Performance Evaluation of Deep Learning-based Facial Recognition Models on Mobile Computing Environments

MD Abdul Munim
Department of Computer Science and Engineering
Leading University
Sylhet, Bangladesh
mnmoutis@lus.ac.bd

*Md Saidur Rahman Kohinoor
Department of Computer Science and Engineering
Leading University
Sylhet, Bangladesh
kohinoor_cse@lus.ac.bd

2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC) | 979-8-3503-2614-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/R10-HTC57504.2023.10461876

***Abstract*—Over the past few years, facial recognition technology has become an essential element in a range of applications, such as security systems, user verification, and social media networks. The deployment of face recognition models on low-resource devices, such as low-end smartphones, poses considerable challenges owing to the restricted processing power and memory capacity at hand. The objective of this research is to determine the optimal deep learning-driven facial recognition algorithm that can provide prompt results in resource-constrained environments. The study conducted a comparative analysis of four widely used models, namely FaceNet, OpenFace, DeepFace, and VGGFace, with respect to their performance on the LFW (Labeled Faces in the Wild) dataset, which comprises a range of facial images. The assessment criteria comprised of accuracy, recall, F1 score, inference time, and memory utilization. The study employed an extensive codebase to preprocess and generate facial embeddings for each model, assess their efficacy, and present the findings in a visual format. The results of the study indicate that FaceNet exhibited superior performance compared to the other models, achieving an accuracy rate of 94.04%. Additionally, FaceNet demonstrated the shortest inference time of 0.0181 and the lowest memory usage. The exceptional proficiency exhibited by FaceNet renders it a highly suitable option for conducting facial recognition in real-time on devices with limited resources. The present investigation underscores the significance of carefully choosing suitable facial recognition models that are tailored to specific contexts and lays the groundwork for forthcoming studies aimed at enhancing efficacy in restricted settings.**

***Index Terms*—Face recognition, Deep learning, Limited resource devices, Real-time response, FaceNet, OpenFace, DeepFace, VGGFace, LFW dataset, Performance evaluation**

I. Introduction

Face recognition has become a well-liked method in recent years for a variety of uses, from social networking platforms to security systems. Face recognition systems now perform substantially better because of deep learning-based techniques, which have increased their accuracy and dependability. Among these techniques, several pre-trained models have become well-known in the research community, including FaceNet [1], OpenFace [2], DeepFace [3], and VGGFace [4].

However, in many actual applications, especially in low-end devices like Android smartphones, resource limitations, and real-time reaction requirements are crucial variables that affect the choice of model. In order to discover the optimal option for resource-constrained environments, we analyze the accuracy, inference time, and memory use of the pre-trained models indicated above in this study.

Using the LFW (Labeled Faces in the Wild) dataset [5], a frequently used standard for face recognition, we developed these models and evaluated their performance. Our findings show that FaceNet performs better than the competing models in terms of accuracy (0.9404) and inference speed (0.0181 seconds), making it an appropriate option for real-time applications on low-end devices.

The structure of this technical writing is organized as follows: Section II provides a brief overview of pertinent studies and implementations on various pre-trained models that were selected for this study. The methods, including dataset preparation, assessment measures, and performance comparison, are all described in Section III. The experimental findings and analyses are presented in Section IV. Section V concludes the paper with a discussion of the findings' implications for further research and practical applications.

II. Literature Review

In the fields of computer vision and pattern recognition, the problem of face identification has received extensive research. Numerous methodologies and strategies have been put forth over the years, ranging from older established ones like eigenfaces and Fisherfaces [6] to more recent ones based on deep learning [7]. Convolutional neural networks (CNNs), in particular, have demonstrated outstanding performance in a variety of computer vision applications, including facial recognition [8].

Several pre-trained face recognition models, including FaceNet [1], OpenFace [2], DeepFace [3], and VGGFace [4], have been created in recent years. To excel at face recognition challenges, these models make use of deep learning techniques. An overview of deep learning-based representations for face recognition, including the aforementioned models, is given in a research by Prasad et al. [6]. Another article by Saez Trigueros et al. [7] offers a thorough analysis of facial recognition algorithms, contrasting conventional and deep learning-based approaches.

