Transfer Learning Technique with VGG-16 for Near-Infrared Facial Expression Recognition

Chukwuemeka C Atabansi\textsuperscript{1,2*}, Tong Chen\textsuperscript{1,2}, Ranlei Cao\textsuperscript{1,2} and Xueming Xu\textsuperscript{1,2}

\textsuperscript{1}School of Electronic and Information Engineering, Southwest University, Beibei, District, Chongqing, 400715, China
\textsuperscript{2}Chongqing Key Laboratory of Nonlinear Circuit and Intelligent Information Processing, Southwest University, Beibei District, Chongqing, 400715, China
\textsuperscript{*}Corresponding author’s e-mail: mecusbans@gmail.com

Abstract. In this paper, we investigate a deep learning vgg-16 network architecture for facial expression recognition under active near-infrared illumination condition and background. In particular, we consider the concept of transfer learning whereby features learned from high resolution images of huge datasets can be used to train a model of relatively small dataset without losing the generalization ability. The pre-trained vgg-16 network architecture with transfer learning technique has been trained and validated on the Oulu-CASIA NIR dataset comprising of six (6) distinct facial expressions, and average test accuracy of 98.11\% was achieved. The validation on our test data using the confusion, the precision, and the recall matrix reveals that our method achieves better results in comparison with the other methods in the literature.

1. Introduction
The identification system for human facial expressions has experienced huge development since its invention and numerous techniques and methods are widely utilized. For this reason, further research on FER is fully in place to provide constant improvement over accuracy of identification. Facial expressions are one of the powerful and natural means for human being to present their intentions and emotions [2]. Emotion recognition perhaps is an extensive, complex, and useful research topic in many fields such as biomedical engineering, health, education, neuroscience, psychology [3] and so on. The ability of computers to differentiate, understand human expression, and take appropriate action is one of the key research areas particularly in human computer interaction (HCI) [4]. Suppose that during online studying, student acceptance of material will rise if the computer knows the state of a student expression. With the knowledge of a student expression, the computer can provide appropriate learning in line with the conditions of the student expression at that particular time [4]. Another useful application is that through an expression recognition system, lies can be detected during crime interrogation [6]. Emotion detection in the area of biomedical engineering focuses on predicting human expression, and computer-assisted diagnosis of psychological disarrays [3]. There are several ways in the recent past to detect emotional states of a person or individuals such as EEG, speech analysis, galvanic skin response (GSR) and so on [3]. Applications can be extended to missions involving patients with Autism, very aged, newborn babies, and so on, who will not be able to express their expressions clearly [4].
Most recent datasets for facial expression recognition were captured under light condition between 380-750 nm [5]. However, several backgrounds produce huge amount of variation for FER [6] and for this reason, the high level of illumination conditions can generate poor results to some extent, and also weaken the performance of FER. Though in the recent past, preprocessing algorithms to reduce the problem associated with illumination have been adopted in some works [7]. Hence, it is still very hard to be applied in nature and obtain commendable results. Therefore, a preferable way for solving this problem is to apply and use near-infrared camera between the range of 780-1100 nm for FER task [8]. Moreover, methods and techniques that were successfully used in recognizing facial expression with transfer learning technique was proposed by Q. Wen et al., they propose deep learning model architecture for FER on three (3) public datasets (CK+, JAFFE and FACES). They trained and used pre-trained CNN network architecture such as Inception-V3, VGG-Face, VGG-19 and VGG-16 model architecture. From their results, they achieved a maximum accuracy of 97.16% on FACES dataset using VGG-19 network architecture [3, 9]. Hai-Duong et al. on the other hand proposed a facial expression recognition system using a multi-level CNN model technique which is similar to VGG network architecture. Their models do not only take features with high level, but also features with mid-level. The plain CNN achieves a maximum accuracy of 69.21% while that of their proposed network reaches 73.03% accuracy on a publicly used dataset known as FER2013 [10].

