Mobile Convolutional Neural Network for Neonatal Pain Assessment

Lucas P. Carlini¹, Leonardo A. Ferreira¹, Gabriel A. S. Coutrin¹, Victor V. Varoto¹, Tatiany M. Heiderich¹, Rita C. X. Balda², Marina C. M. Barros², Ruth Guinsburg², Carlos E. Thomaz¹ ¹University Center of FEI, São Bernardo do Campo SP, Brazil ²Federal University of São Paulo, São Paulo SP, Brazil

lucaspcarlini10,ruth.guinsburg@gmail.com, cet@fei.edu.br

Abstract

More than 500 painful interventions are carried out during the hospitalization of a newborn baby in an Intensive Care Unit. Since a direct and objective verbal communication by neonates is unlikely, this work proposes and implements a computational framework to automatically classify the neonatal pain based on its facial expression. Our findings showed promising results to correctly identify the facial expression of pain in neonates with high accuracy and generalization capability, highlighting as well sound facial regions that agree with pain scales used by neonatologists and with the visual perception of adults when assessing pain in neonates, whether they are health professionals or not.

1. Introduction

More than 500 painful interventions are carried out during the hospitalization of a newborn baby in a neonatal intensive care unit [10, 13]. In the past, due to inability of neonates verbally communicate pain, it was believed that the central nervous system of newborn babies was not fully developed, consequently, not being able to sense and suffer from pain [14]. However, it was observed in the latter years of 1980 that the central nervous and nociceptive systems are sufficiently developed in the sixth month of gestation [9, 11], leading to an increased sensitivity of pain, since inhibitory pathways of painful stimulus are not fully developed yet [2, 3].

During clinical practice, facial expression analysis has been widely used to verify the presence of pain. Although this analysis may be subjective, it is a noninvasive method that delivers valid information regarding the nature and intensity of the pain allowing better communication between the neonate and his/her caregiver [11, 12]. Therefore, methods based on clinical, statistical, deep learning and facial expression recognition [14, 19, 20] allow the implementation of computational frameworks that are specific to pain and enable continuous monitoring of the neonate.

This work unprecedentedly proposes and implements a mobile application for smartphones that uses Articial Intelligence (AI) techniques to automatically identify the facial expression of pain in neonates, presenting feasibility in real clinical situations. Firstly, a Convolutional Neural Network architecture was adapted and trained with face images captured before and after painful clinical procedure carried out routinely. Then, this computational model was optimized to a mobile environment to make it practical for everyday use. Moreover, we used an explainable AI method to identify facial regions that might be relevant to pain assessment.

2. Materials and Methods

This section is divided into three parts: (1) Face image datasets, (2) Computational methods, and (3) Implementation of our mobile application.

2.1. Face Image Datasets

We used two image datasets to design our proposed framework: iCOPE [4, 5, 6] and UNIFESP [14].

The UNIFESP dataset was developed by Heiderich et al. [14] at the Federal University of São Paulo. It is a proprietary dataset that includes 30 healthy neonates (7 late preterms and 23 born at term) of mixed ethnicity with 34 to 41 weeks of gestational age, and 24 to 168 hours of life. All photographs were captured using three Foscam cameras with resolution of 320x233. In total, 360 images were collected, of which: 138 were captured before a painful procedure, 30 during the procedure, and 192 images captured within 10 minutes after a painful procedure, such as blood specimen collection, Hepatitis B vaccine and Inborn Errors of Metabolism (IEMs). Subsequently, each image was randomly submitted for evaluation by health professionals. The assessment led to 164 images classified as "pain" and 196 images classified as "no pain".

The infant Classification of Pain Expression (iCOPE) dataset (available upon request) was developed by Brahnam et al. [4, 5, 6] during a study at the St. John Hospital with the Neonatology Department in Missouri, USA. A total of 200 images were captured from 26 neonates, 13 girls and 13 boys, all Caucasians. The age group of these neonates ranges from 18 hours to 3 days of life. All images were photographed using a Nikon D100 digital camera in ambient light conditions with resolution of 3008x2000. The neonates were photographed during a session in which they experienced 4 different stimuli performed in the following sequence: (1) Transport from one crib to another, (2) air stimulus on the nose, (3) cotton wool friction on the heel and (4) puncture on the heel for blood collection. During this sequence the newborn responses were labeled as either resting, crying, air stimulation, friction or painful procedure, resulting in: 63 images of neonates resting, 18 crying, 23 images of air stimulation, 36 during friction and 60 with neonates during painful procedures. Only resting and painful samples were used as "no pain" and "pain" for our goal, respectively.

2.2. Computational Methods

Firstly, we applied the state-of-the-art Retina Face [8] algorithm, commonly used in similar recognition tasks, in all images of the UNIFESP and iCOPE datasets in order to extract the face from each image. It is a single-stage pixel-wise face localization method which employs a multi-task learning strategy to simultaneously predict face score, face box, five facial landmarks, and 3D position and correspondence of each facial pixel. All faces from both datasets were successfully detected. Then, both datasets were merged adding up to 483 images. The train-test-split adopted was 80/20, next we applied Data Augmentation for each training sample of our dataset, generating a total of 8000 augmented images. Using TensorFlow [1], we performed image manipulations such as rotation, rescaling, horizontal and vertical offsetting and brightness and zoom variations.

