ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research

Vol. 12, Issue, 01, pp. 53399-53404, January, 2022

https://doi.org/10.37118/ijdr.23794.01.2022

RESEARCH ARTICLE OPEN ACCESS

DEEP LEARNING APPROACHES FOR FACIAL EXPRESSION RECOGNITION

Aluizio Haendchen Filho*[1], Rodrigo José Fagundes[1], Hércules Antonio do Prado[2], Edilson Ferneda[2] and Lucas Debatin[1]

[1]Laboratory of Applied Intelligence, University of the Itajaí Valley, Itajaí-SC, Brazil
[2] Catholic University of Brasilia, Brasília-DF, Brazil

ARTICLE INFO

Article History:
Received 10th October, 2021
Received in revised form
08th November, 2021
Accepted 27th December, 2021
Published online 30th January, 2022

Key Words:
Facial Expression Recognition,
Recognition of Emotions,
Convolutional Neural Networks.

Corresponding author:
Aluizio Haendchen Filho

ABSTRACT

Recent studies point to the increasing importance of facial expression recognition (FER) in different application areas like healthcare and assistive technologies. Among these, computational diagnosis and evaluation of mental or facial diseases, machine-assisted rehabilitation, clinical psychology and psychiatry, pain monitoring, can be highlighted. FER also plays a relevant role in the case of diagnosis or assessment of cognitive impairments, such as autism and schizophrenia. The literature shows that FER systems effectively working in a health setting is still an open research problem. This paper presents a study on techniques for FER based on images, applying deep learning algorithms. A comparative analysis is performed, pointing out the combinations of algorithms and repositories that manage to obtain gains in the accuracy index.

Citation: ***Aluizio Haendchen Filho, Rodrigo José Fagundes, Hércules Antonio do Prado, Edilson Ferneda and Lucas Debatin.*** "Deep learning approaches for facial expression recognition*", International Journal of Development Research*, 12, (01), 53399-53404.

INTRODUCTION

Facial expressions are one of the most commonly used forms for the semantic modulation technique (Humphries et al., 2006). This technique is used to study how emotions are expressed by the human body. The semantic modulation proved that facial expressions practically follows a pattern that remains valid regardless of the person, culture, ethnicity and gender. Considering that facial expressions are practically universal, there is, for example, a guarantee that smiling people cannot be confused with angry people.
One of the ways in which human beings communicate non-verbally is by means of facial expressions, a not so easy way to be captured and requires more attention than the verbal communication (Michaud et al., 2015). This information is related to a person's mental state, which is expressed, e.g., by means of their emotions, intentions or physical efforts to perform tasks. As a result, automatic emotion recognition with the help of sensors is quite useful in a variety of fields, such as face analysis in healthcare, robotics, psychological studies and virtual reality applications (Leo et al., 2020; Zharovskikh, 2020). Facial Action Coding System (FACS), the most notorious technological work on facial expressions, was published by Ekman and Friesen (1978). FACS method eliminated the need for a human being to see facial expression to identify what a particular person is feeling. For this, FACS analyses certain groups of facial muscles to identify the feeling involved. So, when a person is smiling, certain muscles in his face move, and they vary according to the person's feeling.

The face appearance of a patient may indeed provide diagnostic clues to the illness, the severity of the disease and some vital patient's signs (Toy, 2014). From people's facial expressions at certain times, computer algorithms can determine their mood, for example. For this reason, since the early studies related to automatic image processing, researchers have investigated the possibility of automatically analysing the face to streamline related processes. Considering this context, this paper aims to present an approach for FER in healthcare. It was emphasized the necessity to design facial recognition as independent from human imprecision and the caregiver's attention level as possible. Additionally, its results can help to build new assistive apps.

Background: This section presents an overview of FER applicability on the healthcare, a description of deep learning algorithms and facial expression's datasets, presenting the most relevant in the literature.

