UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Visión Electrónica
Más que un estado sólido

https://doi.org/10.14483/issn.2248-4728

VISIÓN ELECTRÓNICA

A RESEARCH VISION

Classification of Facial Expression of Post-Surgical Pain in Children: Evaluation of Convolutional Neural Networks

Clasificación de la expresión facial de dolor postquirúrgico infantil: Evaluación de redes neuronales convolucionales

Carolina Jiménez-Moreno [1], *Jenny Kateryne Aristizábal-Nieto* [2], *Olga Lucía Giraldo-Salazar* [3]

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:
Enviado: 09/11/2020
Recibido: 10/11/2020
Aceptado: 22/12/2020

Keywords:
Artificial intelligence
Assessment tools
Facial expression
Pain
Pediatrics

Palabras clave:
Inteligencia artificial
Herramientas de evaluación
Expresión facial
Dolor
Pediatría

ABSTRACT

There are certain difficulties in differentiating between children's facial expression related to pain and other stimuli. In addition, the limited communication ability of children in the preverbal stage leads to misdiagnosis when the child feels pain, for example, post-surgical conditions. In this article, a classification approach of facial expression of child pain is presented based on models of pre-trained convolutional neuronal networks from the study carried out in a Colombian hospital of level 4 (Hospital Universitario San Vicente Fundación), in the recovery areas of child surgery services. AlexNet and VGG (16, 19 and Face) networks are evaluated in the own dataset using the FLACC scale and their performances are compared in three experiments. The results show that the VGG-19 model achieves the best performance (92.9%) compared to the other networks. The effectiveness of the model and transfer learning for the classification of facial expression of child pain shows a promising solution for the assessment of post-surgical pain.

RESUMEN

Existen ciertas dificultades para diferenciar entre la expresión facial infantil relacionada al dolor con la de otros estímulos. Además, la limitada capacidad de comunicación de los niños en la etapa preverbal conlleva a un error de diagnóstico cuando el niño siente dolor, por ejemplo, afecciones posteriores a las cirugías. En este artículo, se presenta un enfoque de clasificación de la expresión facial de dolor infantil basado en modelos de redes neuronales convolucionales pre-entrenadas a partir del estudio realizado en un hospital colombiano de nivel 4 (Hospital Universitario San Vicente Fundación), en las áreas de recuperación de los servicios de cirugía infantil. Se evalúan las redes Alex-Net y VGG (16, 19 y Face) en el conjunto de datos propio utilizando la escala FLACC y se comparan sus rendimientos en tres experimentos. Los resultados muestran que el modelo VGG-19 logra el mejor rendimiento (92.9%) en comparación con las demás redes. La eficacia del modelo y el aprendizaje por transferencia para la clasificación de la expresión facial de dolor infantil muestran una solución prometedora para la evaluación del dolor postquirúrgico.

[1] BSc. In Physics Engineering, Universidad Tecnológica de Pereira, Colombia. MSc. in Engineering, Universidad de Antioquia, Colombia. Current position: Masters student, Universidad de Antioquia, Colombia. E-mail: carolina.jimenezm@udea.edu.co

[2] BSc. in Bioengineering, Universidad de Antioquia, Colombia. MSc. in Science in Biomedical Engineering, Politecnico di Torino, Italy. Current position: Researcher and Professor, Universidad de Antioquia, Colombia. E-mail: jenny.aristizabal@udea.edu.co

[3] BSc. In Telecommunications Medicine, Universidad de Antioquia, Colombia. Specialization in Anesthesiology and Resuscitation, Universidad de Antioquia, Colombia. MSc. in Clinical Epidemiology, Universidad de Antioquia, Colombia. Current position: Professor, Universidad de Antioquia, Colombia and Anesthesiologist, Hospital Universitario San Vicente Fundación, Colombia. E-mail: lucia.giraldo@udea.edu.co

Cite this article as: C. Jiménez Moreno, J. K. Aristizábal Nieto, O. L. Giraldo Salazar, "Classification of Facial Expression of Post-Surgical Pain in Children: Evaluation of Convolutional Neural Networks", *Visión Electrónica*, vol. 15, no. 1, pp. 7-16, 2021. https://doi.org/10.14483/22484728.17425

1. Introduction

The manifestation of pain has a great impact on the patient's environment and on him or herself, even more so when the pain is not well controlled. For this reason, optimum communication is needed between the treating personnel and the patient to make the correct interpretation of the pain at the time of the medical intervention, thus evaluating the intensity of the pain to provide analgesics and formulate the respective diagnosis. This is a starting point that cannot be replaced by advances in pharmacology and technology [1], [2].

