Multi-view Face Recognition Using Novel Convolutional Neural Network-based Deep Learning Architecture

Bharath J[1], Dr R Balakrishna[2]

[1]*Research Scholar, Department of CSE, Rajarajeshwari College of Engineering, Bengaluru.*

2 Research Supervisor, Department of CSE, Rajarajeshwari College of Engineering, Bengaluru.

[1]bharath5911@gmail.com, [2]rayankibala@gmail.com

Abstract:

Face recognition is an important research arena in the domain of human-computer interaction. For multi-view face recognition, algorithms resulting in more accurate outputs often pose a challenging bar. In this paper, a novel convolutional neural network (CNN) has been proposed and applied to multi-view facial recognition. In this proposed method, a large dataset of images is pre-processed, and raw data is altered as features. Different angles of the same image have been analysed with a deep learning model to overcome the clarity issues in the case of a multi-view domain. A CNN algorithm was applied to train the model with this complex dataset to gain an optimal output that has higher accuracy. With this applied dense layer of CNN model- training, the favourable features well suited to classify the data into separate categories were learned. The novel architecture of the model was developed using eight convolutional layers, four max-pooling layers, and a dropout layer. The Adam optimizer was used for the optimization of the performance of the proposed model. This novel CNN approach executed lower false positives and false negative outputs, yielding a high accuracy of this model. Hence, the proposed model can be used for efficient multi-view face recognition.

Keywords: *Convolutional neural network (CNN), multi-view face recognition, false positive, false negative, confusion matrix*

DOI Number: 10.14704/nq.2022.20.13.NQ88234 Neuro Quantology 2022; 20(13):1889-1897

I. INTRODUCTION

One of the ubiquitous things in this modern era of smart techs is facial recognition (FR). From various social networking sites to different types of cloud storage, biometric scans, and control systems monitoring, there is a rapid increase in demand for better facial recognition methods [1] [2]. FR aims to track and localize different human faces using captured or stored images [3]. This is a part of the broader domain of Man-Machine Interaction (MMI) which is getting more challenging in order to meet the requirements of user experience and accuracy. Face recognition is also a big part of face modelling techniques which is also growing parallelly with the advancement of technology [4]. Due to all of these, there is a high demand for systematic research on the improvement of techniques to recognize the face in efficient ways. With the fast development of computation and image processing technologies, FR has remarkably progressed in the last few decades. For real-world applications, advanced artificial intelligence and deep learning algorithms are used to get higher accuracy. The objective of such approaches is to simulate a similar learning process by human cognition and neurons to learn the patterns and rules for large image datasets [5]. These algorithms not only help out in getting successive results for image detection but also in other scientific and industrial application research. More research in different areas is needed to get better results that are more accurate and outputs that have more clarity. The current techniques face more issues in the case of real-time implementation. That is why various algorithm performances have yet to reach a satisfactory level because of the variance of the result. More precisely, actual applications of face angles, expressions, lighting effects, and

aging lead to image clarity issues as the face regions shrink [6]. The reasons behind the image clarity issue have led the multi-view face detection approaches to solve the problems. These approaches include cascade-based strategies, deformable part model techniques, and neural network-based algorithms [7]. The deep neural network methods had significantly given better results when more datasets were trained for different kinds of features and were then combined [8]. They have given better accuracy for face recognition, classification, and detection.

When a large dataset is used as an input for multi-view face recognition, often it has been seen that the implemented algorithm does not reach a satisfactory number of correct recognition of images. The accuracy of a model depends upon the number of false negatives and false positives present in the outcomes. They are also called type I and type II errors. "False alarm" is another term for a "false positive." If you say something is false, while it is the case of an actual sample, you have committed a false negative. This is known as a false negative when something is not discovered when it is present. Reducing the number of these two errors can lead to better model accuracy for face recognition. In this study, a novel CNN algorithm was implemented to recognize 181 different views of a single individual. The proposed study outcomes show a lower number of these two errors, indicating an accuracy of 99.06%, far better than that of state-of-the-art techniques.