In this research paper, we still use pre-trained CNN network architecture for facial expression classification task. The main thing that makes our work different from others is that we employ a deep convolutional neural network (DCNN) architecture and train it directly on our proposed dataset. To improve the efficiency of our model further and avoid over-fitting as well, we employ and use transfer learning technique. The idea of employing this method with vgg-16 network architecture is to further improve the performance of our network and reduce computational complexity. Specifically, vgg-16 network trained on high resolution images was fine-tuned on our proposed dataset. Feature vectors were extracted from vgg-16 model and the weights of all the five (5) convolutional blocks were frozen. A new dense layer consisting of three (3) layers was added right after the last vgg-16 convolutional layers, and the resulted output was given to a new classifier for classification. The experimental results show that our proposed method with Oulu-CASIA NIR dataset, can achieve test accuracy up to 98.11%.

2. Related work

The study of transfer learning is inspired by the fact that human beings can quickly apply knowledge learned in the past to solve new tasks faster or even provides better remedies [11]. The main motivation, and the primary goal for transfer learning tasks in the field of machines specifically machine learning was discussed in a paper written by Pan, Sinno Jialin, and Qiang Yang (2010) [12]. This paper describes the need for lifelong machine learning architectures that retain and reuse knowledge learned in the past. A learning approach to transfer learning is the multi-task learning method as postulated by Caruana, R [13], which tries to learn several tasks at the same time, even when the tasks are not the same. Machines which are used for emotion prediction either by facial expressions only or by vocal sounds are postulated by Albert Mehrabian and Ekman et al. [14, 15]. Gil Levi and Tal Hassner used Mapped Binary Patterns with CNNs to counter illuminations variances and trained an ensemble of vgg-16 networks on CASIA-Webface dataset [16]. In summary, it can be seen that CNN architecture gives optimistic results for classification tasks and can be used for FER application like ours. To reduce the computational complexity and improve accuracy as well, transfer learning can be used in which knowledge obtained from the pre-trained network architecture trained with high resolution images of huge dataset can be transferred to learn on a newly smaller dataset without loosing the generalization ability.

2.1. Transfer learning with VGG-16

Transfer learning is simply a machine learning technique where the knowledge obtained from the previous task can be applied to another related task and at the same time improves the learning
The CNN network architecture like ResNet, VGG, AlexNet and so on, are already trained on a huge image dataset of ImageNet comprising more than one million labelled high resolution images belonging to one thousand (1000) categories. Thus, the knowledge obtained already from a particular task is assigned to learn a new different task. It is especially used where the training data is relatively small. In addition, it shows good performance, especially during classification tasks and the computational complexity is significantly minimized to some extent as the entire operation need not start from the scratch. VGG-16 network architecture is a sixteen (16) layer network used by the VGG group at the University of Oxford to obtain outstanding results in the ILSVRC competition held in 2014. The main feature of the vgg-16 network architecture was the increased depth of the network. It is known for its outstanding performance in several classification tasks like image, object and so on [18]. It was trained on ImageNet dataset.

Table 1. VGG-16 network architecture.

| Layers  | Kernel/filter size | Input size | Layers  | Kernel/filter size | Input size |
|---------|--------------------|------------|---------|--------------------|------------|
| Input   | -                  | 3×224×224  | Conv-9  | 3×3 (512)          | 512×28×28  |
| Conv-1  | 3×3 (64)           | 64×224×224 | Conv-10 | 3×3 (512)          | 512×28×28  |
| Conv-2  | 3×3 (64)           | 64×224×224 | Max-pooling | 2×2             | 512×14×14  |
| Max-pooling | 2×2               | 64×112×112 | Conv-11 | 3×3 (512)          | 512×14×14  |
| Conv-3  | 3×3 (128)          | 128×112×112| Conv-12 | 3×3 (512)          | 512×14×14  |
| Conv-4  | 3×3 (128)          | 128×112×112| Conv-13 | 3×3 (512)          | 512×14×14  |
| Max-pooling | 2×2               | 128×56×56  | Fc      | -                  | 25088      |
| Conv-5  | 3×3 (256)          | 256×56×56  | Fc      | -                  | 4096       |
| Conv-6  | 3×3 (256)          | 256×56×56  | Fc      | -                  | 4096       |
| Conv-7  | 3×3 (256)          | 256×56×56  | Softmax | -                  | 1000       |
| Max-pooling | 2×2               | 256×28×28  |         |                    |            |
| Conv-8  | 3×3 (512)          | 512×28×28  |         |                    |            |