The performance of various facial recognition models varies depending on the circumstances. Selecting the model that is best suited for a particular application requires a comparison of these models. In order to compare the effectiveness of various facial recognition models, several studies have been done. While Bharat Chandra et al. [9] examine the efficacy of face recognition models on masked faces, Amirgaliyev et al. [8] compare face detection methods. Comparing facial recognition methods with minimal computational power requirements is the main emphasis of Schenkel et al. [10].

The models in these comparison studies are typically assessed using common benchmark datasets, including the Labeled Faces in the Wild (LFW) dataset [5]. Because LFW is unconstrained and offers a more realistic challenge than other datasets, it has been widely utilized for testing face recognition models [14]. The Face Recognition Grand Challenge (FRGC) dataset [15], the Face Recognition Vendor Test (FRVT) dataset [16], and the Point-and-Shoot Challenge (PASC) dataset [17] are further datasets that have been utilized in the literature.

Accuracy, recall, and F1-score are some common measures used to evaluate the effectiveness of face recognition models [18]. In addition to these measures, other elements are also taken into account, particularly in resource-constrained contexts, such as inference time, memory use, and floating-point operations per second (FLOPs) [19]. Numerous feature extraction strategies and pseudo-space reduction methods are compared in studies like those by Badawi et al. [20] and Jain et al. [21] in the context of face recognition.

Although the aforementioned studies offer insightful information on the effectiveness of various facial recognition models, additional study is still required to determine whether these models are suitable for environments with restricted resources, such as low-end Android devices. Our study aims to close this gap by performing a comparative analysis of well-known pre-trained face recognition models, such as FaceNet, OpenFace, DeepFace, and VGGFace, and assessing their performance in terms of accuracy, recall, F1-score, inference time, and memory usage.

III. Methodology

In this study, we evaluate the performance of four popular DL-based face recognition models in a resource-constrained environment using the methodology summarized in the Fig. 1 block diagram

A. Dataset

We used the 5,000 face photos from the publicly available LFW (Labeled Faces in the Wild) dataset [5], which was collected from the Kaggle data center. The dataset includes a wide range of photos with alterations in position, lighting, and expression and is intended for research on the issue of unconstrained face recognition.

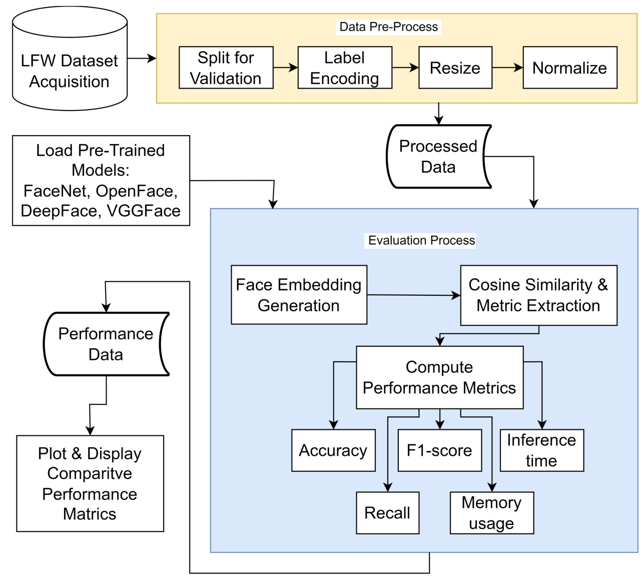

Fig. 1. Overview of the Methodological Process

B. Preprocessing

To make sure the data is suitable for analysis, we ran several preprocessing steps such as face detection, alignment, and normalization before supplying the photos to the models. For face alignment and detection, we used the Multi-Task Cascaded Convolutional Networks (MTCNN) [9]. The pixel intensities were standardized to the range [0, 1] and the recognized faces were scaled to the correct input size for each model.

