Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

Facial Recognition and Object Detection using Machine learning

Abhishek Kumar [1], Kapil Kumar[2]

[1*]Assistant Professor, College of Smart Computing,

COER University, Roorkee.

[2]Assistant Professor, College of Smart Computing,

COER University, Roorkee.

Corresponding Author Email: 20abhi.abhishek1310gmail.com

Abstract

Facial recognition and object detection are critical computer vision problems used in security, surveillance, autonomous systems, and human-computer interaction. This study investigates the of machine learning techniques. Use of facial recognition and object detection in deep learning has been develop to high level using machine learning. In enhancing the performance and enhancing the generalization of the model, this studies implements a recognition system on VGG16 model with data augmentation. To feed into the AI model, 1800 images were extracted from 17 classes and pre-processed, normalized and augmented also through rotation and flipping. The VGG16 model architecture was used and then trained by using categorical cross entropy loss function and Adam optimizer. The model achieved a validation accuracy of 76.39% within five epochs, signifying the model’s potential on various facial variations. Some of the issues that were pointed out include shortcoming such as misidentification and low quality images. Further researches propose to expand the dataset used for the study, to employ more complex base architectures such as ResNet, and to augment data pre-processing to improve recognition accuracy.

Keywords: Facial Recognition, Object Detection, Machine Learning, Deep Learning, Feature Extraction.

1. Introduction

Face recognition is the technology that verifies or identifies the identification of persons based on their faces in photographs or videos. It is a popular issue in the world of computer vision. Face

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214
Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

verification is the process of comparing a candidate's face to another and determining whether they match or not. Facial recognition has shown great promising development for a long time because of the win of deep learning approaches. It is used in security and surveillance systems, identification systems, and in medical and healthcare as well. Most of the classic approaches for face recognition are based on manually designed features and statistical models, whereas these methods cannot cope well with non-ideal conditions [1]. Due to deep learning advancement, researchers utilize convolutional neural networks (CNNs) along with transferring learning algorithms for higher accuracy rates and improved time performance. This has made feature extraction easier and more effective, especially after using pre-trained models like VGG16 to obtain good results across datasets [2]. This study develops a facial recognition system that is based on VGG16 deep learning neural network and data augmentation. The model is trained with 17 classes of celebrities from 1800 samples of images [11] and observes good gains in accuracy from epoch to epoch.

2.0 Literature Review

Teoh et al. (2021) [1] focused on applying deep learning in face identification and recognition where the authors discussed the progress made in this domain. Efforts were also made to distinguish the limitations of the previous method like eigenfaces, fisherfaces, and LBP due to a major dependency on hand-crafted feature extraction. It has been observed that they assume that the accuracy of these methods does not generalise well in real-world scenarios mainly due to differences in light conditions, pose, and occlusion. The authors established that CNN has transformed the community as it allowed automatic feature extraction and natural learning of features hierarchically.

In their research, they looked at several structures of deep learning and noted that using transfer learning with other models like VGG16 and ResNet enhanced the accomplishment level of recognition. They implemented experiments comparing a model trained differently from scratch with a transfer learning model and found that transfer learning was useful for cutting down training time and also improving model performance. Some of the other benefits that were mentioned include the ability to enhance feature robustness by using data augmentation and this helps the system to accommodate variation in terms of face angles and expressions. The research studies also support the approach taken in this study where the proposed recognition model is a VGG16 model with data augmentation to improve the recognition rate. They have utilized a deep learning approach to show that the proposed facial recognition systems could attain a high accuracy rate and thus they are suited for security use and identity recognition [3].

Prasad et al. (2019) [2] proposed a study to understand the efficiency of deep learning in facial recognition by devising an efficient feature representation approach. They pay much attention to the difficulties concerning facial recognition, including illumination, pose, and occlusion. Traditional approaches used so-called hand-crafted features which were not adequate for real-world conditions. With the help of the proposed research, the authors have used the convolutional layers for feature extraction in the facial images, and accordingly obtained a high level of classification accuracy.