Table 1 shows the architecture of vgg-16 network. Moreover, it uses thirteen (13) convolutional layers and three (3) Fc layers (dense layers). The convolutional layers in vgg-16 network architecture have kernel size, of 3×3 with a stride size of 1 and the padding is the same. The pooling layers are all 2×2 layers with a stride size of 2. The input size of the network is 224×224 pixels. After each pooling layer, the size of the feature map is decreased by half (1/2). The last feature map of the network architecture before the Fc layers is 7×7×512 channels. The layer is expanded into a vector with 25,088 (7×7×512) channels respectively. The last layer of the network is a softmax layer that outputs class probabilities. A static feature representation is mainly used where all the images are represented with the same size. It is one of the most effective and efficient used methods to apply transfer learning to image recognition problems and for transfer learning, a pre-trained network architecture of vgg-16 was used with implementation in tensorflow [19].

3. Methodology

Our proposed method aims to detect the emotional state of individuals under active near-infrared illumination condition based on their respective facial expressions manifested in photos. However, It has been seen that features of the previous layers of a DCNN architecture usually contains colour, and edge information. In contrast, the later layers, contain attributes more specific to the details of the categories. For this reason, the parameters of the previous layers need no or minimal fine-tuning [20]. Inspired by this, in our proposed work, we only fine-tuned the dense layers of the vgg-16 network architecture which are the fully connected layers. The description for transforming the vgg-16 network layers is illustrated as follows; The network is trained on more than a million (1,000,000) high resolution images. It has the capability of classifying images into a thousand categories [21]. The last three (3) Fc layers of the network is designed for these one thousand categories. For an expression
recognition task like ours, these layers need to be fine-tuned [22, 23]. The procedure for fine-tuning the network is to extract the entire layers of the network other than the last three (3) dense layers. Move the layers to the new task and insert the new dense layers, a softmax function, and a classification output layer. In our proposed work, the channel size of the Fc layers are 128, 256 and 512 and the number of classes are set to be six (6). To this end, this part of the paper will describe the training techniques, image preprocessing steps, and the dataset used.

3.1. Facial expression dataset
A dataset has to be designed or acquired, such that it is sufficiently apt to train the CNN model [19]. The dataset comprises set of pictures of faces with various expressions [19]. One of the extensive datasets found to be suitable for the purpose of emotion recognition and as well in this research work is the Oulu-CASIA NIR dataset [1]. It consists of six (6) basic facial expressions (anger, happiness, sadness, fear, surprise, and disgust) from eighty (80) subjects between 23 and 58 years old, and 73.8% of the subjects are males. Each and every expression are captured in three different illumination conditions namely; normal, weak, and dark. Normal illumination indicates that good normal lighting is used, weak illumination means that only computer display is ON and dark illumination means near darkness [1]. In our proposed work, we choose only nine thousand five-hundred and four (9,504) near infrared images from normal illumination as shown in figure 1. These facial expressions have been classified as follows; one equals anger, two equals disgust, three equals fear, four equals happy, five equals sad and six equals surprise. In addition to the image class number (a number between one, and six), the final custom dataset contains 9,504 NIR images, and the data was randomly divided into training, validation and testing sets as shown in table 2 and 3 respectively. It was found in this experimental research work that Oulu-CASIA NIR dataset gave better performance for our proposed method.

Figure 1. Anger, disgust, fear, happiness, sadness and surprise images of one subject from oulu-casia near-infrared facial expression dataset.

Table 2. Image distribution of the six (6) basic facial expressions.

|             | Anger | Disgust | Fear | Happy | Sad | Surprise | Total |
|-------------|-------|---------|------|-------|-----|----------|-------|
| Oulu-Casia NIR dataset | 1625  | 1488    | 1670 | 1601  | 1500| 1620     | 9504  |

Table 3. Image distribution of the six (6) basic expressions for training, validation, and testing.

|            | Anger | Disgust | Fear | Happy | Sad | Surprise | Total |
|------------|-------|---------|------|-------|-----|----------|-------|
| Training data | 1316  | 1205    | 1352 | 1297  | 1215| 1312     | 7697  |
| Validation data | 146   | 134     | 151  | 144   | 135 | 146      | 856   |
| Testing data | 163   | 149     | 167  | 160   | 150 | 162      | 951   |

3.2. Data arrangement and preprocessing
Data have been collected from Oulu-CASIA NIR dataset [1]. The data samples of the dataset are shown in figure 1. For image preprocessing, the following steps are considered, namely; face detection, crop and resize.
Figure 2. Image preprocessing steps; face detection, crop and resize.