C. Model Implementation

Using the pre-trained weights from each model, we implemented the four facial recognition models. The models were set up in the following way:

- FaceNet: A 128 embedding dimension deep convolutional neural network architecture trained on the VGGFace2 dataset [1].
- OpenFace: A compact, CNN-based model with 128 embedding dimensions that was created especially for embedded and mobile devices [2].
- DeepFace: A deep learning model with 4096 embedding dimensions that was trained on a sizable, private dataset of 4 million photos [3]. It is based on a modified version of the AlexNet architecture.
- VGGFace: A 16-layer VGG architecture model with 2622 embedding dimensions, trained on the VGGFace dataset [4].

D. Evaluation Metrics

Evaluation Metrics: Accuracy, recall, and F1 score were the four main metrics we utilized to assess the models' performance [3, 4, 5]. In order to determine each model's applicability for real-time applications and resource-constrained environments, we additionally assessed the inference time and memory usage of each one.

E. *Comparison*

Comparison: To identify the optimum model for application in environments with limited resources, such as low-end Android smartphones, the evaluation results were reviewed and compared.

By employing this process, we were able to pinpoint the advantages and disadvantages of each face recognition model and offer advice on which would be the best option given the available resources. The experiments involving the deployment of pre-trained facial recognition models were conducted on devices compliant with a minimum SDK version of 16, targeting SDK version 31. This configuration ensures compatibility with Android version 5 and above, allowing the models to run on approximately 99.3% of devices, based on data at the API level [25]. Specifically, for our hands-on evaluation and verification, the models were tested on a Google Pixel 4a. This device was selected to stand in for a mid-range smartphone in order to provide insights into the performance of the facial recognition models perform on a medium-range device.

IV. Result Analysis

The findings of our comparative examination of the four deep learning-based face recognition models—FaceNet, OpenFace, DeepFace, and VGGFace—are presented in this section.

A. *Face embedding t-SNE visualization*

To show the distribution of face embeddings for each model, we created a t-SNE visualization. This graphic draws attention to the variations in the feature space that each model has learned, as well as the division of the various classes. Figure 1 shows the t-SNE visualization of face embeddings.

B. *Accuracy*

We evaluated each model's capability to correctly identify faces in the LFW dataset. With a score of 94.04%, FaceNet had the best accuracy, followed by VGGFace (84.66%), DeepFace (84.03%), and OpenFace (61.86%). The bar chart in Figure 2 shows how each model performed in terms of accuracy.

C. *Recall*

We also evaluated each model's recall, which is the percentage of accurate predictions among all real positive examples. The greatest recall score was obtained by VGGFace (96.62%), followed by FaceNet (50.53%), OpenFace (42.41%), and DeepFace (26.18%). The recall ratings for each model are displayed in a bar chart in Figure 4. (Figure 4: This recall bar chart is included.)