Face Analysis in Healthcare: Since the demand for smart, interactive healthcare services is increasing, new research challenges are being posed, such as accurate diagnosis, remote monitoring, and cost–benefit rationalization. A research taxonomy (Figure 1) is introduced by Leo et al. (2020) dividing the face in its main features: eyes, mouth, muscles, skin, and shape. For each facial feature, the author details the computer vision-based tasks for analysing it and the possible healthcare goals to be pursued.

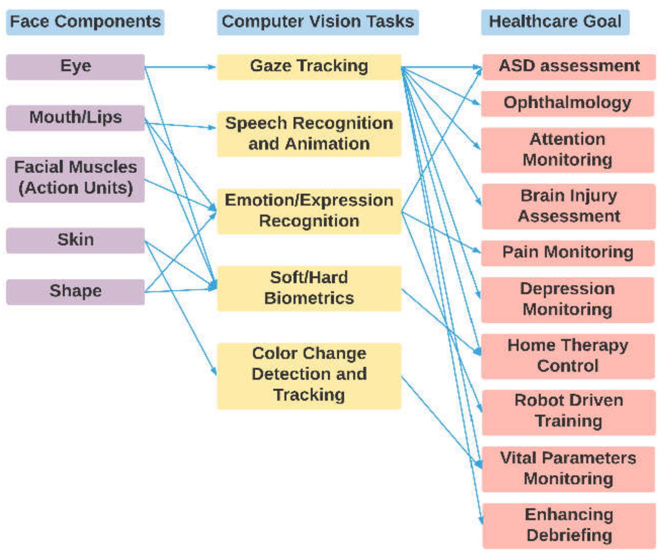

Figure 1. Face features taxonomy (LEO et al., 2020)

For example, when examining the Mouth/Lips Face Component, one can observe the association with three Computer Vision tasks: Speech Recognition, Emotion/Expression Recognition and Soft/Hard Biometrics. It is also interesting to note that Computer Vision Emotion/Expression Recognition, which is one of the objects of this research, is associated with the following healthcare goals: Autism Spectrum Disorder (ASD) Assessment, Pain Monitoring, Depression Monitoring and Robot Driven Training. Six basic emotions, namely happiness, anger, surprise, sadness, fear, and disgust, are well established in the FER context (Mehta et al., 2018). These primary emotions are considered the primary classification for the study of human emotional responses.

Deep Learning: The set of deep learning techniques includes Artificial Neural Networks (ANN) like Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN). They are powering applications like unmanned aerial vehicles, self-driving cars, speech recognition, and now facial expression recognition. Inspired by the mathematical operation known as convolution, CNN are an ANN architecture that aims to create a third function from the values of two other functions. They employ a matrix-based topology for their main intermediate layers. CNN are widely used in the field of computer vision, as they are easy to deal with images represented by a pixel matrix and videos (Goodfellow et al., 2016). CNN is able to assign different importance to each part of the input images (Konam et al., 2018). It consists of convolutive layers, as well as pooling layers, in addition to fully connected layers (Goodfellow et al., 2016). The use of this type of network is due to the fact that it achieves remarkable results in computer vision applications (Krizhevsky et al., 2017). Figure 2 represents a simple architecture for convolutional network for facial analysis.

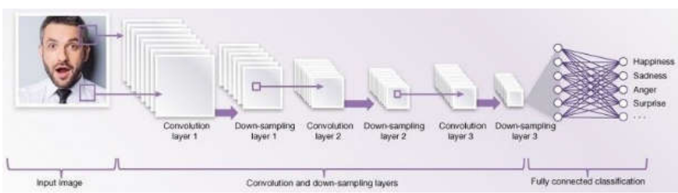

Figure 2. Example of a Convolutional Neural Network architecture for facial analysis. (Cooper, 2017)

The convolution is carried out by applying filters, which are matrices of varying sizes and with different weights. To learn the filters weights for best characterizing the data set is one of the main steps for training these networks. Each filter has the role of generating a map of characteristics, which is the result of multiplying the weights of the filters with the values of the input matrix (Goodfellow et al., 2016).