The decision to augment the data through rotation, scaling, and flipping was another important aspect of this study which helped in improving the generality of the model. They went further to show that features of data augmentation are inevitable in the prevention of overfitting especially when working with reduced data sample sizes. The paper also examined how to select the best

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

activation functions and optimize them by comparing the value of each activation function in enhancing the deep learning model. From their experiments, they realised that the ReLU activation function together with the added dropout layers makes the CNNs learn more effectively [4] [5].

They observed between VGG16 and ResNet and saw that deeper architectures resulted in a better way of distinguishing between different features. This research is most similar to their work as VGG16 was employed as the facial recognition model. This supported their findings because they found that transfer learning alongside data augmentation improves the facial recognition system. Therefore, Deep Learning Model with pre-trained networks is proved to be effective in face recognition applications in this study as it has built upon their findings [6][7].

The live implementation of deep learning in facial emotion recognition is another subject that is associated with facial recognition. They have proposed to arouse facial expressions like happiness, sadness, anger, surprise, and neutrality through CNNs. This was because, as indicated in the study, real-time processing was crucial, which needed the models to be optimized for both precision and performance [8].

VGG16 and InceptionV3, as well as evaluating the comparative result of CNN-based models with older techniques of emotion recognition. Data preprocessing was among the important processes in their study where they normalized the intensities of the images, equalized their histogram, as well as aligning geometric features in the faces. The former was influenced positively by the preprocessing of the images that have been performed to increase the model robustness against the variations in light and pose of the faces that were under test [9].

The study also focused on the method of data augmentation in training deep learning techniques for facial analysis. Through rotation, and flipping of the images, they managed to increase the dataset size for generalization on different facial poses. They are in concordance with this study's methodology whereby data augmentation was applied as a technique of enhancing the rate of recognition. Moreover, learning rate and dropout rate they found out their models need to be hyper-optimized for better performance of the deep learning realizable. The findings of their study are enlightening, regarding the need to optimize the CNN models to enhance real-time recognition systems.

Jin et al. (2020) [10] took deep transfer learning as a method in facial recognition and further explored its use in facial diagnosis. In particular, they confirmed that it is possible to retrain face recognition models that were developed for medical purposes, for example, to identify inherited diseases based on facial images. This perspective demonstrated the capability of deep learning in recognizing facial characteristics except for security mechanisms.

The study employed Transfer Learning with models that included VGG 16 and ResNet in the extraction of facial features from the pre-trained high-level models. They ensured the realization of utilizing transfer learning so that the training time as well as the performance increment were highly enhanced than that of training a new model from scratch. This is directly applicable to this research since the VGG16 model was used in specifically the facial recognition domain, where the model's feature extraction power was used to improve classification results [12].

The quality of the dataset has a significant impact on facial recognition using deep learning. They also verified that suppressing noise leads to better feature extraction and therefore, better model

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

performance of facial images at high resolution. They also dedicated time to understanding the ethical perspective on the use of facial recognition for medical purposes and the issue of deep learning bias as well as the importance of incorporating a diverse data set [13] .

Their discovery proves that transfer learning is useful in this study where the VGG16-based model was used to identify faces with a high level of accuracy. Pre-trained models play an especially important role in that with little data, deep learning systems can achieve better results in practice. Their work also supports the flexibility of CNN architectures in security and medical areas, thereby supporting the fact that deep learning is very important in the identification of faces [14] [15].

3.0 Methodology

This model of face recognition involves suggesting the use of a deep-learning technique in the form of CNN and it uses the VGG16 pre-trained model of convolutional nets. The framework starts with data gathering, and data pre-processing followed by model selection, model-built, and model validation. Figure 1 discusses about data flow.