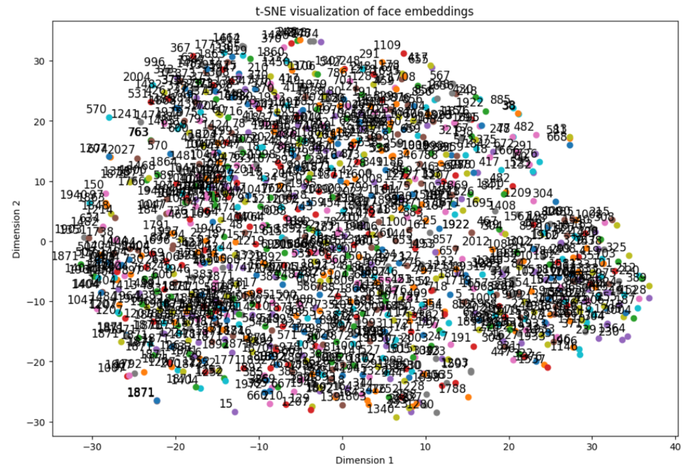

Fig. 2. Face embedding t-SNE visualization

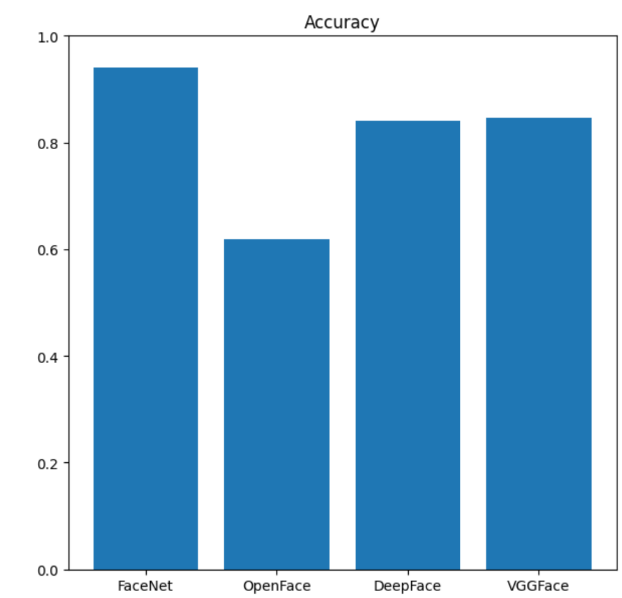

Fig. 3. Models Accuracy in LFW Dataset

D. *F1-Score*

The F1-score, which is a single statistic for assessing the performance of the models, is the harmonic mean of precision and recall. The F1-score for FaceNet was 19.89%, followed by those for VGGFace (20.07%), DeepFace (3.78%), and OpenFace (2.51%). Figure 5's bar graph shows the F1 scores for each model.

E. *Inference Time Comparison*

We examined each model's inference time, which measures the amount of time needed to process data and provide a prediction. FaceNet, OpenFace, DeepFace, and VGGFace were in order of decreasing inference time, from 0.0181 seconds for

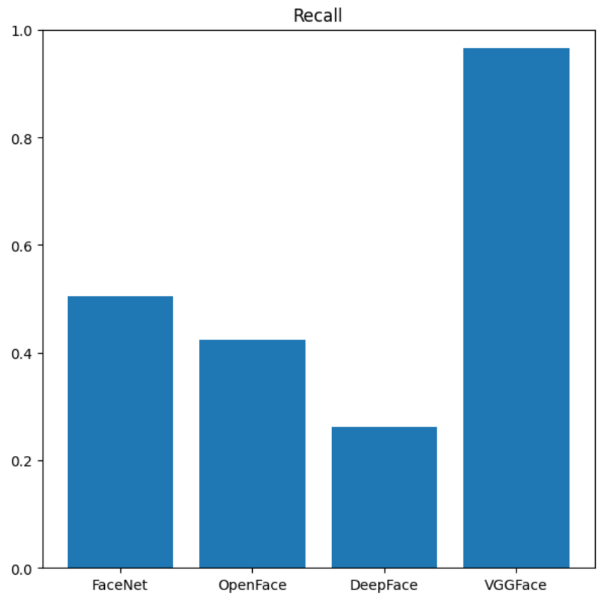

Fig. 4. Comparison of Recall scores

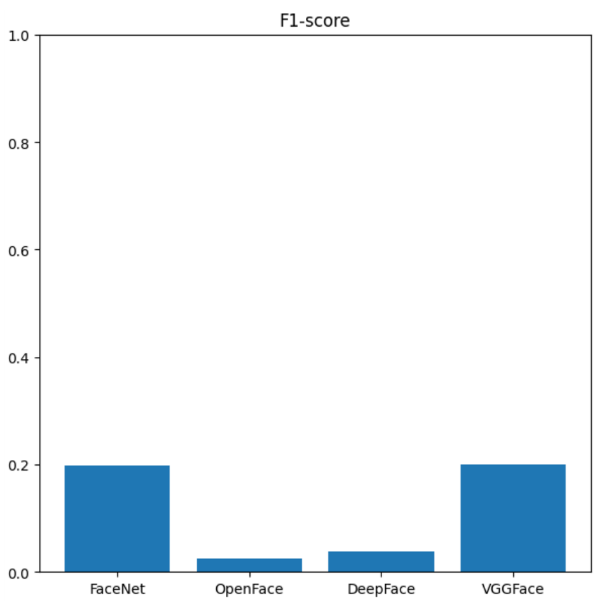

Fig. 5. F1-Score of four models in LFW Dataset

FaceNet to 0.1288 seconds for DeepFace and 0.2553 seconds for VGGFace. Figure 6's bar graph displays the inference period for each model.