II. LITERATURE REVIEW

The appearance of image inputs has always played a significant role in the case of facial recognition algorithms because of the vast number of variances. There may be issues with shadow and lighting. There can be contrast problems. The face angle present in the image can lead to different clarity issues. As a solution to these problems, the image channel is extended to diverse types, like gradient histograms and gradient magnitudes which simplifies the enriched data [9]. After simplifying, aggregate channel features are applied to get a better performing model for multi-view face recognition. From estimation of pose to expression analysis in images, annotation is essential to detect the facial landmarks, i.e., coordinates of the labelled parts. Ma *et al.* (2018) applied a multi-task cascaded CNN to find the target even in complex environments [10]. In this case, the frontal face detector is trained initially using a multi-task CNN algorithm. Then the non-frontal face detection is carried out. However, it needs a huge amount of data to be trained to reduce the fault in target calculation due to large-angle issues for face recognition. After both frontal and non-frontal face detections are completed, the results are concatenated to get the multi-view face recognition. When face recognition applications are used on social networking sites, users try to search for 'images with a friend. This query is quite arduous. The reason is the absence of definite signals correlated to indicate person's identity in the photos. Farfade et al. (2015) use a single model to get a broad range of orientations in a photo based on deep convolutional neural network layers, i.e., Deep Dense Face Detector [6]. Using an innovative face alignment approach, Zhao *et al.* (2020) developed a new framework for extensively embedding facial images [11]. Following this, they used principal component analysis (PCA) to reduce the attributes' dimensionality while eliminating the excessive and corrupted visual characteristics. The authors then suggested a shared Bayesian method for assessing feature vector convergence. The state-of-the-art facial recognition system has an accuracy of 98.52%. Ravidas *et al.* (2019) devised a deep-learning approach using several frameworks to improve the recognition of faces from multiple viewpoints [12]. They developed CNNs using a cascading structure that swiftly discards areas that are not facial-regioms. In order to identify different faces, a probability-based assessment of similarity among the face photos was carried out using a Bayesian approach. Deshmukh *et al.* (2020) presented a method based on a deep CNN for enhancing the clarity of faces in low-quality films [13]. The input picture recognised by the Viola-Jones technique as well as the scaling parameters is then supplied into the CNN model. The outcomes obtained from this approach are then incorporated using a fuzzy model, leading to improved face image resolution.

Various algorithms have been used to extract facial features in recent years. Some of these approaches are – histogram gradients, local face quantization, binary patterns, Gaussian regression, etc. Local binary patterns were used for the features of multi-view detection leading to a better classification model. The scale-invariant feature transform (SIFT) method was used by Zheng *et al.* (2014), in which distinct local feature points were extracted to get a higher classification accuracy [14]. In some approaches, like Zhang *et al.* (2016), SIFT was combined with a deep neural network-guided algorithm to learn and classify prominent features for optimal output of face expression detection [15]. Different datasets with frontal and non-frontal data were used to train the deep learning model to deal with the multi-view inputs. CNNs have outperformed other algorithms given that they are trained with more epochs and with larger training datasets and richer features to aim relations. This can be a better approach with the add-on of Bidirectional Long Short-Term Memory (BLSTM). In previous approaches to facial action detections, fewer facial views were used. When the no. of different facial views was increased, the algorithms needed modifications to deal with them. Here comes BLSTM-RNN. In the work of He *et al.* (2017), many CNN layers were used to recognize the facial action units when the given datasets were complex [16]. Using a large number of CNN models helps in extracting the required features to pin down those action units. In this study, a CNN-based novel deep learning method was tweaked in such a way that it upgraded the accuracy of identifying different views of a person. The datasets were minutely pre-processed to get better feature-target correlations for the algorithm. Data visualization took these pre-processed images as training datasets to get the patterns and relations among the features from multi-view aspects so that the algorithm can perform better during implementation. Here, a multi-view perspective was important; otherwise, the issues with image clarity might mess up the training, causing weaker outputs. Further, these features were chosen or altered in such a way that they made the dataset reliable and strong feature-target correlation, which is necessary to implement deep neural network layers. The convolutional neural network worked as a critical feature selector and thus led the model to better training for classification.