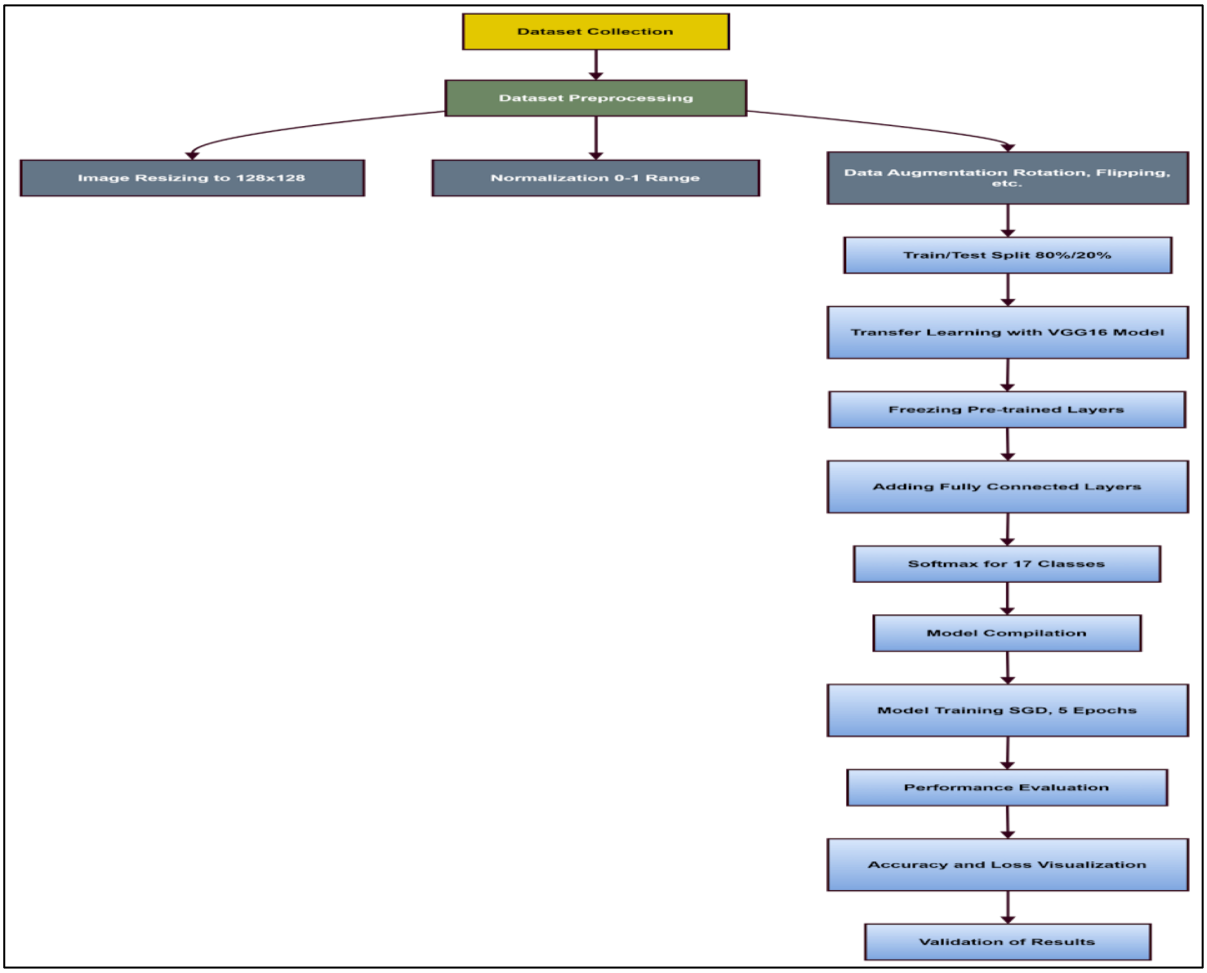

Figure 1: Workflow Diagram

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214
Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