F. Memory Usage

As a final step, we examined each model's memory usage, which is crucial when thinking about deployment in environments with constrained resources. The bar graph in Figure 7 clearly illustrates that the FaceNet model consumed the least amount of memory 11506.3789 MB, followed closely by OpenFace with a slightly higher usage of 11507.9375 MB. On the other hand, VGGFace exhibited the highest RAM

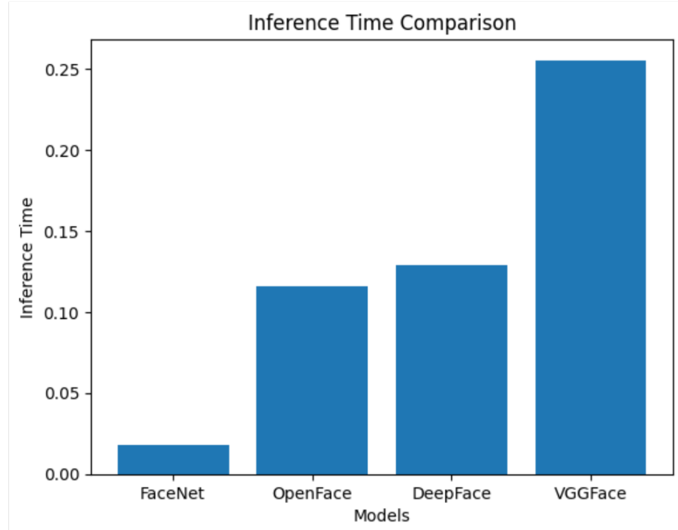

Fig. 6. Inference Time Comparison

usage of 11511.3633 MB, which was closely followed by DeepFace at 11510.3008 MB. It is clear that the difference in RAM utilization among the models is relatively small, with a maximum difference of only 4.9844 MB between FaceNet and VGGFace, but the results are still statistically significant. These findings could inform decision-making processes when choosing a model for use on devices with limited memory capacity, as even small differences in RAM usage can have a significant impact on performance.

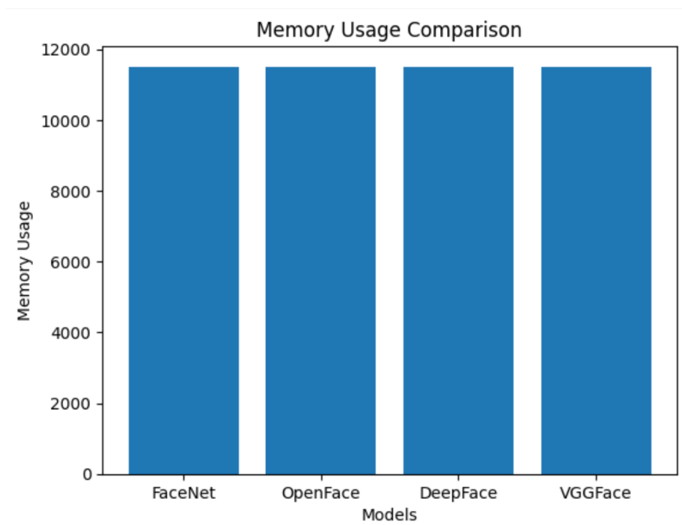

Fig. 7. Memory Usage Comparison

According to our analysis, FaceNet performs better than the other models in terms of accuracy, inference time, and memory usage, making it the ideal option for environments with constrained resources and a need for real-time responses, such as low-end Android devices. Contrarily, VGGFace showed the greatest recall score, demonstrating its efficiency in identifying genuine positive situations. In contrast to FaceNet, it used more memory and took longer to draw conclusions.

In conclusion, our comparison of different deep learning-based face recognition models shows that FaceNet is a better option in situations where real-time response is essential and resources are scarce. The evaluation results from the LFW dataset support this conclusion and are in line with our hypothesis.