To perform the convolution, filter matrix moves through the input matrix at a pre-defined rate. At each shift, the elements between the filter and the input matrix are multiplied in the region of intersection, adding the products and storing the result of this operation in a corresponding position in a new matrix. After all displacements are complete, the resulting matrix becomes the feature map that will be passed on to the subsequent layer. Figure 3 shows an example of convolution operation, showing the first, second and last displacement.

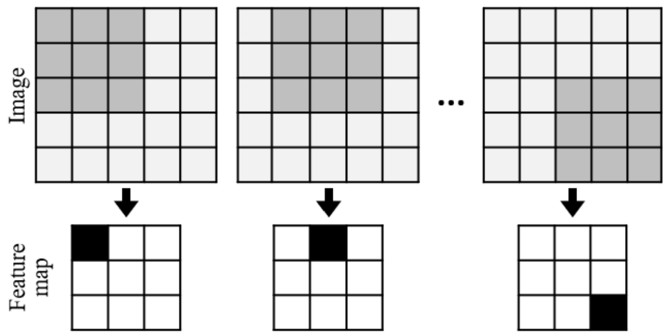

Figure 3. Example of convolution operation, showing the first, second and last displacement

Due to the possibility of obtaining irregular or invalid values in the characteristics map (e.g., negative values for images), an additional step of regularizing their values can be applied. One option is to use ReLU (Rectified Linear Unit), which act through an activation function that aims at zeroing negative values (Zeiler and Fergus, 2014). After the convolution step, a subsampling one is performed with the obtained characteristic maps. The aiming is to reduce the dimensionality of the matrices, while trying to preserve the most relevant aspects. To perform the subsampling, a matrix is used and traverses the characteristics map in a similar way to the filters of the previous step, averaging between neighbouring elements or storing only the element with the highest value in a given region. By only preserving the most characteristic elements and removing those that normally have little or no value, the matrices become less susceptible to small changes in their initial value, increasing the generalization capacity of the network. In the case of images, this helps to preserve the characteristics of objects regardless of their position (Nielsen, 2015). Figure 4 shows the convolution and subsampling processes applied to the image of an animal. Initially, the image has three-color channels that can be explored by different network filters. Each filter seeks to prioritize certain details of the animal to capture different characteristics of the image, thus generating characteristic maps. The maps are regularized by ReLU to adjust possible negative pixels. Then, the maps go through subsampling operations that aim to preserve the most relevant points captured by the filters. This makes it easier for the network to recognize the animal independently of its position in the image. These processes are repeated until the number of convolutional layers is exhausted.

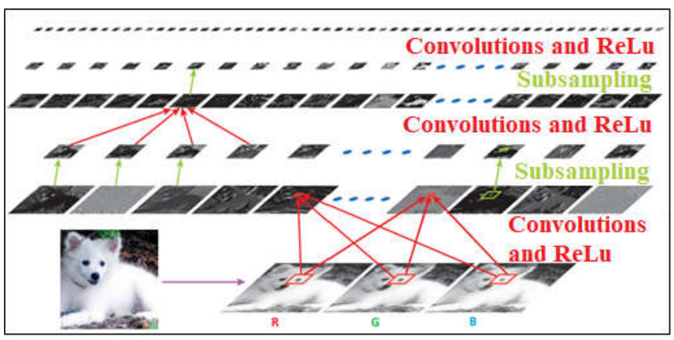

Figure 4. Convolution and subsampling processes applied to an image [adapted from Lecun et al (2015)]

The final step occurs along specific layers to carry out the classification of the characteristic maps obtained previously in one of the possible outputs of the network. These classification layers can work in a similar way to an MLP network, in which all neurons are connected, sending forward signals based on an activation function. During the training stage, this process is repeated in an iterated way through the backpropagation algorithm. So, the network aims to correct the values of its filters and internal weights of its activation neurons to classify the data correctly.