As a solution to issues like position fluctuation, lighting variability, and diffraction concerns in the facial images, Bhangale *et al.* (2021) provided a deep learning model for face recognition built on VGG16 [17]. The suggested solution employs the Multi-Task CNN for facial feature identification and the VGGNet for facial feature interpretation. Jie *et al.* (2020) suggested a multi-view face-expression recognition network (MVFE-LightNet) to solve the problem of recognising facial expressions from several angles [18]. This was accomplished by employing MTCNN for face detection and aligning prior to any pre-treatment, such as normalisation or up sampling. To limit the possibility of overloading and decrease the number of design variables, a depth wise differentiated residue CNN was developed.

III. METHODOLOGY

The current study discusses multi-view face recognition, which is capable of recognizing numerous perspectives of a single individual. For effective multi-view facial recognition, the methodology offers a strong picture pre-processing, feature engineering, and a novel CNN-based deep learning architecture. Data gathering, multiplex data visualization for better comprehension of the data, feature engineering, pre-eminent data production, and building of a deep learning architecture comprised the stages of the technique (Fig. 1). The algorithm (presented in detail in Fig. 2) was also fine-tuned to improve accuracy in detecting several views of a single person. Numerous metrics were also used to evaluate and test the entire deep learning process.

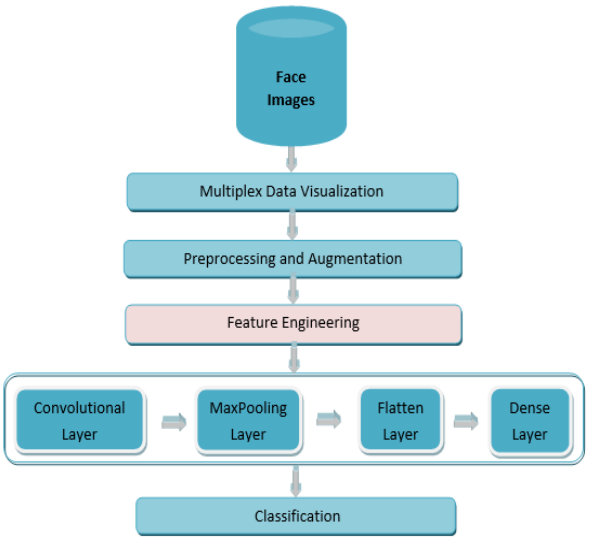

Fig. 1: Novel CNN flowchart for Multi-view Face Recognition

a. Dataset:

The dataset used in our study is the 'FacePix', which has 5,430 photos, sufficient for implementing the novel CNN model [19]. This dataset stores photographs of people's faces and was developed at Arizona State University's Center for Cognitive Ubiquitous Computing. These pictures were taken using a static camcorder that took a series of facial shots of the subject whereas a light moved in a circle across from them. The dataset comprised 181 face images of 30 distinct persons to satisfy the many views of a single subject. Each photo was taken at an angle ranging from -90 degrees to +90 degrees. The images were transformed to a dimension of 60 x 51 pixels and a grayscale type.

b. Data Visualisation:

Data visualization helps in giving a better understanding of the data and the arrangement of the data in the dataset. Data visualization can be defined as the graphical display of information and data. By utilizing visual components such as charts, graphs, and maps, the data visualization methods simplify the process of recognizing and analysing patterns, trends, and outliers in the dataset. This is accomplished through the use of the tools. Multiplex Data Visualization refers to the many or multiple perspectives on a subject that will be fed into the deep learning architecture. The angle and different views of a single subject are emphasized heavily in data visualization. Fig. 2 describes the possible multiplex data visualization of the above-considered dataset. The images show the multiple views and face angles of a single subject. Similarly, the data can be projected for different subjects with different face angles and views.

Fig. 2. Multiplex data visualization of face images

c. Pre-processing:

The process of preparing pictures that can be used in model training and inference is referred to as "pre-processing". This includes things like resizing, rotating, and making color adjustments, amongst other things. Scaling, rotating the photos in the range of 50 degrees, zooming the images with a zoom range of 0.05, and applying a 0.05 range of shear to various photographs of diverse subjects were considered in this study. The images were flipped horizontally and vertically, and the brightness was changed at random to extract better features from them. Fig. 3 graphically describes the pre-processed images with the above-mentioned processing parameters.