The painful experience of each person depends on their personal and subjective value based on age, culture, previous experience, context-derived senses, among other factors. For this reason, no two people experience pain under the same physiological conditions and mechanisms. This is a problem for the health personnel involved in pain management, since the evaluation of pain intensity depends on their criteria as well as on the patient's verbal report, and there are not strictly objective and precise measures to establish the degree of pain suffered by the patient. Such an assessment is complicated when dealing with children or people with limited ability to communicate.

The main problem that arises in the management of pediatric pain is the assessment and self-perception of it [3], [4]. Pain in children has been associated with physiological changes and behavioral patterns, which are indicators of pain that can be recorded and therefore quantified [5] - [7]. In that sense, it is evidenced in the literature, the development of several traditional scales of pain assessment to estimate the intensity of this.

Patient self-assessment is the most reliable and valid measure for assessing pain. The patient can express the intensity of his or her pain and the location of the pain. However, it is not possible to use it in people with communication or neurological impairments or in infants [8], [9], since they cannot quantify its severity and inform medical personnel about the effectiveness of the analgesia [4], [5]. To evaluate pain in children, the indicators summarized in Table I, which are related to pain, were defined. It is necessary to emphasize that changes in the child's facial expression in response to pain are considered the most reliable and consistent indicator [10].

Table 1. Indicators that determine the presence of pain in children [4], [10]–[12].

Children's response to pain			
Physiological Changes		**Behavioral Patterns**	
Vital Signs	Blood pressure	**Behavior**	Changes in facial expression
	Heart rate		Movement of the legs
	Breathing rate		Crying
	Oxygen saturation		Frequent body movements

In young children, verbal skills remain limited and quite inconsistent. Pain-related behaviors are the main indicator for assessments in this age group. Nonverbal behaviors, such as facial expression, limb movement, grasping, and crying, are considered more reliable and objective measures of pain than self-evaluation. The most used pain assessment scales for this age group are [12]: *The Children's Hospital of Eastern Ontario Pain Scale (CHEOPS), Face Legs Arms Cry Consolability (FLACC), COMFORT Scale, The Observational Scale of Behavioral Distress (OSBD), Observational Pain Scale (OPS), y The Toddler-Preschooler Postoperative Pain Scale (TPPPS).*

One possible way to provide an objective and continuous assessment of pain is to develop an automated system that observes and analyzes different behavioral/physiological indicators related to pain [13], [14].

It is for these reasons that there is interest in having new techniques and strategies that allow doctors and nurses to better diagnose postoperative pain and identify its levels.

Recent innovations in the field of computer vision have facilitated the development of automated approaches to evaluating facial expressions. In order to minimize errors in the recognition of facial expressions given by complex image backgrounds, techniques such as the background subtraction technique are used in [15], where authors explore the potential of this technique, which allowed them to design and implement a motion detection algorithm. Some other works have considered the use of the image segmentation technique [16].

The increase of computing power in GPUs and the creation of large image data sets have allowed convolutional neural networks (CNN) to show an outstanding performance in the challenges of computer vision, as evidenced in [17].

Thus, the present study was based on the premise that the incidence of severe pain in post-surgical patients of moderate to severe intensity is high, and that facial expressions are considered the fundamental pillar in the evaluation of pain since they constitute one of the most significant pain indicators [18]. Therefore, it is proposed to **evaluate different architectures of convolutional neural networks (CNNs), widely used in the recognition of emotions, for the classification of facial expression of child pain.**

The paper is organized as follows. In the second chapter, the methods used for the construction of the data set are described. The implementation is described in detail in the third chapter. The experimental results are discussed in the fourth chapter. Finally, the conclusions are presented in the fifth chapter.

2. Methodology

2.1. Definition of the population

A proprietary data set was built with images of pediatric patients from the Hospital Universitario San Vicente Fundación (HUSVF) in Medellín, Colombia. This study was approved by the Ethics and Research Committee of the HUSVF and by the Biomedical Committee of the University of Antioquia (UdeA), Medellín, Colombia.

The sample size was defined for 50 pediatric patients (39 boys and 11 girls), who were registered after undergoing surgical procedures such as general pediatric surgery, orthopedic surgery, or plastic surgery. The average age of the children is 16.84 months, which varies from 1 to 36 months (standard deviation = 10.58). Any child who received surgery and whose age was within the range was eligible for data recording, after obtaining the respective informed consent of the child's parents and/or guardians. Children with neurological diseases and facial dimorphism or with a facial handicap were excluded.