Auxiliary block: Based on the CNN architecture, several papers have proposed the use of auxiliary blocks or layers of CNN to increase the capacity of representation related to the facial expression (Li and Deng, 2020). For this purpose, the CNN HoloNet architecture (Yao et al., 2016) was developed for FER applying CReLU (Concatenated ReLU) activation function (Shang et al., 2016). It was attached to the residual structure to increase the number of deep layers of the network without reducing efficiency, an initial residual block was developed exclusively for FER to learn multiscale resources. Aiming at increasing the degree of supervision for FER, three types of supervised blocks were incorporated early into CNN mainstream hidden layers (Hu et al., 2017): for shallow, intermediate and deep supervision. Figure 5 shows the three supervised blocks. SS block supervises the surface layer, IS block checks the intermediate layer and DS block supervises the deep layer. The blocks were developed according to the ability to describe resources in the layers of the original network.

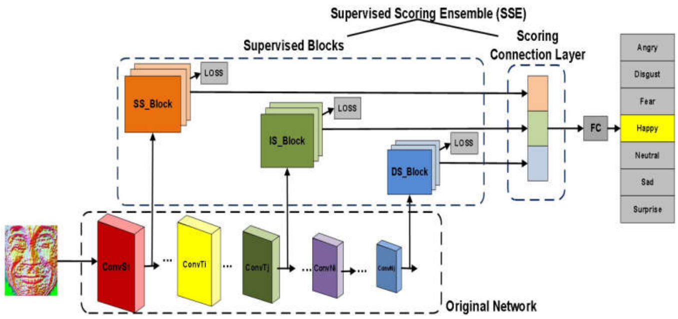

Figure 5. HoloNet architecture with the three supervision blocks (Hu et al., 2017)

Facial Expressions Dataset: In order to create systems capable of identifying facial expressions, a database with samples of different facial expressions is required. Variety in the actors is also important, since degrees of expressiveness as well as ethnicity and features can interfere with the detection of emotions (Valstar et al., 2012). In this sense, the most significant databases found in the literature are JAFFE, FERPlus, and AffectNet, described next. JAFFE, the Japanese Female Facial Expression dataset, is commonly used in facial expression recognition jobs. This dataset was developed in a laboratory for the analysis and testing of the work methodology and includes 213 images from 10 Japanese women (Barman and Dutta, 2019). The emotions described in this dataset are neutral, joy, sadness, surprise, anger, disgust and fear. Figure 6 shows one example of expression of feelings (*anger*) for an individual.

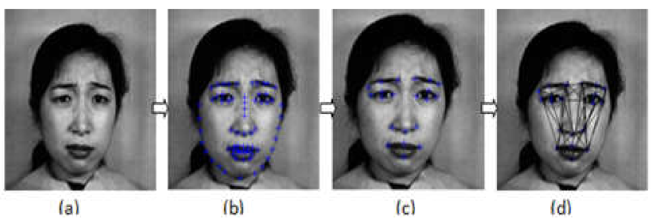

Figure 6. Example of landmarks detection for *anger* in JAFFE dataset (Barman and Dutta, 2019)

The person's photo has only one-color tone and has a resolution of 256x256 pixels. For detecting each emotion in the original image (a), a set of processes is necessary for identifying the corresponding landmarks (b, c and d). FERPlus (Barsoum et al., 2016) is an enhanced version of FER2013, a dataset published after a challenge from the Kaggle platform that took place in 2013 [6]. It has more than 30,000 images of in-the-wild faces that are divided into training, testing and validation folders. The dataset has already been applied in several works associated with the recognition of facial expressions (Pramendorfer and Kampel, 2016; Barsoum et al., 2016). Currently it is one of the most recognized datasets in the literature for the application in recognizing feelings from publicly available facial expressions. The improvement in FER2013 to create FERPlus comprises distributing the probability of emotions for each image. Figure 7 shows some new labels that FERPlus brings to different emotions.