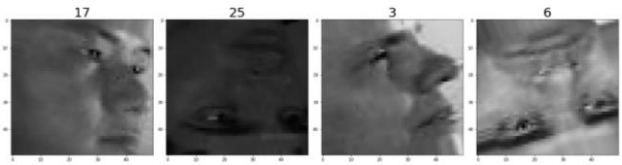

Fig.3. Pre-processed images

d. Feature Engineering:

Feature engineering refers to the act of choosing, altering, and converting raw data into features that can be utilized in machine learning. It is necessary to build better features for machine learning to perform well on new jobs. In this study, the pre-processed photos were used to generate duplications, which might enhance the learning of the deep learning architecture. This, in turn, can result in a high level of accuracy in categorizing the individuals' varied states and facial angles.

e. Deep Learning Architecture:

Deep learning is a special category of machine learning technique that has the capabilities to learn the features of the given data by itself. In the most basic form, it is a neural network consisting of three layers. Although they are not fool proof, these neural networks make an attempt to mimic how the human brain functions by enabling itself to "learn" from enormous quantities of data. However, this does not mean that they are successful. Figure 4 illustrates a typical deep learning model which was incorporated in our proposed novel model architecture.

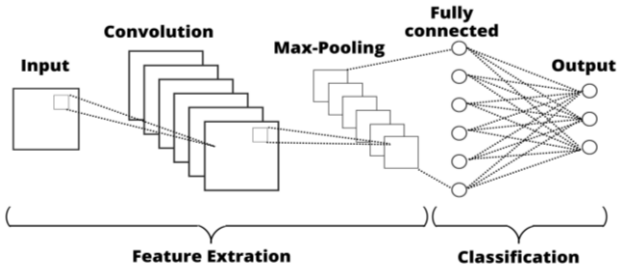

Fig. 4. Typical deep learning model architecture

A standard CNN model consists of three layers: convolutional, pooling, and fully connected (Fig. 4). The neural networks that belong to this category are characterised by the grid-like structure of the data they process. The convolution stage is the foundation of a convolutional neural network and is responsible for the bulk of its computations. The number of necessary calculations is decreased and the spatial size of the depiction is decreased by pooling. The Fully Connected Layer, on the other hand, has connections to both the previous and the current layers.

In this research, the deep learning model was built up of various layers, such as convolutional layers, which could act as one of the critical feature selection layers in deep neural networks. There were eight convolutional layers in the deep learning implementation of this paper. In the convolution process, a small matrix of numbers, which is called a kernel, is passed over the fed image and is transformed based on the values from the kernel (Eq. 1).

$$G[m,n]=(f*h)[m,n]=\sum_j \sum_k h[j,k]f[m-j,n-k] \quad (1)$$

where *f = input image, h = kernel, m,n = indexes of rows and columns of the result matrix*

After placing the kernel over a selected pixel of the image, each value from the kernel is taken and multiplied in pairs with the corresponding values from the image. In the end, the summation is done, and the results are entered into the feature map. Then, the bordering is done around it to solve the problems like image shrinking due to convolution and the impact of outskirt pixels being smaller than the center ones (Eq. 2).

$$p=\frac{f-1}{2} \quad (2)$$

where *p = padding and f = kernel* dimension

In addition, there were MaxPooling layers that had the capability of selecting the most significant element from the portion of the feature map that was being affected by the filter. As a consequence of this, the outcome of the max-pooling layer corresponded to a feature map that contains the most notable features from the prior feature map. As a result, the dimension of the output matrix becomes n_{out}.

$$n_{out}=\left[\frac{n_{in}+2p-f}{s}+1\right] \quad (3)$$

where n_{in} *= input matrix, p = padding and s= stride.*

In this process, batch normalization is done after two convolution layers. It is used to

reparametrize the deep neural network so that we get standardized data which stabilizes the model training and thus speeds it up. It also smoothens the loss function. In batch normalization, the activation vector (Z) of hidden layers is standardized through the use of mean (μ) and variance (δ^2). The procedure is presented by the four equations (Eq. 4-7) given below.

$$\mu = \frac{1}{n}\sum_i Z^{(i)} \tag{4}$$

$$\sigma^2 = \frac{1}{n}\sum_i (Z^{(i)} - \mu)^2 \tag{5}$$

$$Z_{norm}^{(i)} = \frac{Z^{(i)} - \mu}{\sqrt{\sigma^2 - \varepsilon}} \tag{6}$$

$$Z = \gamma * Z_{norm}^{(i)} + \beta$$

The flatten layer was responsible for flattening the outcome from the feature selecting layers, which passed into the dense layers of the neural network model. The flatten layer unstacks all the tensor values of the previous layer outputs into a 1-dimensional tensor shape so that it can be used as the input for the next layer, which is the dense layer. There are four dense layers in this deep learning model, which learned to flatten critical features from the feature selecting layers, i.e., convolutional layers. Finally, the design of the deep learning architecture was completed to classify the 30 separate categories.

