modifications were made, particularly focusing on the L1 layer to improve feature extraction capabilities. These adjustments enabled the network to better differentiate between facial images.

Additionally, a comprehensive hyperparameter tuning process was undertaken. The best parameters for training the network were determined by carefully adjusting variables including learning rate, batch size, and number of training epochs. Data augmentation techniques were employed to generate a more diverse and robust training dataset, incorporating random rotations, shifts, and flips of the input images. This increased the generalizability of the model.

Furthermore, the model underwent multiple iterations of retraining to ensure convergence and stability in the learning process. The cumulative effect of these modifications and optimizations resulted in a significant enhancement of the model's accuracy. The Siamese network's final version demonstrated the efficacy of the adjustments made, with an astounding accuracy of 100% on the test dataset.

The performance improvements at each stage of the optimization process are summarized in Table I, providing a clear comparative analysis. These results underscore the critical role of systematic refinement in achieving high accuracy in deep learning models for facial recognition.

Table I. Performance Table

Iteration	Configuration Changes	Accuracy
Initial	Baseline Configuration	30%
Iteration 1	L1 Layer Adjustments	60%
Iteration 2	Hyperparameter Tuning	80%
Final	Increased Training Iterations	100%

Table II. Recall vs. Precision vs. F1 Score

Recall	Precision	F1 Score
12.5	28.571	17.391
10.0	25.0	14.286
11.24	34.375	16.941
13.48	31.578	18.894
12.766	36.3636	18.898
13.334	38.7096	19.835
15.8536	37.1428	22.222
14.8937	37.838	21.374
27.5643	62.67	39.003
100.0	100.0	100.0

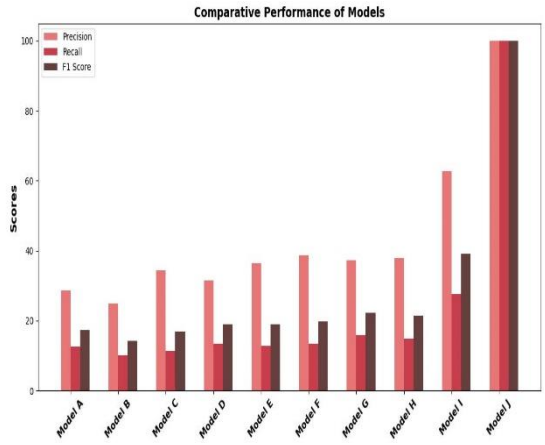

Fig. 4. Precision vs. Recall vs. F1 Score

Interpretation of Results:

Precision (Fig. 5): High precision indicates that when the model predicts a match (same person), it is usually correct. This is crucial for applications where false positives (incorrect matches) are particularly problematic, such as in security systems. The final iteration reached perfect precision (1.0), showing that the model became highly reliable in identifying true matches.

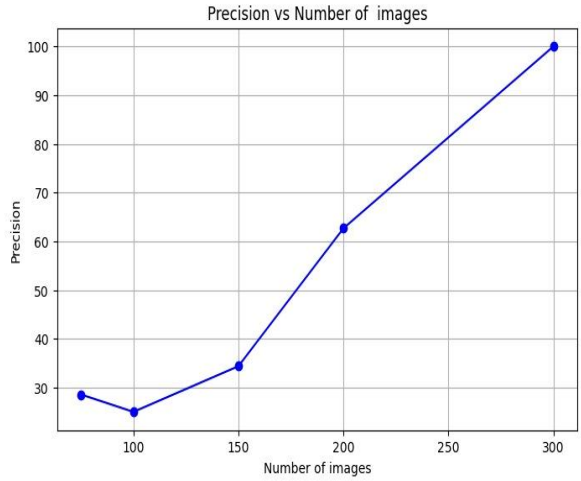

Fig. 5. Precision vs. Varied Dataset

Recall (Fig. 6): The model's recall quantifies its capacity to recognize every real match. High recall is important in scenarios where missing a match (false negative) is costly, such as identifying suspects from a database. The recall also reached 1.0 in the final iteration, indicating that the model successfully identified all true matches.