Figure 7. Example of FERPlus images (Barsoum *et al.*, 2016)

Affect Net (Mollahosseini et al., 2017) is the largest publicly available FER dataset on the internet. It has more than a million images, obtained from search engines using tags related to each emotion. The database comprises images referring to two kinds of information, categorical and dimensional. Categorical images represent information related to the seven main discrete facial expressions (*Happy*, *Sad*, *Surprise*, *Fear*, *Disgust*, *Anger*, and *Contempt*). The dimensional information refers to the intensity of valence and arousal. Beyond 450,000 images for the seven discrete facial, the data set includes images labelled as *Neutral*, *None*, *Uncertain*, and *Non-Face*. Figure 8 shows some examples of AffectNet dataset images.

Figure 8. Example of AffectNet dataset images (Mahoor, 2017)

METHODOLOGY

Literature review is a way of identifying, evaluating and interpreting all available and relevant studies for a specific research question or subject or phenomenon of interest (Kitchenham and Charters, 2007).

The most common reasons for conducting a literature review are to summarize the evidence on technology treatment; identify gaps in current research; and provide a background for appropriately positioning new research activities. The methodology applied in this work consists of: *(i)* delimiting the research questions; *(ii)* defining the repositories and search strategies; and *(iii)* selecting the relevant papers. This systematic review is driven by the research questions: *(i)* What are the most used computer vision techniques for FER? *(ii)* What datasets were used in the found approaches? *(iii)* How to associate facial expression patterns to sentiment analysis?

Repositories and Search Engines: To answer these questions, four research repositories were accessed: IEEE (www.ieeeexplore. ieee.org), Science Direct (www.sciencedirect.com), DIVA (www.diva-portal. org), and ACM (www.acm.org). For searching the defined repositories, a non-redundant quest expression was derived from the keywords. The goal was to include the largest number of papers with a filter that returns the most relevant results on the subject. The search expressions applied were: ((“facial expression recognition” OR FER OR EMFACS OR “Emotional Facial Action Coding System”) AND (“convolutional neural networks” OR CNN OR “deep learning” OR “machine learning”)).

keywords, summary, introduction and conclusion. Table 1 shows the number of papers discovered and selected per repository, and Table 2 presents the selected papers, identifying the authors, the repository and the publication year, and first two research questions. From the 10 selected papers, 9 applies CNN for composing the learning architecture. We focus on the characteristics of the different CNN architecture models, their accuracies and FER datasets. For the sake of space limitations, five papers are presented by considering a trade-off between the best accurate results with the smaller processing time are next presented. Shao and Qian (2019) approach adopts three CNN models: *(i)* a shallow network, named the Light-CNN, which is a fully convolutional neural network consisting of six depth wise separable residual convolution modules to solve the problem of complex topology and overfitting; *(ii)* a dual-branch CNN which extracts traditional Local Binary Patterns and deep learning features in parallel; and *(iii)* a pre-trained CNN which is designed by transfer learning technique to overcome the shortage of training samples. The evaluations are performed on three popular datasets (public CK+, multi-view BU-3DEF and FER2013, including 7 basic emotion categories). Table 3 displays the respective accuracy and datasets models. The best results were obtained with the CK+ dataset.