IV. RESULT AND DISCUSSION

For the successful implementation of the deep learning architecture, the different layers involved in the algorithm handle the entire crucial feature selection and learning process. In our study, the convolutional layers, MaxPooling layers, flatten layers, and dense layers were the deep learning model layers for multi-view facial recognition. Concentrating on convolutional layers, the essential features from the extracted features were chosen using the convolutional layer mathematical standards. In this research, the artificial intelligence frame is entirely built with eight convolutional layers, which are responsible for the possible critical feature selections. The MaxPooling layers select the maximum important elements from the feature map of the convolutional layers. In this research, four MaxPooling layers were used. The training of the deep learning model for identifying the different facial angles and the multiple views plays a vital role. This study implemented a state-of-the-art training model for obtaining the possible highest accuracy in recognizing the multiple facial views and the angles of the various subjects. Fig. 5 explains the deep learning model architecture of the classification and recognition of various views and facial angles of 30 different subjects by providing a detailed idea about the multiple layers in the deep learning model.

conv2d_19 (Conv2D)	(None, 50, 50, 64)	102464
batch_normalization_3 (Batch	(None, 50, 50, 64)	256
max_pooling2d_9 (MaxPooling2	(None, 25, 25, 64)	0
conv2d_20 (Conv2D)	(None, 25, 25, 128)	204928
conv2d_21 (Conv2D)	(None, 25, 25, 128)	409728
max_pooling2d_10 (MaxPooling	(None, 12, 12, 128)	0
conv2d_22 (Conv2D)	(None, 12, 12, 256)	295168
conv2d_23 (Conv2D)	(None, 12, 12, 256)	590080
max_pooling2d_11 (MaxPooling	(None, 6, 6, 256)	0
conv2d_24 (Conv2D)	(None, 6, 6, 512)	1180160
conv2d_25 (Conv2D)	(None, 6, 6, 512)	2359808
max_pooling2d_12 (MaxPooling	(None, 3, 3, 512)	0
dropout_3 (Dropout)	(None, 3, 3, 512)	0
flatten_3 (Flatten)	(None, 4608)	0
dense_10 (Dense)	(None, 256)	1179904
dense_11 (Dense)	(None, 128)	32896
dense_12 (Dense)	(None, 64)	8256
dense_13 (Dense)	(None, 30)	1950

Total params: 6,370,462
Trainable params: 6,370,334
Non-trainable params: 128

Fig. 5. Proposed deep learning model architecture diagram

Out of the 5430 images, the training data set comprised 4350 images, the validation data set contained 1080 images, and the entire 5430 images were used as the testing data in our study. The model was trained for the epoch of 15. The proposed novel CNN model involved 6 x 370 x 334 trainable parameters and 128 non-trainable parameters. The "Adam" optimizer was used to optimize each epoch while training. During the training stage, the loss was inversely proportional to the accuracy of the model. The validation loss also decreased with each epoch but in a zig-zag manner. The reduction in the training and validation losses with the increase in the number of epochs can be observed in Fig. 6.

Fig. 6. Training Graph of deep learning model

The performance of the model proposed in this study was examined by various performance metrics. A confusion matrix, also known as an error matrix, is a special table structure that permits visualization of the performance of an algorithm in the field of machine learning and specifically for the problem of statistical classification. The matrix represents the factors such as true positive, true negative, false positive, and false negative classified by the model. Fig. 7 describes the confusion matrix for this research. The confusion matrix in Fig. 7 shows that the classification of multiple views and the different facial angles of the subjects have low false positives and false negatives. This, in turn, improves its accuracy in classifying the given data.