3.2.1. Face detection, crop and resize. The process of face detection which is sometimes called face registration is a method of detecting the face portion (region of interest) from a given image [25]. OpenCV LBP algorithm [24] has been used to detect these faces from the images. After this process, the face portion of the image is being cropped out to avoid background complexity, so that it will be very sufficient for the model to train data efficiently. The cropped images were finally resized into 224×224 pixels as shown in figure 2.

Figure 3. Processed images..

3.3. Training procedures

The data and their respective labels were imported from Oulu-CASIA NIR dataset. They were shuffled and split into training, validation and testing sets in the ration 80:10:10. The training and the validation data was used during the training process. Moreover, the training data was specifically used to train the values of the network parameters while the validation data was used to measure the accuracy of the network after each epoch to prevent over-fitting. The model was tested with the unknown test set immediately after the training was completed, and the performance of the network was measured accordingly. In addition, the trained model was saved after the whole training is completed so that it can be used for future predictions if the needs arises. The architecture used is composed of two stages: (1) preprocessing and (2) feature extraction and classification. Here, the proposed architecture uses transfer learning technique with vgg-16 network to extract features and feed its output to a softmax classifier to classify those expressions. Firstly, the region of interest (face portion) is extracted from the given face images by using LBP algorithm implemented by OpenCV to detect the frontal face. Then from the cropped face regions, the image was resized into 224×224 pixels as shown in figure 2 and 3 respectively. Secondly, we included weights equals ‘imagenet’ to download vgg-16 network architecture which was trained on a huge dataset of ImageNet and at the same time, we also set include_top equals ‘False’ to skip the Fc layers of the pre-trained network from downloading. Then after this process, the bottleneck features for training, validation and testing sets were created from the vgg-16 ImageNet model respectively. The input shape of the bottleneck feature of each image which comes after passing the image through vgg-16 is 7×7×512 with batch size equals 64. Feature vectors were extracted from the proposed vgg-16 network architecture and the weights of all the five (5) convolutional blocks were frozen. A new dense layer was introduced and added right
after the last vgg-16 convolutional layers which comprises three (3) layers 512, 256 and 128. The output from the network is given to a new classifier and the images were classified into one of the six (6) basic facial expressions using the softmax function as shown in equation (3). The proposed model that was used to classify these six (6) basic facial expressions contains 13 convolution layers of vgg-16 network architecture as illustrated in table 1. Convolution over an image \( f(x, y) \) using a filter \( w(x, y) \) can be calculated as shown in equation (1). The activation function that was used in the hidden layers is the rectified linear unit activation function (ReLU).

\[
w(x, y) \otimes f(x, y) = \sum_{s=-j}^{j} \sum_{t=-k}^{k} w(s,t) f(x-s, y-t)
\]  

Moreover, relu is applied to introduce the non-linearity of a model [26] because it is known to better transmit the error than the sigmoid function [27]. Equation (2) depicts ReLu activation function.

\[
\varphi (x) = \begin{cases} 
  x, & x > 0 \\
  0, & x \leq 0
\end{cases} = \max(0, x)
\]  

The network has been fed with 224×224 pixels sized images as default input images. There are three (3) Fc layers of size 128, 256 and 512 nodes right after vgg-16 convolutional layers. Like convolution layer, the relu activation function has also been applied in those layers and then after each hidden layer, we inserted a dropout layer and the value of our dropout layers are set to 0.25. It randomly deactivates 25% nodes from the hidden layer to avoid over-fitting the network [28]. At the last stage, the output layer of the network comprises six (6) nodes as it has six (6) classes and the activation function that was used in the output layer as shown in the equation three (3) is the softmax classifier. Adam has been used as the model optimizer while categorical cross entropy was used as the network loss function. Figure 4, shows the proposed architecture that has been designed for this very research work.