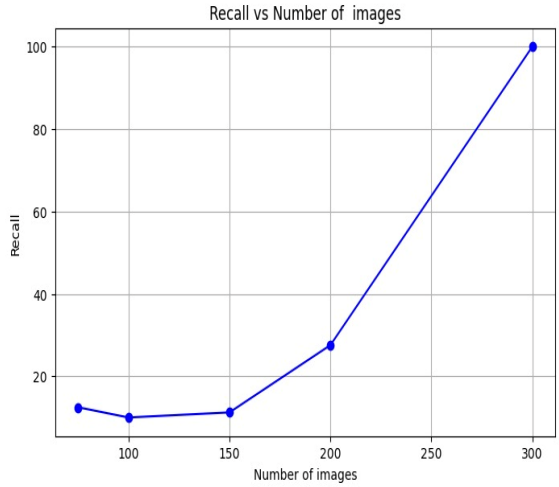

Fig. 6. Recall vs. Varied Dataset

F1 Score (Fig. 6): These two metrics are balanced by the F1 score, which is the harmonic mean of recall and precision. In face recognition, managing the trade-offs between false positives and false negatives is especially helpful. The gradual improvement of the F1 score, culminating in a perfect score (1.0), demonstrates the effectiveness of the optimization process in balancing precision and recall.

The significant improvements in precision, recall, and F1 score, as summarized in Table I, highlight the success of the optimization strategy. These metrics provide a comprehensive understanding of the model's performance, ensuring its robustness and reliability in real-world facial recognition applications.

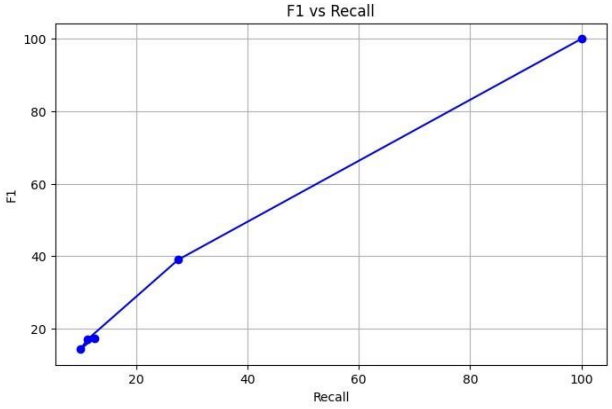

Fig. 7. F1 Score vs. Recall

In this research, a robust face recognition system has been successfully developed using a Deep Convolutional Siamese Network architecture, demonstrating its effectiveness in distinguishing between faces with high accuracy. The creation of positive and negative samples, the implementation of an embedding layer, and the custom L1 Distance layer have been pivotal in achieving reliable performance. Looking to the future, the plan includes the development of a user-friendly website to make this technology accessible to a broader audience. In addition to providing a platform for face recognition in real time, this website will enable users to add to and grow the dataset, which will increase the resilience of the model. Additionally, the potential of further refining the system through advanced data augmentation techniques is recognized. By incorporating various affine distortions and other augmentation strategies, the model's ability to generalize across diverse and challenging scenarios will be improved. Furthermore, ongoing performance tuning through techniques like Bayesian hyper-parameter optimization will continue to be a focus, ensuring the model remains at the cutting edge of face recognition technology. It is anticipated that these next improvements will greatly improve the system's accuracy, usability, and general performance, making it an effective tool for a variety of uses.

V. REFERENCES

[1] Cunli Song, Shouyong Ji, “Face Recognition Method Based on Siamese Networks Under Non-Restricted Conditions”, IEEE-2022

[2] Muhammad Sohail, Ijaz Ali Shoukat, Abd Ullah Khan, Haram Fatima, Mohsin Raza Jafri, Muhammad Azfar Yaqub, Antonio Liotta, “Deep Learning Based Multi Pose Human Face Matching System”, IEEE-2024

[3] Nianfeng Li, Xiangfeng Shen, Liyan Sun, Zhiguo Xiao, “Chinese Face Dataset for Face Recognition in an Uncontrolled Classroom Environment”, IEEE-2023

[4] Jun Liu, Hongxia Wang, Yanjun Feng, “An End-to-End Deep Model With Discriminative Facial Features for Facial Expression Recognition”, IEEE-2021