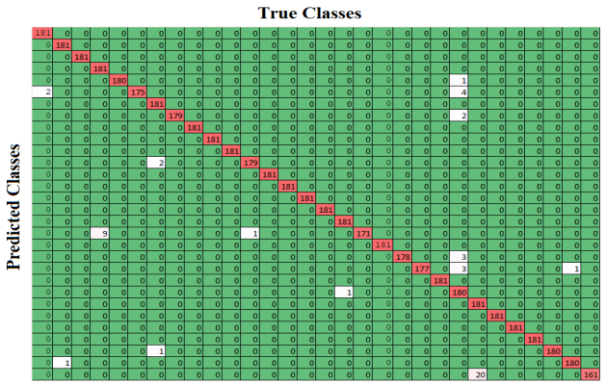

Fig. 7. Confusion matrix of the proposed model

The amount of precision that a deep learning model possesses is one of the metrics that is used to measure how successful the model is. Precision refers to the dependability of a positive prediction that the model has made. The ratio of the number of actual positive results to the total number of successful forecasts is the concept that is meant to be referred to when using the term "precision." To calculate the recall, compute the ratio of the total number of positive samples to the total number of samples that were correctly classified as positive. This will give you the recall percentage. After that, multiply this ratio by 100 to get the final answer. The recall measure is used to evaluate the ability of the model to recognize positive samples. The number of positive samples recovered increases in proportion to the size of the recall. The F1 Score is a statistic that is calculated by calculating the harmonic mean of a classifier's accuracy and recall scores and then merging those two values into a single result. Its major goal is to assess how well two distinct classifiers compare to one another in terms of their efficiency. Let's say, for the sake of this discussion, that classifier A has a higher recall and classifier B has higher precision. In this situation, the F1 scores for each of these classifiers may be evaluated against one another to see which one produces greater results. Support is referred to as the number of observations of the genuine response that fits into each category of objective values. Table 1 displays the relevant values for accuracy, F1 Score, recall, and support that our suggested CNN model has.

Table 1. Class-wise performance parameters of the proposed model

Class no	Precision	Recall	F1-score	Support
1	0.99	1	0.99	181
2	0.99	1	1	181
3	1	1	1	181
4	0.95	1	0.98	181
5	1	0.99	1	181
6	1	0.97	0.98	181
7	0.98	1	0.99	181
8	1	0.99	0.99	181
9	1	1	1	181
10	1	1	1	181
11	1	1	1	181
12	0.99	0.99	0.99	181
13	1	1	1	181
14	1	1	1	181
15	1	1	1	181
16	1	1	1	181
17	0.99	1	1	181
18	1	0.94	0.97	181
19	1	1	1	181
20	1	0.98	0.99	181
21	1	0.98	0.99	181
22	1	1	1	181
23	0.93	0.99	0.96	181
24	0.9	1	0.95	181
25	1	1	1	181
26	1	1	1	181
27	1	1	1	181
28	1	0.99	1	181
29	0.99	0.99	0.99	181
30	1	0.89	0.94	181

The accuracy of our proposed model was compared with other state-of-the-art models for face recognition available in the literature (Table 2). While comparing multi-view algorithms, it was found that the detection algorithm based on CNN and BLSTM proposed by He *et al.* (2017) had a performance accuracy of 73.50% [16]. The deep neural network suggested by Zheng *et al.* (2016) showed an overall accuracy of 85.20% [20]. The Deep Dense Face Detection algorithm developed by Farfade *et al.* (2015) improved the accuracy to 94.02%. Further, the deep neural network (MVP) algorithm proposed by Zhu *et al.* (2014) could take the accuracy to 98.10% [21]. On the other hand, our proposed model could achieve an accuracy of 99.06%, which is higher than the previously reported algorithms. However, the comparison may feel unfair as different models used different image pre-processing methods, and that could lead to deviation in accuracy for different algorithms.