Finally, following Zamzmi et al. [20], we applied the VGG-Face architecture with 16 layers, originally proposed and implemented by Parkhi [17], as our classification model. Since we are dealing with a small dataset its common to use the transfer-learning method, where we take advantage of an already pre-trained VGG-Face model, adding a fully-connected classifier on top, specifically trained with neonatal face images captured before and after painful clinical procedure to classify the facial expression into two classes "pain" or "no pain", without pain level detection. We used TensorFlow [1] for training and testing our proposed CNN. We performed a random search with parameters ranging from 50 to 2048 neurons and 1 to 3 fully connected layers. Using both datasets, the result suggested 2 fully connected layers with 512 neurons each. We used the Categorical Cross-Entropy as the loss function with L1 regularization penalty and the RMSprop [15] as our gradient descent optimization algorithm. Its noteworthy that all images must match VGG-Face input size of 224 x 224 x 3. Also, Parkhi [17] changed RGB channels to BGR and centralized values of each channel on zero.

Training process was carried out to obtain three classification models: (1) with the UNIFESP dataset only, (2) with the iCOPE dataset only, and (3) with the UNIFESP + iCOPE dataset. However, all models were tested on both datasets. Each model was evaluated with the k-Fold Cross Validation technique (k = 10), randomly choosing independent test sets (20% of original datasets) for each fold. The best model was optimized using the TensorFlow Lite Dynamic Range Quantization function. The embedded model was quantized from floating point weights (*float32*) to integers (*int8*) and during the inference process, weights are converted back from integers to floating point and cached in memory to reduce latency. The memory space required for the model decreased from 160Mb to 26Mb

2.3. Mobile Application

We developed a mobile application capable of detecting a neonate face and classifying it with low latency and offline. Also, our app registers the performed analysis and make them available for query, allowing metadata analysis and the collection of new face images. The application was developed for Android OS using the Android Studio IDE.

On the app's home screen (Figure 1a) there are three buttons that activate different functionalities of the application: instructions, camera and history. In Instructions, the user will find a quick guide for the other two functionalities (camera and history) of the app. The selection of the Camera button activates the app's main feature. The realtime analysis starts right after the neonate's name insertion. The camera automatically captures a picture, which is then processed by the face detector algorithm. In the mobile environment the face detection is carried out by the Face Detection API from the Firebase's ML Kit by Google¹. When a face is successfully detected, the picture's region corresponding to the detected face is cropped out and inputted in the classification model, which will determinate if the facial expression indicates the presence of pain or not. The user will see the image's classification, the result's confidence score and the processing time (Figure 1b). At last, by selecting the History button at the home screen, the application will present a list containing the performed analyses and a search tool, which can be used to find a specific neonate by typing it's name or unique ID.

¹More information on https://developers.google.com/ ml-kit/vision/face-detection



(a) Home Screen.

(b) Pain assessment.

Figure 1: Mobile application screens.

	UNIFESP	iCOPE	Both
Accuracy	$72\%\pm3\%$	$83\%\pm2\%$	$89\%\pm4\%$
F1 Score	0.72 ± 0.04	0.86 ± 0.02	0.89 ± 0.03
AUC	0.74 ± 0.03	0.82 ± 0.03	0.86 ± 0.05

Table 1: Evaluation metrics results for each model.

3. Results and Discussion

In this section we present the evaluation of our classification model and, then, the results of model interpretability.

3.1. Evaluation Metrics

Table 1 shows the results obtained for each classification model. Comparing the UNIFESP model with the iCOPE, we can observe that the second one outperforms the first in all evaluation metrics. We believe that this better performance is due to the higher resolution of the iCOPE dataset. Our framework achieved 89% Accuracy, 0.89 F1 Score and 0.86 AUC when trained in both datasets. We believe that these results are suited for neonatal pain assessment since commonly used clinical pain scales achieved similar results, such as the NFCS [12] that achieved reliability of 0.86.

3.2. Model Interpretability

To understand the relationship between our classification model prediction in terms of its features, we applied an interpretability method, named Integrated Gradients [18].

Observing Figure 2, we can see that the most highlighted facial features are the forehead, upper contour of the nose and nasolabial groove. Specifically to images classified as "pain", the mouth with tongue protrusion was highlighted.



Figure 2: Examples of interpretability on iCOPE dataset and its corresponding classification probabilities.

Analyzing all results, we believe that these features may be the most discriminating facial regions to pain assessment. Moreover, its interesting that these features are deemed clinically relevant and agrees with the visual perception of adults when assessing pain, whether they are health professionals or not [7, 12].

4. Conclusion

This work presents a computational framework implemented on mobile environment that uses AI to automatically identify the facial expression of pain in neonates, presenting feasibility in real clinical situations and practical for everyday use. Our findings showed promising results to correctly identify the facial expression of pain in neonates with high accuracy and generalization capability.