2.2. Image acquisition

The integrated camera of an iPad Mini 4 was used to record videos of the children's facial expression and the FLACC (Face, Leg, Activity, Cry, Consolability) pain assessment scale was used to record changes in body movement. All recordings were made in the HUSVF postoperative clinical setting.

Each child was recorded for four time periods: 1) Right after the surgical procedure for the first observation (take ZERO); 2) Ten minutes after the surgical procedure (take ONE); 3) Twenty minutes after the surgical procedure (take TWO); and 4) Thirty minutes after the completion of the painful procedure (take THREE). Each period was observed by trained Nurse Practitioners and Anesthesiology Residents to provide pain assessment using the FLACC scale and to perform vital sign measurements (variation in blood pressure, heart rate and oxygen saturation), which helped to supplement the assessment.

2.3. FLACC Pain Assessment Scale

There is no universally accepted standard measurement instrument for assessing and measuring childhood pain. The basic principle of pain measurement is to choose the right instrument for the right patient, which means that it should be based on developmental age and the type of pain or medical condition (i.e., procedural pain versus postoperative pain) [12]-[14], [19]-[21].

For these reasons, the FLACC pain assessment scale was used since according to the literature [14], [19], [20], [22]-[23] it is the recommended and best validated scale for assessing postoperative pain in infants because it is reproducible and simple to use in a clinical setting and it assesses the child's face, leg movement, body activity, crying, and whether he is easy to comfort, which are all observable variables associated with pain.

Each component of the FLACC scale is scored between 0 and 2 points, with 0 being an overall indicator that the child is calm, 1 being very restless, and 2 being desperate. A total of 1 to 3 points represents mild pain, 4 to 6 points represents moderate pain, and 7 or more points represents severe pain.

The scores obtained (0, 1, 2, 3, 4 and 5) for each of the recorded shots in this study were used as the label for the evaluation of the models. It should be noted, the exclusion of the records labeled with the scores from 6 to 10, since the number of valid records and images was too small.

3. Implementation

The proposed process for the classification of pain expression consists of two main stages: 1) pre-processing of the images and 2) adjustment and training of pre-trained CNN architectures. Each stage is described in detail below.

3.1. Image pre-processing

Before starting to pre-process the images, the OpenCV Library [26] was first used to extract the frames from each of the 200 videos purchased with the iPad Mini 4's built-in camera. The next step was to implement the Oriented Gradient Histogram (HOG) descriptor offered by the DLIB library [27], [28], in each of the extracted images to detect the face. Images where faces are not detected by the algorithm were excluded from further analysis. In addition, a correlation analysis was performed to verify the matching of the images and to process only the relatively different images, thus selecting only the key frames of each video. Using again the DLIB library [29], 68 facial reference points were obtained which allow to identify the sketch of the face, eyebrows, eyes, nose, and mouth. From the information of the coordinates of the sketch of the face and the eyebrows, a mask was created to segment the face from the background of the image.

Since the total number of frames is too small (i.e., 2730 frames) to retrain a CNN and to make the model robust to the characteristics that the set of images might have, such as the angles of the shots, the illumination, the similarity of the images, among others, a series of transformations to the set of training and validation images were performed as follows.

First, the images were flipped horizontally at random with a 50% probability. These types of transformations are optimal for the data set since the facial expressions, in these cases, of the babies and children, are quite symmetrical. Following this, the images were resized to a size of 256xN, with N being the ratio of dimensions of the images. And finally, each image was cropped to the size 224x224 and normalized ([0.0996611, 0.0800176, 0.06390216], [0.16571397, 0.14057845, 0.12316495]) to scale the image values in the range of [0.1].

3.2 Adjustment and training of pre-trained CNN architectures

Four CNNs architectures were used for pain classification in the relatively small data set (50 subjects, 2730 images). The first three architectures, as seen in Figure 1, AlexNet, VGG16, and VGG19, were previously trained on the ImageNet dataset [30] which contains more than 1.2 million images for the 1000 class classification.

The AlexNet architecture, which has five convolutional layers and three fully connected layers, has promoted the development of deep learning in the field of facial expression recognition, specifically, emotion recognition. In the present study, the last connection layer was modified to change from classifying 1000 classes to 6 classes corresponding to the scores obtained by the FLACC scale (0, 1, 2, 3, 4, and 5).