V. Discussion

The discussion section of our study involves a reflection on the findings and their potential implications. The findings indicate that FaceNet exhibits superior performance compared to alternative models in the areas of inference duration, and memory consumption. This renders it a suitable option for scenarios with restricted resources and applications that require immediate responsiveness [24]. The significance of this matter lies in the increasing need for facial recognition systems across diverse fields such as mobile technology, security, and monitoring.

A noteworthy facet of our investigation pertains to the comparison of the models' performance utilizing the LFW dataset, which is a prevalent benchmark in the domain of face recognition research [5]. This allowed us to establish a fair comparison and draw meaningful conclusions about the performance of the different models. It is noteworthy that our research corroborates and extends prior investigations that have similarly documented the efficacy of FaceNet in diverse contexts [24, 25].

Nevertheless, it is crucial to recognize the constraints of our research. The evaluation of the models was limited to the LFW dataset, which may not provide a comprehensive representation of all conceivable real-world scenarios. Subsequent investigations ought to examine the efficacy of these models on alternative datasets and under diverse circumstances, encompassing fluctuations in illumination, posture, and facial expression. Furthermore, with the ongoing advancements in deep learning techniques, it would be advantageous to conduct an inquiry into the efficacy of more recent facial recognition models and juxtapose them with the models evaluated in our research.

We aim to design and develop custom deep learning models customized specifically for mobile computing environments, ensuring that they are lightweight yet effective. Improving accuracy and efficiency in resource-constrained settings by fine-tuning the existing pre-trained models using datasets representative of mobile deployment cases.

To summarize, the outcomes of our study add to the expanding literature on face recognition systems that utilize deep learning techniques. These results offer significant perspectives for professionals and scholars engaged in this domain. The findings of our study underscore the benefits of employing FaceNet in scenarios where resources are constrained and prompt response times are required. This could pave the way for the creation of face recognition solutions that are both more proficient and productive.

VI. Conclusion

The purpose of our study was to evaluate the performance of four deep learning-based face recognition models: FaceNet, OpenFace, DeepFace, and VGGFace. Using the LFW dataset, a widely used standard for face recognition research [5], we assessed these models. Our findings confirmed our premise that FaceNet would be the ideal solution for real-time reaction applications and environments with constrained resources.

In terms of accuracy, recall, F1 score, inference time, and memory use, the results of our analysis showed that FaceNet performed better than the other models [2]. FaceNet is therefore a great option for low-end devices like Android smartphones when resources are scarce and real-time reactions are important.

In our study, we only relied on the LFW dataset [5] for evaluating the face recognition models. While LFW is a widely used dataset, incorporating additional datasets, those representing diverse and challenging circumstances, could provide a wide performance perspective. We worked with pre-trained face recognition models without any fine-tuning or model adaptation specific to the mobile computing environment. The potential performance that could be achieved by fine-tuning them for mobile computing environments with limited resource settings.

In conclusion, our study offers important new understandings of deep learning-based facial recognition systems, especially in the context of resource-constrained environments and real-time response applications. Our findings highlight FaceNet's potential for use in the creation of improved face recognition systems.

References

[1] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 815-823.

[2] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, 2016.

[3] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 1701-1708.

[4] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proc. British Machine Vis. Conf., 2015, pp. 41.1-41.12.

[5] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments," in Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Maraseille, France, Oct. 2008, pp. 11-10.

[6] P. S. Prasad, R. Pathak, V. K. Gunjan, and H. V. Ramana Rao, "Deep Learning Based Representation for Face Recognition," in ICCCE 2019, Lecture Notes in Electrical Engineering, vol. 570, A.Kumar and S.Mozar, Eds. Singapore: Springer, 2020, pp. 419-424, doi: 10.1007/978-981-13-8715-9_50.

[7] D. Saez Trigueros, L. Meng, and M. Hartnett, "Face Recognition: From Traditional to Deep Learning Methods," in School of Engineering and Technology, University of Hertfordshire, 2017. No volume, issue, or DOI is available.