Moreover, to the best of our knowledge, this is the first work to apply explainable AI techniques in neonatal pain classification. Our methodology presented novel results highlighting as well sound facial regions that agree with pain scales used by neonatologists and with the visual perception of adults when assessing pain in neonates, whether they are health professionals or not.

As future work, we intend to carry out practical tests of our mobile application in the Neonatal Intensive Care Unit at the Hospital of São Paulo. We believe that these practical tests are needed to identify limitations of the proposed solution when dealing with difficulties of real situations, such as the presence of artifacts on the face of the newborn and physiological signal measuring instruments. Also, we intend to implement our framework using other CNN architectures, such as AlexNet [16] and N-CNN [21].

Acknowledgements

The authors would like to thank the financial support provided by the Brazilian funding agencies FAPESP (2018/13076-9), CNPq (401059/2019-7) and CAPES.

References

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensor-Flow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 2
- [2] Kanwaljeet JS Anand and David B Carr. The neuroanatomy, neurophysiology, and neurochemistry of pain, stress, and analgesia in newborns and children. *Pediatric Clinics of North America*, 36(4):795–822, 1989. 1
- [3] Kanwaljeet JS Anand, Paul R Hickey, et al. Pain and its effects in the human neonate and fetus. N Engl j Med, 317(21):1321–1329, 1987. 1
- [4] Sheryl Brahnam, Chao-Fa Chuang, Randall S Sexton, and Frank Y Shih. Machine assessment of neonatal facial expressions of acute pain. *Decision Support Systems*, 43(4):1242– 1254, 2007. 1, 2
- [5] Sheryl Brahnam, Chao-Fa Chuang, Frank Y Shih, and Melinda R Slack. Svm classification of neonatal facial images of pain. *International Workshop on Fuzzy Logic and Applications*, pages 121–128, 2005. 1, 2
- [6] Sheryl Brahnam, Chao-Fa Chuang, Frank Y Shih, and Melinda R Slack. Machine recognition and representation of neonatal facial displays of acute pain. *Artificial intelligence in medicine*, 36(3):211–222, 2006. 1, 2
- [7] Lucas Pereira Carlini, Juliana C. A. Soares, Giselle V. T. Silva, Tatiany M. Heideirich, Rita C. X. Balda, Marina C. M. Barros, Ruth Guinsburg, and Carlos Eduardo Thomaz. A visual perception framework to analyse neonatal pain in face images. In Aurélio Campilho, Fakhri Karray, and Zhou Wang, editors, *Image Analysis and Recognition*, volume 12131 of *Lecture Notes in Computer Science*, pages 233–243, Cham, 2020. Springer International Publishing. 3
- [8] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multilevel face localisation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5203–5212, 2020. 2
- [9] Brenda Golianu, Elliot J Krane, Karen S Galloway, and Myron Yaster. Pediatric acute pain management. *Pediatric Clinics of North America*, 47(3):559–587, 2000. 1
- [10] Ruth Eckstein Grunau. Neonatal pain in very preterm infants: long-term effects on brain, neurodevelopment and pain reactivity. *Rambam Maimonides medical journal*, 4(4), 2013. 1
- [11] Ruth VE Grunau and Kenneth D Craig. Pain expression in neonates: facial action and cry. *Pain*, 28(3):395–410, 1987.
 1

- [12] Ruth VE Grunau, C Celeste Johnston, and Kenneth D Craig. Neonatal facial and cry responses to invasive and noninvasive procedures. *Pain*, 42(3):295–305, 1990. 1, 3
- [13] Ruth Guinsburg. Avaliação e tratamento da dor no recémnascido. J Pediatr (Rio J), 75(3):149–60, 1999. 1
- [14] Tatiany Marcondes Heiderich, Ana Teresa Figueiredo Stochero Leslie, and Ruth Guinsburg. Neonatal procedural pain can be assessed by computer software that has good sensitivity and specificity to detect facial movements. Acta Paediatrica, 104(2):e63–e69, 2015. 1
- [15] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. Overview of mini-batch gradient descent, 2012. [Online; accessed 23-Julho-2020]. 2
- [16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25:1097–1105, 2012. 3
- [17] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *British Machine Vision Confer*ence, 2015. 2
- [18] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. arXiv preprint arXiv:1703.01365, 2017. 3
- [19] G. F. Teruel, T. M. Heiderich, R. Guinsburg, and C. E. Thomaz. Analysis and recognition of pain in 2d face images of full term and healthy newborns. *Proceedings of the XV Encontro Nacional de Inteligencia Artificial, ENIAC 2018*, pages 228–239, 2018. 1
- [20] Ghada Zamzmi, Dmitry Goldgof, Rangachar Kasturi, and Yu Sun. Neonatal pain expression recognition using transfer learning. arXiv preprint arXiv:1807.01631, 2018. 1, 2
- [21] Ghada Zamzmi, Rahul Paul, Dmitry Goldgof, Rangachar Kasturi, and Yu Sun. Pain assessment from facial expression: Neonatal convolutional neural network (n-cnn). In 2019 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2019. 3