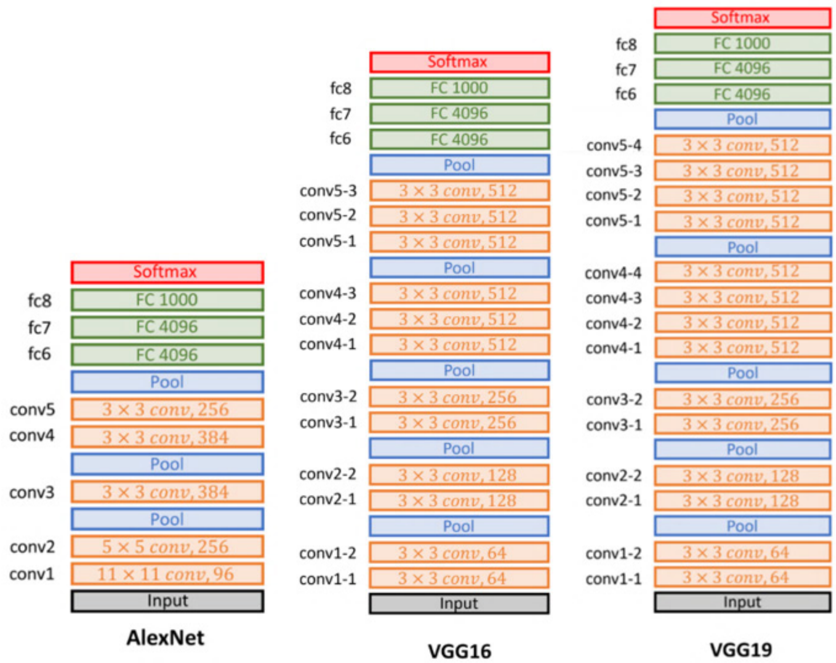
The fourth architecture is the VGG-FACE descriptor. The VGG family of architectures, as shown in Figures 1 and 2, have the same structure in the first three sets of convolutional layers, and the overall structure contains five sets of convolutional layers. The VGG16 and VGG19 networks are also widely used for the task of emotion classification. In the present study, the last connecting layer of both networks was modified from 1000 to 6 classes. The VGG-FACE network was previously trained on a large set of face images [31], which contains approximately 2.6 million face images to classify 2622 identities in the face recognition task. The output neurons of the last layer (fc8 layer) were also replaced by 6 classes.

The choice of these pre-trained CNNs allows to investigate the difference between using networks trained on a relatively similar data set (i.e., VGG-FACE, Face Dataset) and networks trained on a relatively different data set (i.e., AlexNet, VGG16 and VGG19, ImageNet) to one's own data set.

The image set was randomly divided into training set, validation set, and test set. The test data set was used to select the best classifier, where the loss function would reach the minimum.

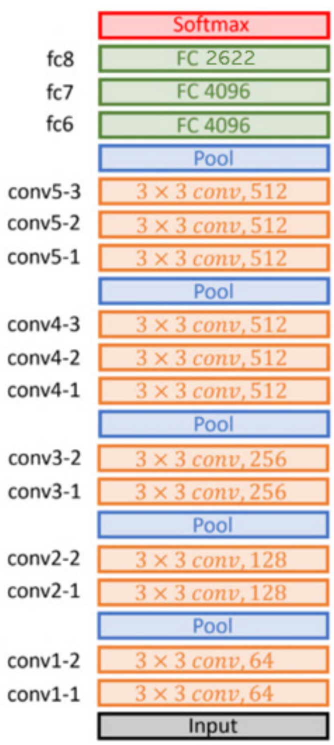
The challenge that arose in the development of this study was the limited number of face images. The proposed solution was to use learning by transfer [32], [33] to address the problem of limited availability of tagged data. By making use of this technique, it is possible, as a first option, to preserve all the previously trained layers before the last output layer and to connect these intermediate layers to a new layer designed for the new classification problem.

Figure 1. Architectures AlexNet, VGG-16 and VGG-19.



Source: own.

Figure 2. Architecture VGG-Face.



Source: own.

The second option is to adjust more layers, or even the entire set of pre-trained network layers. It is also possible to keep the first convolutional layer fixed, as this layer is often used for edge extraction, which is common for generic image processing problems.

In this study, it was decided to adjust all the parameters of each of the pre-trained models. In addition, to obtain an unbiased evaluation in the ratings, three experiments were used to evaluate the performance of the model by adjusting the hyperparameters. The selection of the hyperparameters was based on several studies [34]-[40] focused on the area of facial expression recognition, emotions, and pain. In these studies, the policy of updating the stepwise learning rate is developed, specifying the values for each of the hyperparameters and the combination of these.