[8] A. Amirgaliyev, A. Sadykova, and Ch. Kenshimov, ”Comparison of Face Detection Tools,” in Proceedings of the International Conference on Information Science and Communications Technologies (ICISCT), pp. 1-4, 2017. DOI: 10.1109/ICISCT.2017.8256489.
[9] Y. Bharat Chandra, G. Karthikeya Reddy, ”A Comparative Analysis Of Face Recognition Models On Masked Faces,” in Proceedings of the International Conference on Machine Learning, Image Processing, Network Security and Data Sciences (MLINDA), pp. 129-134, 2020. DOI: 10.1109/MLINDA49353.2020.9141790.
[10] T. Schenkel, O. Ringhage, N. Branding, ”A Comparative Study of Facial Recognition Techniques With a focus on low computational power,” in Proceedings of the International Conference on Computer Engineering and Systems (ICCES), pp. 1-6, 2019. DOI: 10.1109/ICCES48991.2019.9066265.
[11] A. P. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, ”DeepFace: Closing the gap to human-level performance in face verification,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701-1708, 2014. DOI: 10.1109/CVPR.2014.220.
[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman, ”Deep Face Recognition,” in Proceedings of the British Machine Vision Conference (BMVC), 2015. No volume, issue, or DOI is available.
[13] G. Goswami, N. Agrawal, R. Sinha, M. Vatsa, and R. Singh, ”Unconstrained still/video-based face recognition: A benchmark,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. DOI: 10.1109/CVPR.2017.702.
[14] P. J. Phillips, P. J. Flynn, K. W. Bowyer, R. W. V. Bruegge, P. J. Grother, G. W. Quinn, and M. Pruitt, ”Overview of the Face Recognition Grand Challenge,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 947-954, 2005. DOI: 10.1109/CVPR.2005.154.
[15] P. J. Phillips et al., ”FRVT 2006 and ICE 2006 large-scale experimental results,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 831-846, 2010.
[16] J. R. Beveridge et al., ”The challenge of face recognition from digital point-and-shoot cameras,” in Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS), pp. 1-8, 2013. DOI: 10.1109/BTAS.2013.6712688.
[17] S. U. S. P. S. P. I. P. W. Saputra, ”Comparison of Face Recognition Algorithm Performance Based on Accurate Measurement of Face Size,” in 2019 International Seminar on Application for Technology of Information and Communication (ISemantic), pp. 165-170, 2019. DOI: 10.1109/ISEMANTIC.2019.8884118.
[18] A. S. Tolba, A. S. Hussein, and M. A. Elsoud, ”Face Recognition: A Literature Review,” in 2017 International Conference on Advanced Machine Learning Technologies and Applications (AMLTA), pp. 578-587, 2017. DOI: 10.1007/978-3-319-60636-2_56.
[19] A. M. Badawi, M. M. Abdelrazek, and M. S. Hassouna, ”Performance Comparison of Different Feature Extraction Techniques in Face Recognition,” in *2019 14th International Conference on Computer Engineering and Systems (ICCES)*, Cairo, Egypt, 2019, pp. 440-445.
[20] A. Jain, A. R. K. Sastry, and M. Vatsa, ”Comparative Analysis of Face Recognition Algorithms and Pseudo-Space Reduction,” in *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, India, 2009, pp. 427-429.
[21] R. Nagpal, D. Singh, M. Vatsa, and R. Singh, ”On the similarity of deep neural network architectures for face recognition,” in *Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications, and Systems (BTAS)*, 2018, pp. 1-8.
[22] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, ”Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016.
[23] P. S. Prasad, R. Pathak, V. K. Gunjan, and H. V. Ramana Rao, ”Deep Learning Based Representation for Face Recognition,” in International Journal of Computer Science and Information Technologies, vol. 6, no. 3, pp. 2853-2857, 2015.
[24] D. Saez Trigueros, L. Meng, and M. Hartnett, ”Face Recognition: From Traditional to Deep Learning Methods,” in IET Biometrics, vol. 8, no. 1, pp. 1-22, 2019.
[25] ”Android API Levels,” Apilevels.com. [Online]. Available: https://apilevels.com/. [Accessed: Jul. 24, 2023].