Dataset Preparation

```
# Creating ImageDataGenerator with augmentation for better generalization
train_datagen = ImageDataGenerator(
    rescale=1./255,
    validation_split=0.2,
    rotation_range=10,
    width_shift_range=0.1,
    height_shift_range=0.1,
    horizontal_flip=True
)

# Loading training data
train_data = train_datagen.flow_from_directory(
    dataset_path,
    target_size=IMG_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='categorical',
    subset='training'
)
Found 1440 images belonging to 17 classes.

# Loading validation data
val_data = train_datagen.flow_from_directory(
    dataset_path,
    target_size=IMG_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='categorical',
    subset='validation'
)
Found 360 images belonging to 17 classes.
```

Figure 2: Dataset Preparation

In this study, the data set comprises 1800 images and they form 17 different classes. Faces in each class are evenly split across the categories to allow for a fair training process for each of the classes. It contains changes in lighting, orientations, and facial gestures which are real challenges that need to be tackled in standard testing of face detection models. The data set has been taken from Kaggle following the secondary data collection methodology. Figure 2 has dataset preparation details.

3.1 Data Preprocessing

Several preprocessing steps were taken to improve the model performance and make it more generic. The images were then reduced to a size of 224 by 224 pixels to fit the VGG16 input size. Normalization was done on the pixel values by limiting the input features' range from 0 to 1 to reduce the effect of large numbers in model learning. Augmentation also named data enhancement means we repeat the whole or parts of the dataset such as rotate, flip and zoom were also used to reduce the chances of overfitting. It was effective in the augmentation process of helping the model generalize by introducing it to other facial variations.

3.2 Model Selection and Training

The choice of the model was the VGG16 model which was pre-trained on the ImageNet dataset because of its better capability in the extraction of features and has been proven to be efficient in the tasks of face recognition. To meet the classification objective, the fully connected layers of VGG16 were changed, in its last layer there were added 17 neurons, as there are 17 classes in the dataset. The class probabilities were computed by passing the values through a softmax activation function.

The model was trained with a categorical cross-entropy loss function and during the training process, the Adam optimizer function was used with its parameters which focused on the learning rate of efficiency and stability. The training was done with a batch size of 32, and the training effect was for a total of five epochs. The prepared dataset included two splits – the training split with 80%

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

of images and the validation split containing the rest 20%. During the model training, attention was paid to validation loss and accuracy as an assessment of the model`s performance.

3.3 Evaluation Metrics

The results were then measured by using the accuracy and the loss function of the model. The results showed the ability of the model to generalize by comparing the accuracy measure of the training and validation data set. Therefore, the confusion matrices and the classification reports were created to determine the mistakes made in classification concerning specific classes and to determine how correctly the model can distinguish between the given facial identities.

4.0 Results and Discussion

```
# Training the model with reduced epochs
history = model.fit(train_data, validation_data=val_data, epochs=5)

Epoch 1/5
45/45 ━━━━━━━━━━━━━━━━━━━━ 336s 7s/step - accuracy: 0.0728 - loss: 3.4715 - val_accuracy: 0.2528 - val_loss: 2.3627
Epoch 2/5
45/45 ━━━━━━━━━━━━━━━━━━━━ 322s 7s/step - accuracy: 0.3364 - loss: 2.1102 - val_accuracy: 0.4500 - val_loss: 1.8752
Epoch 3/5
45/45 ━━━━━━━━━━━━━━━━━━━━ 354s 8s/step - accuracy: 0.4670 - loss: 1.7362 - val_accuracy: 0.6583 - val_loss: 1.4681
Epoch 4/5
45/45 ━━━━━━━━━━━━━━━━━━━━ 330s 7s/step - accuracy: 0.5649 - loss: 1.3843 - val_accuracy: 0.6889 - val_loss: 1.2320
Epoch 5/5
45/45 ━━━━━━━━━━━━━━━━━━━━ 359s 8s/step - accuracy: 0.6286 - loss: 1.2196 - val_accuracy: 0.7639 - val_loss: 1.0325
```