Table 1. Discovered and selected papers

Repository	Discovered	Selected
ACM	6	2
IEEE	11	2
ScienceDirect	10	5
DIVA	2	1

Table 2. List of selected papers

Reference ID	Repository	Technique	Dataset
Shao and Qian (2019)	ScienceDirect	CNN	CK+, BU-3DFE and FER2013
Barsoum et al. (2016)	ACM	CNN	FER+
S. Xie et al. (2019)	ACM	CNN	FER2013, CK+ and Oulu-CASIA
Yang et al. (2017)	ScienceDirect	CNN	CK+, JAFFE and Oulu-CASIA
Phan-Xuan et al. (2019)	ScienceDirect	CNN	FER2013
Nordén and Marlevi (2019)	DIVA	SVM and CNN	FER2013 and JAFFE
W. Xie et al. (2019)	ScienceDirect	CNN	CK+, JAFFE, TFEID and FER2013
Barman and Dutta (2019)	ScienceDirect	MLP and CNN	CK+, JAFFE, MMI and MUG
Zeng et al. (2018)	IEEE	DSAE	CK+
Li and Deng (2020)	IEEE	CNN	CK+, MMI, Oulu-CASIA, JAFFE, FER2013, AFEW, Multi-PIE, BU-3DFE, EmotionNet, RAF-DB, AffectNet, ExpW, 4DFAB

Table 3. Shao and Qian (2019) results (accuracy)

Dataset CNN	CK+	BU-3DFE*	FER2013
Light-CNN	92.86	86.20	68.00
Dual-branch CNN	85.71	48.17	54.64
Pre-trained CNN	95.29	86.50	71.14

* Binghamton University 3D Facial Expression

Selection of State-of-the-Art Works: The inclusion criteria for the state-of-the-art selection were: *(i)* papers published between Jan/01/2016 and Jan/25/2019; *(ii)* papers that matches the search expression in title, abstract or keywords; and *(iii)* papers in English. The exclusion criteria were: *(i)* short papers; *(ii)* conferences without public review committee; and *(iii)* papers that do not feature the use of artificial neural networks. Exclusion criteria *(iii)* was used to select the latest relevant documents on the topic. On the basis of these criteria, 28 papers related to the research question were identified. After that, a set of 10 papers remained.

RESULTS ANALYSIS

When applying the search expression and using the inclusion and exclusion criteria in each repository, 29 papers were discovered. Then, the selection of these papers was made, analysing the title,

For Barsoum et al. (2016), to discriminate emotion based on appearance is essentially an image classification problem. In this sense, it is reasonable to expect that a state-of-the-art CNN model that performs well in image classification should also perform well in facial expression recognition. The authors adopt a Visual Geometry Group (VGG) network to study emotion recognition performance on the FERPlus data set. VGG is a very deep convolutional networks for large-scale image recognition network. The input to the emotion recognition model is a grey scale image at 64×64 resolution. Table 4 shows the respective accuracy and dataset for different variants of VGG: majority voting (MV), multi-label learning (ML), probabilistic label drawing (PLD) and cross-entropy loss (CEL). The resulting accuracies are close, with a slight superiority of VGG-PLD and VGG-CEL models. Both variants of CNN proved to be quite efficient in the training times, what points out the possibility of selecting the results taking into account both the accuracy and the processing efficiency. It is also possible to improve this model accuracy introducing some features in the CNN middle layers.

Table 4. Barsoum et al. (2016) results (accuracy)

CNN \ Dataset	FERPlus
VGG-MV	83.60
VGG-ML	83.69
VGG-PLD	85.43
VGG-CEL	85.01

S. Xie et al. (2019) present a novel model, named Deep Attentive Multipath Convolutional Neural Network (DAM-CNN), for FER. Unlike from most existing models, DAM-CNN can automatically locate expression-related regions in an expressional image and yield a robust image representation for FER. The proposed model contains two novel modules: an attention-based Salient Expressional Region Descriptor (SERD) and the Multi-Path Variation-Suppressing Network (MPVS-Net). SERD can adaptively estimate the importance of different image regions for FER task, while MPVS-Net disentangles expressional information from irrelevant variations.
DAM-CNN consists of three parts, including the feature extraction module, the attention-based Salient Expressional Region Descriptor (SERD) and the Multi-Path Variation-Suppressing Network (MPVS-Net). The model consists of three modules, i.e., the VGG-Face network for extracting features, SERD for refining CNN features and highlighting salient expressional regions, and MPVS-Net for generating a high-level representation robust to multiple variations. Table 5 shows the respective accuracy and dataset adopted.