The total number of times for training was 100. The training algorithm applied was the stochastic gradient descent with the hyperparameters defined in Table 2 (momentum, weight decay, initial learning rate). The learning rate was reduced by a specific gamma factor every certain number of times established by the size of the steps. A lot size of 32 and 16 was used for the training and validation set, respectively.

Table 2. Hyperparameters established by each proposed experiment.

Hyperparameters \ No Experiments	Initial learning rate	Gamma factor	Momentum	Step size	Decay of weights
Experiment 1	0.000001	10.0	0.99	5.0	0.0005
Experiment 2	0.001	10.0	0.9	5.0	0.0005
Experiment 3	0.001	0.1	0.9	5.0	0.0005

Source: own.

The entire data set was randomly divided into a training set (±50%, 1388 frames), a validation set (±20%, 529 frames), and a test set (±30%, 813 frames).

4. Analysis of results

To classify the facial expression of pain of babies and children, a total of 2730 facial images were entered as input to the four CNNs architectures mentioned above for the final classification. All the networks were implemented in the Google Collaboratory environment using the Python programming language and the PyTorch library. Training performance was reported using accuracy and loss.

The first column of results in Table 3, reports the performance of the pain assessment by applying the hyperparameters established for experiment #1 (Initial learning rate: 0.000001, Gamma factor: 10.0, Time: 0.99, Step size: 5.0, Weight Decay: 0.0005). The evaluation of the networks was carried out in this way because we wanted to evaluate how the choice of the hyperparameters affects the performance of the classification. The AlexNet network, for experiment #1, had the best performance, obtaining good accuracy and relatively low loss.

Table 3. Pain Assessment Performance with AlexNet, VGG-16, VGG-19 and VGG-FACE.

	Experiment #1		**Experiment #2**		**Experiment #3**	
	Accuracy	***Loss***	***Accuracy***	***Loss***	***Accuracy***	***Loss***
AlexNet	**0.641**	**0.238**	0.679	0.199	0.583	1.106
VGG-16	0.171	2.355	0.937	1.075	**0.623**	**0.342**
VGG-19	0.344	1.612	**0.929**	**0.062**	0.536	1.223
VGG-FACE	0.295	3.126	0.836	0.178	0.468	1.136

Source: own.

The second column of Table 3 shows the performance of the pain assessment by applying the hyperparameters established for experiment #2 (Initial learning rate: 0.001, Gamma factor: 10.0, Time: 0.9, Step size: 5.0, Weight Decay: 0.0005). Comparing the performance of the first and second column the hyperparameters chosen for experiment #2 significantly improved the overall performance for each of the models. The accuracy of the pain assessment was improved for the VGG16, VGG19 and VGG-FACE networks. However, the loss obtained for the VGG-16 network was too high. Therefore, it can be concluded that the VGG-19 and VGG-FACE networks achieved the best overall performance, since they have high accuracies and low losses, with VGG-19 being the best option.

The last column of Table 3 provides the performance of the networks using the hyperparameters chosen for experiment #3 (Initial learning rate: 0.001, Gamma factor: 0.1, Time: 0.9, Step size: 5.0, Weight Decay: 0.0005). The VGG-16 network, for experiment #3, had the best performance, obtaining good accuracy and relatively low loss.

The next step was to evaluate the VGG-19 model, in the test phase, on each of the classes in the data set, the statistical accuracy metrics, given in equation (1); sensitivity (True Positive Rate - TPR), equation (2); specificity (True Negative Rate - TNR), equation (3); false positive rate or 1-Sensitivity (False Positive Rate - FPR), equation (4) and false negative rate (False Negative Rate - FNR), shown in equation (5).

- **Accuracy:** Represents the overall performance of the model through the total percentage of hits.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- **Sensitivity (TPR):** Represents the fraction of positive tests that are correctly labeled.

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

- **Specificity (TNR):** Represents the fraction of negative tests that are correctly labeled.

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (3)$$

- **False positive rate (FPR):** Represents the fraction of negative tests that are incorrectly labeled as positive.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (4)$$

- **False negative rate (FNR):** Represents the fraction of positive tests that are incorrectly labeled as negative.

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}} \quad (5)$$

Where the values of TP, FP, FN, TN are explained graphically with the confusion matrix, which can be seen in Table 4.

Table 4. Confusion matrix for classification.

		Manual classification (reality)	
		Positive	**Negative**
Automatic classification (predictions)	**Positive**	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

Source: own.