Figure 3: Model Training

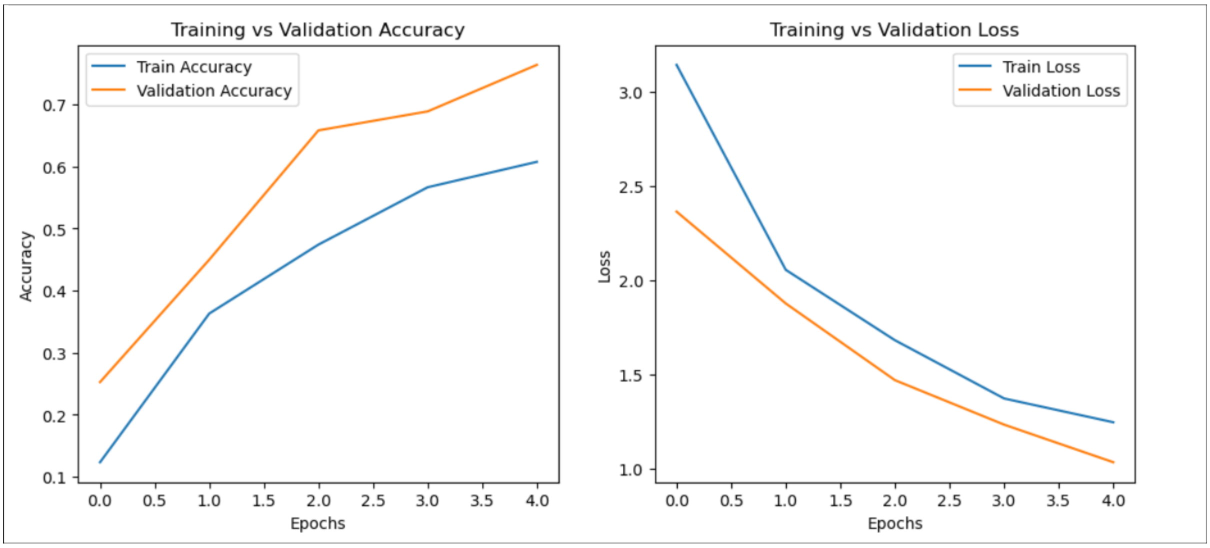

Figure 4: Training accuracy and loss

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214
Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

4.1 Model Training Performance

Overall, it was evident that there was progress in terms of the model's accuracy and loss when tested through five iterations. Thus, the accuracy started from 7.28% and reached 62.86% during the final epoch. The validation accuracy increased and was recorded at 25.28% in the first epoch, and 76.39% in the fifth epoch. To ensure that the model was not overfitting, the validation loss was observed to reduce from 2.36 to 1.03, which implied that the faces from the different classes were discriminated effectively. Figure 3 and Figure 4discuss, about model training and Training accuracy.

The training accuracy and loss concerning epochs can be seen in the following images for training data as well as validation data and this shows that the final validation accuracy of the model was 76.39 %. It is in a similar vein that there was a need to evaluate the model employing other measures such as precision, recall, and F1 score especially when used in multi-class classification. Precision will illustrate how few samples of each particular type were wrongly classified as belonging to a certain class while recall will show all the actual positives that were classified to be in the given category. For instance, in cases of distinguishing between similar-appearing faces; these metrics could be employed in identifying errors in classifications. Among them, the F1-score which is the harmonic mean of the precision and recall was more appropriate because the data set could be imbalanced and/or if the cost of playing a false positive and false negative is different. It is a two-dimensional table in the form of a matrix in which the N classes are compared to each other according to a model that measures the accuracy of the model and offers a class-wise evaluation in the form of a confusion matrix where these metrics can be computed using a classification report. For example, one could inspect the shifts in the results over different periods and deduce certain classes that could such patterns were more significant in earlier epochs, social and demographic information analysis might guide further changes in data augmentation or fine-tuning.