Table 5: Effect of SERD by recognition accuracy

CNN \ Dataset	CK+	JAFFE	FER2013
VGG-Face	93.59	95.05	62.99
VGG-SERD	93.98	96.37	62.22

In this approach, the authors used features in intermediate layers, and obtained a higher accuracy than that presented by Barsoum et al. (2016). The results obtained with the refinement with the use of the SERD layer were also significant. Yang et al. (2017) proposed WMDNN, a weighted mixture deep neural network for automatically extracting features. WMDNN outperforms the FER state-of-the-art based on hand-crafted features or deep networks using one channel. Compared with deep networks that use multiple channels, the network proposed hereby can achieve comparable performance using instead easier procedures. Fine-tuning in WMDNN is effective for FER tasks with an already trained model if sufficient samples cannot be collected. The authors argue that accurate hand-crafted features with high correlation to changes in expression are difficult to extract due to the influences of individual differences and variations in emotional intensity. Features for describing accurately changes in facial expressions are required. CNN for feature extraction in WMDNN is based on the VGG16 network of Simonyan and Zisserman (2014). VGG16 is chosen due to its effective performance in visual detection and fast convergence. A shallow CNN is constructed to automatically extract features of facial expressions. Table 6 shows the obtained accuracy results in two datasets.

Table 6. Yang et al. (2017) results (accuracy)

CNN \ Dataset	CK+	JAFFE
VGG16	97.02	92.21

In addition to the CK+ and JAFFE datasets, the authors explored Oulu-CASIA dataset, however, with no significant results. So, it is not presented in this review. Li and Deng (2020) paper is characterized by extensive analysis of FER's algorithms and datasets. They argue that current Deep Learning algorithms for FER commonly focus on two important issues, namely: (*i*) overfitting due to the lack of training data and (*ii*) an unrelated variation such as expressions, and problems related to lighting, head posture, among others. They describe how the process of identifying the characteristics of an image is done. For the state of the art in FER, techniques applied are presented to facilitate the recognition of facial expressions in images and videos, based on training with static images. Among the datasets tested the best results were obtained with Affect Net. Without presenting numbers, they state that the accuracy results achieved with the auxiliary blocks, network ensemble and multitask network were significant. Regarding efficiency, they achieved a range of results, using auxiliary blocks and multitask networks, varying since from irrelevant to good. They also tested with an ensemble, with weak results.

Final Remarks: This research purpose is to provide the background for developing assistive applications applying facial emotion analysis. It was first studied the best combinations of deep learning algorithms and datasets for FER. The ongoing work includes *(i)* testing the most promising algorithms against static image datasets and *(ii)* the application of these techniques in real world situations on the healthcare domain.

REFERENCES

Barman, A., Dutta, P., 2019. Facial expression recognition using distance and texture signature relevant features. *Applied Soft Computing Journal*, 77:88-105.

Barsoum, E., Zhang, C., Ferrer, C. C., Zhang, Z., 2016. Training deep networks for facial expression recognition with crowd-sourced label distribution. *18thACM International Conference on Multimodal Interaction* (ICMI '16), pp. 279-283.

Ekman, P., Friesen, W. V., 1978. *Facial action coding system*. Consulting Psychologist Press.

Goodfellow, I. Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press.

Hu, P., Cai, D., Wang, S., Yao, A., Chen, Y., 2017. Learning Supervised Scoring Ensemble for Emotion Recognition in the Wild. *19th ACM International Conference on Multimodal Interaction* (ICMI'16), pp. 553-560.

Humphries, C., Binder, J. R., Medler, D. A., Liebenthal, E., 2006. Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension. *Journal of Cognitive Neuroscience*, 18(4):665-679.

Kitchenham, B., Charters, S., 2007. *Guidelines for performing systematic literature reviews in software engineering*. EBSE Technical Report.