Table 5 presents the performance of the VGG-19 network under experiment #2 in the test phase. The purpose of examining these metrics was to evaluate how the selected hyperparameters affect the performance of the classification along with the true positive rate (TPR) and the false positive rate (FPR). As shown in Table 5, particularly good accuracy values are obtained for each of the classes to be predicted by the model. The TNR and FPR values are generally good, however, the TPR values are low and the FNR is remarkably high. It is important to clarify that the purpose is to obtain high values for accuracy, TPR and TNR and to achieve low values for FPR and FNR. Analyzing these last two metrics is crucial in the case of pain assessment, since, in the literature, there are many pediatric studies where overtraining (associated with TPR) and undertraining (associated with NRF) are present. Therefore, it is concluded that the CNN VGG-19 model achieved a good performance in a general way obtaining good specificity, however, it could have suffered from mismatch. In addition to this, it can also be observed that the results obtained for classes 3, mild pain, and 4, moderate pain, are not optimal.

The possible reasons for these results were:

1. The fact of having 50 patients, which is a weakness of the project

2. The pain assessment by the Anesthesiology nurses and resident doctors was not independent, therefore, the categorization of the images could have been biased.

Table 5. Confusion matrix for classification.

	Accuracy	TPR	TNR	FPR	FNR
0 vs All	85.4%	56.1%	94.1%	5.9%	26.1%
1 vs All	86.7%	55.3%	96.3%	3.7%	18.0%
2 vs All	79.5%	34.6%	95.5%	4.5%	26.7%
3 vs All	72.7%	43.7%	12.2%	25.5%	90.28%
4 vs All	89.9%	0%	91.4%	8.6%	100%
5 vs All	99.5%	97.5%	1%	0%	0%

Source: own.

3. Much of the bias may also have occurred because of the design of the data set.

4. The images categorized in classes 3, mild pain, and 4, moderate pain, probably cannot be differentiated well, so the model cannot distinguish between these classes.

5. Conclusions

Assessment of childhood pain can be inconsistent since it depends largely on medical judgment, and medical personnel are required to be well trained to ensure proper use of the assessment scales. This can result in late intervention and inadequate pain management. Because pain assessment is crucial to pain management, automatic tools need to be developed to allow optimal pain assessment.

This work evaluates different convolutional neural network architectures, widely used for the classification and detection of emotions, in the task of automatic pain classification in three different experiments. All networks, AlexNet, VGG16, VGG19 and VGG-FACE were evaluated using the proprietary dataset. The experimental results showed that the selection of hyperparameters influences the performance of the models. The selected hyperparameters for experiment #2 (Initial learning rate: 0.001, Gamma factor: 10.0, Momentum: 0.9, Step size: 5.0, Weight Decay: 0.0005) influenced to obtain the best results with respect to the other two experiments. With the VGG-19 network, the best performance was obtained in comparison with the other networks, achieving an accuracy of 92.9% and a loss of 6.2% for the validation phase. However, when analyzing the accuracy, precision, TPR, TNR, FPR and FNR metrics in the test phase, it could be observed that the model, despite having a good performance at a general level and achieving good specificity, did not achieve good sensitivity and possibly presented training mismatch. The reasons for this are considered to correspond to the distribution of the images in the different classes and/or in the divisions of the training, validation, and test sets. To solve this problem, it is proposed, before improving the technique, to improve the data set, and to make an exhaustive analysis of the error in a manual way, analyzing image by image, taking advantage of the fact that it is a relatively small data set, and to confirm that the labeling of the data has been correct and to carry out the necessary measures, such as merging classes that may not be differentiable. This will possibly help to improve the performance of the model and not suffer from either over- or under-training.

These results are encouraging and suggest that automatic recognition of childhood pain is a viable and more efficient alternative to the current standard of pain assessment. By following the proposed improvements, it is expected to have a robust system capable of classifying the level of child pain with particularly good results, thus solving the problem of biased pain assessment that occurs every day.

Acknowledgments

This work was supported by the Research Group on Bioinstrumentation and Clinical Engineering (GIBIC), the Research Group on Perioperative Medicine (GRIMPA) and the University of Antioquia (UdeA) through the project: 2018-23358 - "Performance and design of a model of post-surgical pain in children under 36 months for the classification of pain compared to the traditional pediatric scale FLACC" and funded by the University of Antioquia (UdeA) through the academic encouragement of Student Instructor.