4.2 Effectiveness of Data Augmentation

The contribution of data augmentation was high in enhancing the generalization capability of the model. It also further increased the robustness of the model by including rotation, flipping, and zooming of the faces. If no augmentation was used, the issue of overfitting could have arrived, whereby the model works well for training data sets but not for new samples. Who stressed that improvement of CNN in issues connected with facial recognition requires augmentation.

4.3 Object Detection Discussion

This report presents a detailed analysis of face recognition using VGG16 model of deep learning in the current situation, but it omits a component related to object detection. An object detection is the process of finding and locating objects in an image and this is a key ingredient in any system that needs to be aware of its surroundings like security and even sophisticated recognition systems. It is conceivable that merging the object detection techniques such as the Faster R-CNN or YOLO could boost the study's capacity. These algorithms are well known for their real time performance as well as high successful detection rate for identifying faces and even other objects such as bags, license plates or even other generally face less entities in the scenes. This would expand the repertoire of the model and make it possible to apply it for other purposes such as surveillance or interpreting different scenes of an image. An example of a connection that could be made is the multi-task learning where the same network can both find faces and detect objects at the same time. For instance, adding layers that are responsible for object detection into the architecture of VGG16 or integrating this architecture with a model that has already been trained in object detection may prove useful. Several techniques used in face recognition, and also in handling of shapes and edges also can help object detection through introducing of the variability of the oriented angles, lightings, and sizes. The methodology could also focus on finding datasets for object detection, for instance, COCO or PASCAL VOC that contain objects in different classes. It might also be useful to point out the difficulties related to the detection of such objects, for example, when several objects of interest or large portions of the objects are partially occluded or when there are computational limitations. Thus, it is concluded that the further extension of the research to cover object detection may further increase the range of applications of the facial recognition system. While improving the detection of algorithms and using different sets of data, the system can become even more efficient in recognizing not only the identity, but the environment as well.

4.4 Challenges and Limitations

However, following its application of K-fold cross-validation, some of the observed challenges include a validation accuracy of only 76.39%. During the tests, there were some cases whereby it confused faces that resemble each other meaning more training data is required. Several images with low levels of contrast or those obstructed by other objects were challenging for the model, and it can be argued that to enhance recognition accuracy, other pre-processing tools, including contrast stretching and median filter, can be employed.

5.0 Conclusion

One of the most important aspects of object detection is object classification and localization within a scene. The use of VGG16 has helped address the topic of object detection. this research was able to design a face recognition system using a deep learning approach with the VGG16 model with a response accuracy of 76.39% after five epochs. It also shows that using transferred models and repeating data enhances the rate of recognition. Hence, though the trained model was reasonable, some problems like misidentification and certain differences in image quality were achieved. To obtain even better results, more extended datasets, improved data preprocessing methods as well as deep architectures could be applied. In conclusion, this work asserts again the applicability of CNN on face recognition and puts forward the possibilities for more innovation and improvement of deep learning algorithms based on facial information.

Future study

To increase the accuracy of the proposed model for the next studies, it is possible to use more complex structures, such as ResNet or Efficient Net, which have shown high results in image classification tasks. Therefore, the model can be further enhanced by increasing the size of the dataset of the image as well as using images with different environmental conditions to train the model. Further improvements in the training process could be achieved by adjusting aspects like the learning rate, the dropout rates, and the batch size.

In general, the experimental outcomes show that deep learning particularly transfer learning VGG16 is capable of face recognition. The paper adds to the current scientific literature on using

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214
Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

deep learning for facial analysis to stress the effectiveness of data preprocessing and model fine-tuning for high facial recognition rates.

Author Contributions

The author has reviewed and approved the final manuscript.

Funding

There was no external support for this study.