Konam, S., Quah, I., Rosenthal, S., Veloso, M., 2018. Understanding convolutional networks with APPLE: Automatic Patch Pattern Labeling for Explanation. *AAAI/ACM Conference on AI, Ethics, and Society*.

Krizhevsky, A., Sutskever I., Hinton, G. E., 2017. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84:90.

Leo, M., Carcagnì, P., Mazzeo, P. L., Spagnolo, P., Cazzato, D., Distante, C., 2020. Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches. *Information*, 11:128.

Li, S., Deng, W., 2020. Deep facial expression recognition: a survey. *IEEE Transactions on Affective Computing*. DOI: 10.1109/TAFFC.2020.2981446

Mahoor, M. H., 2017. AffectNet. http:// mohammadmahoor. com/affectnet/

Mehta, D., Siddiqui, M. F. H., Javaid, A. Y., 2018. Facial Emotion Recognition: A survey and real-world user experiences in mixed reality. *Sensors*, 18(2):416.

Michaud, T., Gassia, V., Belhaouari, L., 2015. Facial dynamics and emotional expressions in facial aging treatments. *Journal of Cosmetic Dermatology*, 14(1):9-21.

Mollahosseini, A., Hasani, B., Mahoor, M. H., 2017. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10:18-31.

Nielsen, M. A., 2015. *Neural Networks and Deep Learning*. Determination Press.
Nordén, F., Marlevi, F. R., 2019. A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition. G4A: Facial Expression Classification.
Phan-Xuan, H., Le-Tien, T., Nguyen-Tan, S., 2019. FPGA Platform applied for Facial Expression Recognition System using Convolutional Neural Networks. *Procedia Computer Science*, 151:651-658.
Pramendorfer, C., Kampel, M., 2016. *Facial expression recognition using convolutional neural networks: State of the art*. https://arxiv.org/pdf/1612.02903.pdf
Shang, W., Sohn, K., Almeida, D., Lee, H., 2016. Understanding and improving convolutional neural networks via concatenated rectified linear units. *33rd International Conference on Machine Learning*, pp. 2217-2225.
Shao, J., Qian, Y., 2019. Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing*, 355:82-92.
Simonyan K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *3rdInternational Conference on Learning Representations*.
Toy, B. L., 2014. Applying behavior models in a system architecture (Applying an engineering protocol structure to AI). *IEEE Symposium on Computational Intelligence for Human-like Intelligence* (CIHLI), Orlando, FL, pp. 1-10.
Valstar, M. F., Mehu, M., Jiang, B., Pantic, M., Scherer, K., 2012. Meta-analysis of the first facial expression recognition challenge. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4):966-979.
Xie, S., Hu, H., Wu, Y., 2019. Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. *Pattern Recognition*, 92:177-191.
Xie, W. Jia, X. Shen L., Yang, M., 2019. Sparse deep feature learning for facial expression recognition. *Pattern Recognition*, 96.
Yang, B., Cao, J., Ni, R., Zhang, Y., 2017. Facial Expression Recognition using Weighted Mixture Deep Neural Network Based on Double-channel Facial Images. *IEEE Access*, 6:4630-4640.
Yao, A., Wang, S., Cai, D., Sha, L., Hu, P., Chen, Y., 2016. HoloNet: Towards Robust Emotion Recognition in the Wild. *18th ACM International Conference on Multimodal Interaction* (ICMI'16), pp. 472-478.
Zeiler, M. D., Fergus, R., 2014. Visualizing and Understanding Convolutional Networks. *European Conference on Computer Vision*, pp. 818-833.
Zeng, N., Zhang, H., Song, B., Liu, W., Li Y., Dobaie, A. M., 2018. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing*, 273:643-649.
Zharovskikh, A., 2020. *Facial Recognition for Healthcare Disruption. Key Use Cases*. InData Labs. https:// indatalabs.com/blog/ai-face-recognition-in-healthcare.

****