References

[1] H. Y. Vivian-Ip, A. Abrishami, P. W. H. Peng, J. Wong, F. Chung, "Predictors of Postoperative Pain and Analgesic Consumption: A Qualitative Systematic review", *Anesthesiology*, vol. 111, no. 3, pp. 657–677, 2009. https://doi.org/10.1097/ALN.0b013e3181aae87a

[2] O. L. Elvir-Lazo, P. F. White, "Postoperative pain management after ambulatory surgery: role of multimodal analgesia", *Anesthesiology Clinics*, vol. 28, no. 2, pp. 217–224, 2010. https://doi.org/10.1016/j.anclin.2010.02.011

[3] American Academy of Pain Medicine, "Get the facts on pain". [Online]. Available at http://www.painmed.org/patientcenter/facts-on-pain/

[4] P. J. Mathew, J. L. Mathew, "Assessment and management of pain in infants", *Postgraduate Medical Journal*, vol. 79, no. 934, pp. 438–43, 2003. http://dx.doi.org/10.1136/pmj.79.934.438

[5] M. Clarett, "Escalas de evaluación de dolor y protocolo de analgesia en terapia intensiva", Clínica y Maternidad Suizo Argentina Instituto Argentino de Diagnóstico y Tratamiento, Buenos Aires, Argentina, 2012.

[6] L. J. Duhn, J. M. Medves, "A systematic integrative review of infant pain assessment tools", *Advance in Neonatal Care*, vol. 4, no. 3, pp. 126–140, 2004. 10.1016/j.adnc.2004.04.005

[7] R. Slater, A. Cantarella, L. Franck, J. Meek, M. Fitzgerald, "How Well Do Clinical Pain Assessment Tools Reflect Pain in Infants?" *PLoS Medicine*, vol. 5, no. 6, p. 129, 2008. https://doi.org/10.1371/journal.pmed.0050129

[8] N. C. de Knegt. et al., "Behavioral Pain Indicators in People With Intellectual Disabilities: A Systematic Review", *The Journal of Pain*, vol. 14, no. 9, pp. 885–896, 2013. https://doi.org/10.1016/j.jpain.2013.04.016

[9] G. Zamzmi, "An approach for automated multimodal analysis of infants' pain", 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 4148–4153, 2016.

[10] V. Guruswamy, “Assessment of pain in nonverbal children”, Association of Paediatric Anaesthetists of Great Britain and Ireland, no. 41, p. 33, 2014.

[11] Registered Nurses’ Association of Ontario, “Assessment and management of pain”, vol. 3. Toronto, Canada, 2013.

[12] R. Srouji, S. Ratnapalan, S. Schneeweiss, “Pain in Children: Assessment and Nonpharmacological Management”, *International Journal of Pediatrics*, 2010. https://doi.org/10.1155/2010/474838

[13] K. Brand, A. Al-Rais, “Pain assessment in children”, *Anaesthesia and Intensive Care Medicine*, vol. 20, no. 6, pp. 314–317, 2019. https://doi.org/10.1016/j.mpaic.2019.03.003

[14] D. Freund, B. N. Bolick, “Assessing a Child’s Pain”, *AJN, American Journal of Nursing*, vol. 119, no. 5, pp. 34–41, 2019.

[15] M. Pérez, G. A. Cavanzo Nisso F. Villavisán Buitrago, “Sistema embebido de detección de movimiento mediante visión artificial “, *Visión Electrónica*, vol. 12, no. 1, pp. 97-101, 2018. https://doi.org/10.14483/22484728.15087

[16] J. F. Pantoja Benavides, F. N. Giraldo Ramos, Y. S. Rubio Valderrama, V. M. Rojas Lara, “Segmentación de imágenes utilizando campos aleatorios de Markov”, *Visión Electrónica*, vol. 4, no. 2, pp. 5-16, 2010. https://doi.org/10.14483/22484728.432

[17] J. Forero C., C. Bohórquez, V. H. Ruiz, “Medición automatizada de piezas torneadas usando visión artificial”, *Visión Electrónica*, vol. 7, no. 2, pp. 36-44, 2013. https://doi.org/10.14483/22484728.5507

[18] S. Brahnam, C.-F. Chuang, R. S. Sexton, F. Y. Shih, “Machine assessment of neonatal facial expressions of acute pain”, *Decision Support System*, vol. 43, no. 4, pp. 1242–1254, 2007. https://doi.org/10.1016/j.dss.2006.02.004

[19] A. Beltramini, K. Milojevic, D. Pateron, “Pain Assessment in Newborns, Infants, and Children”, *Pediatric. Annals*, vol. 46, no. 10, pp. 387–395, 2017. https://doi.org/10.3928/19382359-20170921-03

[20] X. Cong, J. M. McGrath, R. M. Cusson, D. Zhang, “Pain Assessment and Measurement in Neonates: An Ipdated Review”, *Advances in Neonatal Care*, vol. 13, no. 6, pp. 379–395, 2013.