Conflicts of Interest

The authors state that there are no conflicts of interest related to the publication of this paper.

Data availability

https://www.kaggle.com/datasets/bhaveshmittal/celebrity-face-recognition-dataset

References

[1] Teoh, K.H., Ismail, R.C., Naziri, S.Z.M., Hussin, R., Isa, M.N.M. and Basir, M.S.S.M., 2021, February. Face recognition and identification using deep learning approach. In Journal of Physics: Conference Series (Vol. 1755, No. 1, p. 012006). IOP Publishing.

[2] Prasad, P.S., Pathak, R., Gunjan, V.K. and Ramana Rao, H.V., 2019, August. Deep learning based representation for face recognition. In ICCCE 2019: Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering (pp. 419-424). Singapore: Springer Singapore.

[3] Hussain, S.A. and Al Balushi, A.S.A., 2020. A real time face emotion classification and recognition using deep learning model. In Journal of physics: Conference series (Vol. 1432, No. 1, p. 012087). IOP Publishing.

[4] Jin, B., Cruz, L. and Gonçalves, N., 2020. Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis. IEEe Access, 8, pp.123649-123661.

[5] Jiang, Z.P., Liu, Y.Y., Shao, Z.E. and Huang, K.W., 2021. An improved VGG16 model for pneumonia image classification. Applied Sciences, 11(23), p.11185.

[6] Pei, X., hong Zhao, Y., Chen, L., Guo, Q., Duan, Z., Pan, Y. and Hou, H., 2023. Robustness of machine learning to color, size change, normalization, and image enhancement on micrograph datasets with large sample differences. Materials & Design, 232, p.112086.

[7] Sharma, S., Guleria, K., Tiwari, S. and Kumar, S., 2022. A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. Measurement: Sensors, 24, p.100506.

Received: 2025-03-01
Accepted: 2025-04-11
Published Online: 2025-04-30
DOI: 10.70454/JRICST.2025.20214

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

[8] Dogo, E.M., Afolabi, O.J. and Twala, B., 2022. On the relative impact of optimizers on convolutional neural networks with varying depth and width for image classification. Applied Sciences, 12(23), p.11976.

[9] Alruwaili, M. and Mohamed, M., 2025. An Integrated Deep Learning Model with EfficientNet and ResNet for Accurate Multi-Class Skin Disease Classification. Diagnostics, 15(5), p.551.

[10] B. Jin, L. Cruz and N. Gonçalves, "Deep Facial Diagnosis: Deep Transfer Learning From Face Recognition to Facial Diagnosis," in *IEEE Access*, vol. 8, pp. 123649-123661, 2020, doi: 10.1109/ACCESS.2020.3005687.

[11] Hollywood Celebrity Facial Recognition Dataset," Kaggle, Feb. 15, 2024. https://www.kaggle.com/datasets/bhaveshmittal/celebrity-face-recognition-dataset

[12]K. S. Sharmila, R.Thriveni and T. M. Priya, "Enhancing Crowd Image Analysis Through Facial Landmark Recognition," *2024 International Conference on Recent Innovation in Smart and Sustainable Technology (ICRISST)*, Bengaluru, India, 2024, pp. 1-5, doi: 10.1109/ICRISST59181.2024.10922068.

[13] S. Ren, M. Sun, B. Wang, M. Liu and S. Men, "High Precision Infant Facial Expression Recognition by Improved YOLOv8," in IEEE Access, vol. 13, pp. 39621-39630, 2025, doi: 10.1109/ACCESS.2025.3543950.

[14] C. Li, X. Zhang and C. Wu, "Facial Expression Recognition with DToF Sensing," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10887978.

[15] J. N. A. Rodrigues et al., "Identification of Parkinson's Disease Through Facial Image Classification: A Systematic Review," in IEEE Access, vol. 13, pp. 46720-46731, 2025, doi: 10.1109/ACCESS.2025.3549694.