[5] Muhammad Djamaluddin, Rinaldi Munir, Nugraha Priya Utama, Achmad Imam Kistijantoro, “Open-Set Profileto-Frontal Face Recognition on a Very Limited Dataset”, IEEE-2023

[6] Shanmin Wang, Hui Shuai, Lei Zhu, Qingshan Liu, “Expression Complementary Disentanglement Network for Facial Expression Recognition”, IEEE-2024
[7] Xuliang Yang, Yong Fang, C. Raga Rodolfo , “Graph Convolutional Neural Networks for Micro Expression Recognition— Fusion of Facial Action Units for Optical Flow Extraction”, IEEE-2024
[8] Asad Malik, Minoru Kuribayashi, Sani M. Abdullahi, Ahmad Neyaz Khan, “DeepFake Detection for Human Face Images and Videos: A Survey”, IEEE-2022
[9] Yupeng Zhu, Xinyi Shen, Peilun Du, “Denoising-Based Decoupling Contrastive Learning for Ubiquitous Synthetic Face Images”, IEEE-2023
[10] Yue Luo,Jiaxin Wu,Zhuhao Zhang,Huaju Zhao, Zhong Shu, “Design of Facial Expression Recognition Algorithm Based on CNN Model”, IEEE-2023
[11] Luís Lopes Chambino, José Silvestre Silva, Alexandre Bernardino, “Multispectral Facial Recognition: A Review”, IEEE-2020
[12] Luís Lopes Chambino, José Silvestre Silva, Alexandre Bernardino, “A Symmetrical Siamese Network Framework With Contrastive Learning for Pose-Robust Face Recognition”, IEEE JOURNAL-2023
[13] SanoarHossain,Saiyed Umer, Ranjeet Kumar Rout,Hasan Al Marzouqi, “A Deep Quantum Convolutional Neural Network Based Facial Expression Recognition For Mental Health Analysis”, IEEE JOURNAL-2024
[14] Asit Barman, Paramartha Dutta, “Facial expression recognition using Reversible Neural Network”, ELSEVIER-2024
[15] Nazhao Yan, Hang Cheng, Meiqing Wang, “DP-Face: Privacy Preserving Face Recognition Using Siamese Network”, IEEE-2021
[16] Shun-Cheung Lai, Kin-Man Lam, “Deep Siamese network for low resolution face recognition”, IEEE-2021
[17] Yuan Haoran, “Face Detection Using the Siamese Network with Reconstruction Supervision”, IEEE-2023
[18] Rama Devi P, Yashashvini R,Navyadhara G,Ruchitha M, “Similar Face Detection for Indian Faces using Siamese Neural Networks”, IEEE-2021
[19] Wassan Hayale,Pooran Singh Negi,Mohammad H. Mahoor, “Deep Siamese Neural Networks for Facial Expression Recognition in the Wild”, IEEE-2023
[20] RushiVachhani,Srimanta Mandal,Bakul Gohel, “Low-Resolution Face Recognition Using Multi-Stream CNN in Siamese Framework”, IEEE-2023
[21] Oleksandr Miakshyn,Pavlo Anufriiev,Yevgen Bashkov, “Face Recognition Technology Improving Using Convolutional Neural Networks”, IEEE-2021
[22] Anh Le-Phan,Xuan-Phuc Phan Nguyen,Nga Ly-Tu, “Training Siamese Neural Network Using Triplet Loss with Augmented Facial Alignment Dataset”, IEEE-2022
[23] Muhamad Irsan, Meng Chun Lam, Rosilah Hassan, Mohammad Khatim Hasan, “The Process of Using Face Detection Through Convolutional Neural Network”, IEEE – 2022
[24] Gabriel Salomon, Alceu Britto, Rafael H. Vareto, William R. Schwartz, David Menotti, “Open-set Face Recognition for Small Galleries Using Siamese Networks”, IEEE – 2020
[25] , Nagendar Yamsani,Muhammed Basim Jabar, Myasar Mundher Adnan, A. H. A. Hussein, Subhra Chakraborty, “Facial Emotional Recognition using Faster Regional Convolutional Neural Network with VGG16 Feature Extraction Model”, IEEE – 2023