![Figure 4. Proposed system architecture: default input image; 224x224 pixels near-infrared images; feature vectors were extracted from the vgg-16 network as the weights of all 5 convolutional blocks were frozen. The output from the network was given to a new classifier and classification was carried out with the softmax function as it has six (6) classes (happiness, sadness surprise, anger, disgust, and fear).](image)

4. Experimental results and discussions
This part of our paper gives the details of the classification results on Oulu-CASIA NIR dataset. The prime aim and objective of this very research work is to identify the expression of the human face
given an image of that person. Because of FER applications in the real time world, it has become so
famous and popular in the recent past years. This research work proposes a model for facial expression
recognition task that makes use of transfer learning technique with vgg-16 model. To prove the
efficiency of our proposed model, the Oulu-CASIA NIR dataset was trained on the model and the
results show good recognition rate across all the six (6) classes as shown in figure 7, 8 and 9
respectively. For the model performance evaluation, we adopt accuracy, confusion matrix, precision
matrix and recall matrix as the evaluation index. A confusion matrix is actually a kind of table layout
that enables us to calculate, compute and also allows visualization of the performance of an algorithm.
Confusion matrix is also known as an error matrix [29] and it is mostly used in machine learning and
specifically the problem of statistical classification. Each and every column represents the instances in
an original, or actual class while each row of the matrix represents the instances in a predicted class
(or vice versa) [30].

---

**Figure 5.** Training and validation metric losses.

**Figure 6.** Training and validation accuracy plots.

**Figure 7.** Confusion matrix for the six (6) basic expressions.

**Figure 8.** Precision matrix for the six (6) basic expressions.

**Figure 9.** Recall matrix for the six (6) basic expressions.
Precision matrix and recall matrix on the other hand are measurements for the accuracy of information retrieval, identification, and classification within a computer program. In short, precision and recall matrix are measurements of relevance. The proposed model is carried out with Keras library using python programming language, and Adam optimization algorithm is used. In addition to Adam optimization, the loss function that was used in the research work is the categorical cross entropy and the results obtained are given in figure 7, 8 and 9 respectively. Even though the Oulu-CASIA NIR dataset has been trained on our proposed network, it has successfully achieved an average accuracy of 98.11% and has also succeeded to sustain higher, nearly equal classification rate for every class as it delivers a firm classifications output as shown in figure 10, 11, 12, 13, 14 and 15 respectively. It can be seen from the results that our proposed approach of FER using vgg-16 pre-trained network with transfer learning technique successfully achieves better recognition accuracy compared to the other existing models. The training and validation progression curve demonstrating the loss, and also the accuracy values after each epoch is given in figure 5 and 6. Analyzing figure 5 and 6, it is evident to see from these figures that our proposed network reaches 98.11% accuracy within 100 epochs. Our proposed method improved the accuracy of network to some extent. Figure 7, shows the confusion matrix, figure 8, shows the precision matrix and figure 9, shows the recall matrix for the six (6) classes. Figure 10, 11, 12, 13, 14 and 15 shows the classified results from Oulu-CASIA NIR dataset.

4.1. Comparison with other existing methods

Our method and other existing methods of facial expression recognition with the same dataset under normal illumination condition was compared. The classification test accuracy achieved by our method under normal illumination condition for FER task are comparable with the existing methods in the literature. The average test accuracy for each class is listed in table 4. The prime advantage of our proposed model in comparison with other methods in the literature is that, there is no need for feature extraction method and as well as intermediary feature selection phase.

| Anger | Disgust | Fear | Happy | Sad | Surprise | Total |
|-------|--------|------|-------|-----|----------|-------|
| Laszlo et al. [32] | 66.15% | 70.77% | 57.81% | 83.08% | 62.50% | 89.23% | 71.59% |
| Taini et al. [31] | 78.26% | 80.00% | 70.73% | 84.00% | 72.50% | 88.00% | 79.40% |
| Taini et al. [33] | 84.78% | 90.00% | 73.17% | 84.00% | 72.50% | 90.00% | 82.77% |
| Our method | 98.16% | 95.97% | 98.20% | 99.38% | 98.67% | 98.15% | 98.11% |

Figure 10. Classified result from oulu-casia near-infrared dataset; anger.