Table 2: Comparison of Accuracy for Different Algorithms

Methods	Accuracy (%)
CNN with BLSTM- RNN	73.50
Zheng DNN	85.20
Farfade DDFD	94.02
Zhu DNN (MVP)	98.10
Our method (Novel CNN)	99.06

V. CONCLUSION

In this study, a CNN-based deep learning method has been applied to recognize faces from a multi-view perspective. The max-pooling layers extracted the maximum elements from convolutional layers to get the higher correlated features essential to face recognition. In this way, the model learned not only distinct spatial particulars but also extracted prominent features yielding fewer errors in results. This proposed model also reduced type I and type II errors which is a great step towards an efficiently performing recognition technique. Hence, this research concludes by suggesting the start of art solutions for the pre-processing of the multiple views and facial angles of the subjects. The deep learning architecture stands better in classifying the different subjects with nearly zero false positives and negatives, resulting in high accuracy.

V. REFERENCES

[1] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, "Identification of human faces," *Proceedings of the IEEE,* vol. 59, no. 5, pp. 748-760, 1971.

[2] N. Hesse, T. Gehrig, H. Gao, and H. K. Ekenel, "Multi-view facial expression recognition using local appearance features," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 3533-3536: IEEE.

[3] F. Zhao, J. Li, L. Zhang, Z. Li, and S.-G. J. F. G. C. S. Na, "Multi-view face recognition using deep neural networks," vol. 111, pp. 375-380, 2020.

[4] M. Sharif, F. Naz, M. Yasmin, M. A. Shahid, A. J. J. o. E. S. Rehman, and T. Review, "Face Recognition: A Survey," vol. 10, no. 2, 2017.

[5] K. R. Storrs and N. J. a. p. a. Kriegeskorte, "Deep learning for cognitive neuroscience," 2019.

[6] S. S. Farfade, M. J. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015, pp. 643-650.

[7] D. Triantafyllidou and A. Tefas, "A fast deep convolutional neural network for face detection in big visual data," in *INNS conference on big data*, 2016, pp. 61-70: Springer.

[8] A. N. Razzaq, R. Ghazali, N. K. El Abbadi, and M. J. W. Dosh, "A Comprehensive Survey on Face Detection Techniques," vol. 19, no. 1, 2022.

[9] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Aggregate channel features for multi-view face detection," in *IEEE international joint conference on biometrics*, 2014, pp. 1-8: IEEE.

[10] M. Ma and J. Wang, "Multi-view face detection and landmark localization based on MTCNN," in *2018 Chinese Automation Congress (CAC)*, 2018, pp. 4200-4205: IEEE.

[11] F. Zhao, J. Li, L. Zhang, Z. Li, and S.-G. Na, "Multi-view face recognition using deep neural networks," *Future Generation Computer Systems,* vol. 111, pp. 375-380, 2020.

[12] S. Ravidas and M. J. Ansari, "An Efficient Scheme of Deep Convolution Neural Network for Multi View Face Detection," *International Journal of Intelligent Systems Applications,* vol. 11, no. 3, p. 53, 2019.

[13] A. B. Deshmukh and N. U. Rani, "Optimization-Driven Kernel and Deep Convolutional Neural Network for Multi-View Face Video Super Resolution," *International Journal of Digital Crime Forensics,* vol. 12, no. 3, pp. 77-95, 2020.

[14] W. J. I. T. o. A. C. Zheng, "Multi-view facial expression recognition based on group sparse reduced-rank regression," vol. 5, no. 1, pp. 71-85, 2014.

[15] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," *IEEE Transactions on Multimedia,* vol. 18, no. 12, pp. 2528-2536, 2016.

[16] J. He, D. Li, B. Yang, S. Cao, B. Sun, and L. Yu, "Multi view facial action unit detection based on CNN and BLSTM-RNN," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 848-853: IEEE.

[17] K. Bhangale, P. Ingle, R. Kanase, and D. Desale, "Multi-view multi-pose robust face recognition based on VGGNet," in *International Conference on Image Processing and Capsule Networks*, 2021, pp. 414-421: Springer.

[18] S. Jie and Q. Yongsheng, "Multi-view facial expression recognition with multi-view facial expression light weight network," *Pattern Recognition Image Analysis,* vol. 30, no. 4, pp. 805-814, 2020.

[19] A. Farahani. *FacePix*. Available: https://data.world/cptmazi/facepix

[20] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. J. I. T. o. M. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," vol. 18, no. 12, pp. 2528-2536, 2016.

[21] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning multi-view representation for face recognition," 2014.