[21] C. L. von Baeyer L. J. Spagrud, “Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years”, *Pain*, vol. 127, no. 1–2, pp. 140–150, 2007. https://doi.org/10.1016/j.pain.2006.08.014

[22] J. Zieliński, M. Morawska-Kochman, T. Zatoński, “Pain assessment and management in children in the postoperative period: A review of the most commonly used postoperative pain assessment tools, new diagnostic methods and the latest guidelines for postoperative pain therapy in children”, *Advances in Clinical and Experimental Medicine*, vol. 29, no. 3, pp. 365–374, 2020.

[23] C. Greco, C. Berde, “Pain Management in Children”, Gregory’s Pediatric Anesthesia, Wiley, pp. 929–954, 2020. https://doi.org/10.1002/9781119371533.ch37

[24] G. Zamzmi, R. Kasturi, D. Goldgof, R. Zhi, T. Ashmeade, Y. Sun, “A Review of Automated Pain Assessment in Infants: Features, Classification Tasks, and Databases”, *IEEE Reviews in Biomedical. Engineering*, vol. 11, pp. 77–96, 2017.

[25] T. Voepel-Lewis, J. Zanotti, J. A. Dammeyer, S. Merkel, “Reliability and Validity of the Face, Legs, Activity, Cry, Consolability Behavioral Tool in Assessing Acute Pain in Critically Ill Patients”, *American Journal of Critical Care*, vol. 19, no. 1, pp. 55–61, 2010. https://doi.org/10.4037/ajcc2010624

[26] G. Guillen, “Digital Image Processing with Python and OpenCV”, *Sensor Projects with Raspberry Pi*, pp. 97–140, 2019. https://doi.org/10.1007/978-1-4842-5299-4_5

[27] R. Momtahina, M. Hossain, “Image Capturing and Automatic Face Recognition”, Dhaka, Bangladesh, 2019.

[28] O. Subea, G. Suciu, “Facial Analysis Method for Pain Detection”, International Conference on Future

Access Enablers of Ubiquitous and Intelligent Infrastructures, pp. 167–180, 2019. https://doi.org/10.1007/978-3-030-23976-3_17

[29] D. E. King, "Dlib-ml: A Machine Learning Toolkit", *The Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[30] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[31] O. M. Parkhi, A. Vedaldi, A. Zisserman, "Deep face recognition", Proceedings of the British Machine Vision Conference (BMVC), vol. 1, no. 3, p. 6, 2015. https://dx.doi.org/10.5244/C.29.41

[32] S. J. Pan, Q. Yang, "A Survey on Transfer Learning", *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.

[33] F. Zhuang, "A Comprehensive Survey on Transfer Learning", *Proceedings of the IEEE*, vol. 109, no. 1, pp. 1-34, 2019.

[34] H.-W. Ng, V. D. Nguyen, V. Vonikakis, S. Winkler, "Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning", Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15), pp. 443–449, 2015. https://doi.org/10.1145/2818346.2830593

[35] W. Ding, "Audio and face video emotion recognition in the wild using deep neural networks and small datasets", Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '1), pp. 506–513, 2016. https://doi.org/10.1145/2993148.2997637

[36] K. Zhang, L. Tan, Z. Li, Y. Qiao, "Gender and smile classification using deep convolutional neural networks", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[37] V. Campos, A. Salvador, B. Jou, X. Giró-i-Nieto, B. Jou, "Diving Deep into Sentiment: Understanding Fine-tuned CNNs for Visual Sentiment Prediction", *Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia (ASM '15)*, pp. 57-62, 2015. https://doi.org/10.1145/2813524.2813530

[38] H. Ding, S. K. Zhou, R. Chellappa, "FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition", 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 118–126, 2017.

[39] F. Wang, "Regularizing face verification nets for pain intensity regression", in 2017 IEEE International Conference on Image Processing (ICIP), pp. 1087–1091, 2017.

[40] M. S. Hossain, G. Muhammad, "Emotion recognition using deep learning approach from audio–visual emotional big data", *Information Fusion*, vol. 49, pp. 69–78, 2019. https://doi.org/10.1016/j.inffus.2018.09.008