Figure 11. Classified result from oulu-casia near-infrared dataset; disgust.
5. Conclusion and future work
In this research paper, we propose a transfer learning technique for facial expression recognition system under active near-infrared illumination condition. The pre-trained vgg-16 DCNN network architecture which was trained on a huge dataset of ImageNet is used to initialize the weights and therefore, the entire network is trained on the input FER image dataset. This network with transfer learning technique was successful in producing a test accuracy of 98.11%. Several experiments were carried out on the Oulu-CASIA NIR dataset and test accuracies were measured. The results obtained from these experiments show that our method gives better classification results in comparison with other methods in the literature. In addition, we believe that more robust results can be achieved by investigating other pre-trained DCNN network architecture with transfer learning technique for the FER task. This can help us to compare their effectiveness, efficiency, and computational complexity. Our future research work will centre mainly on this research direction.
Acknowledgments

We are deeply grateful and highly indebted to our amiable supervisor as well as International College
and College of Electronic and Information Engineering, Southwest University, Chongqing, 400715,
China for all their support towards the success of this research work. We would also like to express
our thanks to the anonymous reviewers and editors of this research paper.

References

[1] Guoying Zhao, Xiaohua Huang, Matti Taini, Stan Z Li, and Matti Pietikäinen. (2011) “Facial
expression recognition from near-infrared videos”. Image and Vision Computing, 29(9):
page 607–619.

[2] E. Sariyanidi, H. Gunes, and A. Cavallaro. (2015) “Automatic analysis of facial affect: A survey
of registration, representation, and recognition”, Pattern Analysis and Machine Intelligence
(PAMI 15), page 1113–1133, doi: 10.1109/TPAMI.2014.2366127.

[3] Özdemir, Mehmet & Elagöz, Berkay & Alaybeyoglu, Aysegul & Sadighzadeh, Reza & Akan,
Aydin. (2019). Real Time Emotion Recognition from Facial Expressions Using CNN
Architecture. 10.1109/TIPETKNO.2019.8895215.

[4] Aris Haris Rismayana. (2018) “Face Expression Recognition Using Artificial Neural Network
(ANN) Model Back Propagation”. ICASI 2018, page 126 DOI: 10.4108/eai.23-4-
2018.2277588.

[5] Zhang Z, Wang Y, Zhang Z. (2014) Face synthesis from low-resolution near-infrared to high-
resolution visual light spectrum based on tensor analysis. Neurocomputing. pp. 140 -154.

[6] Wang S, Liu Z, Lv S, Lv Y, Wu G, Peng P, Wang X. (2010) A natural visible and infrared
facial expression database for expression recognition and emotion inference, IEEE Trans.
Multimed. 2010, 12, pp. 682-691.

[7] Tan X., Triggs B. (2010) “Enhanced Local Texture Feature Sets for Face Recognition under
Difficult Lighting Conditions”, IEEE Trans. Image Process. 2010,19, pp. 1635-1650.

[8] Zhan Wu, Tong Chen, Ying Chen, Zhihao Zhang and Guangyuan Liu. ( 2017) “NIRWxpNet:
Three-Stream 3D Convolutional Neural Network For Near Infrared Facial Expression
Recognition”. Appl. Sci. 2017,7,11184; doi:10.3390/app7111184.

[9] A. Sajjanhar, Z. Wu and Q. Wen. (2018) “Deep Learning Models for Facial Expression
Recognition”, 2018 Digital Image Computing: Techniques and Applications (DICTA),
Canberra, Australia. pp. 1-6, doi: 10.1109/DICTA.2018.8615843.

[10] Hai-Duong Nguyen, Soonja Yeom, Guee-Sang Lee, Hyung-Jeong Yang, In-Seop Na and Soo-
Hyung Kim. (2019) “Facial Emotion Recognition Using an Ensemble of Multi-Level
Convolutional Neural Networks”, International Journal of Pattern Recognition and Artificial
Intelligence. Vol. 33, No. 11, 1940015. https://doi.org/10.1142/S0218001419400159.

[11] Henry C. Ellis. (1965) “The Transfer of Learning. The Macmillan Company”, New York.

[12] Pan, Sinno Jialin, and Qiang Yang. (2010) “A Survey on Transfer Learning”, IEEE
Transactions on Knowledge and Data Engineering 22.10 (2010): 1345-359.

[13] Caruana, R. (1997) “Multitask Learning. Machine Learning”, page 28, 41–75.
https://doi.org/10.1023/A:100739606734.

[14] Albert Mehrabian. (2015) “Silent Messages”, University of California Los Angeles.

[15] P. Ekman and W. V. Friesen. (2017) “Emotional facial action coding system. Unpublished
manuscript”, University of California at San Francisco.

[16] G. Levi and T. Hassner. (2015) “Emotion recognition in the wild via convolutional neural
networks and mapped binary patterns”. Proceedings of the 2015 ACM, page. 503–510.

[17] R. M. Prakash, N. Thenmozhi and M. Gayathri. (2019) “Face Recognition with Convolutional
Neural Network and Transfer Learning”, 2019 International Conference on Smart Systems
and Inventive Technology (ICSSIT), Tirunelveli, India, pp. 861-864.
[18] Johnson, Justin, Andrej Karpathy, and Li Fei-Fei. (2016) Densecap: “Fully convolutional localization networks for dense captioning”. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[19] Ashesh Vaidya, Hitee Sachdeva, R. Brindha. (2018) “Facial Expression Recognition using Convolutional Neural Networks”. 2018 International Journal of Pure and Applied Mathematics Volume 118 No. 22, pp. 245-250.

[20] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, (2014) “How transferable are features in deep neural networks?”, in Advances in neural information processing systems. pp. 3320–3328.

[21] K. Simonyan and A. Zisserman. (2014) “Very deep convolutional networks for large-scale image recognition”. arXiv Prepr. arXiv1409.1556.

[22] P. K. Sonawane and S. Shelke. (2018) “Handwritten Devanagari character Classification using Deep Learning”, in 2018 International Conference on Information, Communication, Engineering and Technology (ICICET), pp. 1–4.

[23] S. Lu, Z. Lu, and Y.-D. Zhang. (2019) “Pathological brain detection based on AlexNet and transfer learning”. J. Comput. Sci., vol. 30, pp. 41–47.

[24] Jia M., Zhang Z., Song P., Du J. (2014) “Research of Improved Algorithm Based on LBP for Face Recognition”. In: Sun Z., Shan S., Sang H., Zhou J., Wang Y., Yuan W. (eds) Biometric Recognition. CCBR 2014. Lecture Notes in Computer Science, vol 8833. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-12484-1_12.

[25] T. U. Ahmed, S. Hossain, M. S. Hossain, R. ul Islam and K. Andersson. (2019) “Facial Expression Recognition using Convolutional Neural Network with Data Augmentation”, 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (iciVPR), Spokane, WA, USA, pp. 336-341, doi: 10.1109/ICIEV.2019.8858529.

[26] X. Glorot, A. Bordes, and Y. Bengio. (2011) “Deep sparse rectifier neural networks, in Proceedings of the fourteenth international conference on artificial intelligence and statistics”. pp. 315–323.

[27] P. Kim (2017) “Matlab Deep Learning: with Machine Learning”, Neural Networks and Artificial Intelligence, ISBN-13 (electronic): 978-1-4842-2845-6, pp. 106-107.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R.Salakhutdinov. (2014) “Dropout: a simple way to prevent neural networks from over-fitting”. The Journal of Machine Learning Research, vol. 15, no. 1, pp.1929–1958.

[29] Stehman, Stephen V. (1997) “Selecting and interpreting measures of thematic classification accuracy”. Remote Sensing of Environment. 62 (1): 77-89.

[30] David MW. (2001) “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation”. Journal of Machine Learning Technologies. 2(1): 37-63, 2001.

[31] M. Taini, G. Zhao, S. Z. Li and M. Pietikainen. (2008) “Facial expression recognition from near-infrared video sequences”, 2008 19th International Conference on Pattern Recognition, Tampa, FL, pp. 1-4, doi: 10.1109/ICPR.2008.4761697.

[32] Jeni, L. A., Hashimoto, H., and Kubota, T. (2012) “Robust Facial Expression Recognition Using Near Infrared Cameras”. Journal of Advanced Computational Intelligence and Intelligent Informatics, Paper: jc16-2-5364, Vol.16 No.2.

[33] Taini M., Zhao G., Pietikäinen M. (2009) “Weight-Based Facial Expression Recognition from Near-Infrared Video Sequences”. In: Salberg AB., Hardeberg J.Y., Jenssen R. (eds) Image Analysis. SCIA 2009. Lecture Notes in Computer Science, vol 5575. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02230-2_25.