Neonatal Pain Detection from Facial Expressions Using Deep Learning

Fei Zhao
The University of Alabama at Birmingham
Chengcui Zhang
The University of Alabama at Birmingham
Katherine M. Dudding
kmdudding@ua.edu
The University of Alabama at Birmingham
Allyson N. Sanders
The University of Alabama at Birmingham
Peyton Lewis-Chumley
The University of Alabama at Birmingham
Lenington Kathryn
The University of Alabama at Birmingham

Research Article

Keywords: Deep Learning, Neonatal, Pain, Facial Expression, Neonatal Pain Detection

Posted Date: February 29th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3979706/v1

Additional Declarations: No competing interests reported.

Abstract

Current clinical tools to assess neonatal pain, including various pain scales such as Neonatal Infant Pain Scale (NIPS) and Neonatal Pain, Agitation, and Sedation Scale (N-PASS), are overly reliant on nurses' subjective observation and analysis. Emerging deep learning approaches seek to fully automate this, but face chal- lenges including massive training data and computational resources, and potential public mistrust. Our study prioritizes facial information for pain detection, as facial muscles exhibit distinct patterns during pain events. This approach, using a single camera, avoids challenges associated with multimodal methods, such as data synchronization, larger training datasets, deployment issues, and high computational costs. We propose a deep learning-based neonatal pain detection framework that can alert a neonate pain management team when a pain event occurs, consisting of two main components: a transfer learning-based end-to-end pain detection neural network, and a manual assessment branch. The proposed neural network requires much less data to train and can evaluate whether a neonate is in a pain state based on facial information only. Additionally, the man- ual assessment branch can specifically handle the borderline/hard cases where the pain detection network is less confident. The integration of both machine detection and manual evaluation can increase the recall rate of true pain events, reduce the manual evaluation effort, and increase public trust in such applications. Experimental results show our neural network sur- passes state-of-the-art algorithms by at least 25% in accuracy on the MNPAD dataset, with overall framework accuracy reaching 82.35% with integration of manual assessment branch.

1 Introduction

Neonatal pain, recognized for over 36 years, has profound implications for long-term health outcomes [1]. Prolonged exposure to pain during early neurodevelopment can lead to neurodevelopmental impairments and physiological consequences in later life [2]. Alarmingly, neonates in Newborn Intensive Care Units (NICUs) undergo an average of 14 painful procedures daily [3], accumulating to nearly 185 painful events during an average NICU stay of 13.2 days [4]. With around 13 million newborns admitted to NICUs globally each year, the challenge of timely pain detection and management is paramount. Despite advancements in care standards, only 33% of neonates receive adequate pain management [5]. In most cases, a nurse is required to take care of up to four neonates depending on their acuity, which is difficult when these neonates are often critically ill and require intensive care. Since there is no mature algorithm alerting the nurse which neonate is suffering from a pain event, especially a mild pain event (borderline/hard case), the nurse may miss the window of opportunity to relieve the pain while managing other neonates. Therefore, promptly detecting and alerting neonatal pain events, especially mild pain events (borderline/hard cases), has become increasingly more important to neonatal pain management. The alert of pain events can be helpful for nurses to prioritize their tasks and manage pain promptly.

There are three main challenges to detecting and alerting neonatal pain events including the subtle communication of pain that varies for every neonate. Subjective bias is associated with the assessment of pain and limited technologies within the NICU. Neonates communicate their pain through subtle changes by physiological, behavioral, and stress hormone indicators. Physiologically, neonates have changes in heart rate, blood pressure, and oxygen saturation levels; behaviorally, neonates display changes in their facial features (e.g., eyebrow furrowing, grimace) and body movements (e.g., flying of arms and legs). In response to painful events, the neonate also releases several hormones (e.g., cortisol, catecholamines, growth hormone, glucagon) to compensate. Crying is a typical way neonates could communicate their pain. But unfortunately, some neonates are so immature that they cannot sustain crying to com- municate or alert the nurse that they are in pain. The communication of pain by the neonate is determined by the age of the neonate, where the more premature the neonate the more subtle the communication of pain.

Subjective bias presents challenges to detect pain in the neonate. Nurses receive extensive training for assessing pain in the neonate [6]. Nurses assess pain according to physiological measures and behavioral characteristics then use a pain scale to quantify the pain level. Despite this multimodal approach, pain assessment remains subjective. Therefore, nurses have the potential to introduce implicit bias into their assessments. What one nurse may perceive the same neonate with no pain where another nurse may perceive the same pain, as moderate pain level.

There are also a limited number of technologies in NICU to communicate neonatal pain. In general, NICU consists of various types of equipment monitoring the status of neonates, e.g., hemodynamic monitors detecting changes in physiological measures, electronic health records (EHRs) for recording nurse assessments and collecting data, and neonatal pain scale tools to guide and quantify pain assessments. However, none of them were designed to analyze or report pain data to inform nursing care [7, 8]. Current NICU technologies continue to struggle to effectively detect and communicate the neonate's pain.

Machine learning techniques have advanced rapidly, leading to an increased focus on automated neonatal pain detection. Brahnam et al. in [9] were pioneers in this area, emphasizing the importance of facial information for objective pain detection, a departure from traditional nurse-based assessments. However, early methods relied on hand-crafted features, which were sensitive to environmental changes and required extensive data preprocessing. By 2019, Zamzmi et al. in [10] introduced deep learn- ing to this domain. While powerful, these deep learning-based models demanded vast training data and computational resources. Given the scarcity of comprehen- sive neonatal pain datasets, such models are easily prone to overfitting, resulting in a poor performance in terms of generalizability. Additionally, fully automated deep learning-based models may deliver erroneous medical evaluations on mild pain events (**borderline/hard cases**), which could erode the public's trust in such models and healthcare institutions. Therefore, there remains a need for a new framework/workflow more robust to changes in environmental parameters, less dependent on the amount of training data, and more trustworthy to the public.

In this work, we aim to develop a deep learning-based neonatal pain detection framework to explore the full potential of facial information in such pain event detection and to alert nurses to manage neonatal pain events. This framework, consisting of a transfer learning-based end-to-end pain detection neural network and a branch for manual assessment, can address issues of **data scarcity** (using data collected from a single, portable, and low-cost sensor) and **public trust, while also reducing cost associated with**

computation and deployment. In this framework, the transfer learning-based end-to-end neural network can predict the probability that an event belongs to a certain pain class, i.e., “pain” or “non-pain”. Based on the neural network’s prediction, the pain management team will be alerted. If the neural net- work’s output probability for the predicted class is lower than a threshold, a nurse will be informed to manually assess the pain state of the subject in this case and decide whether the pain management team should be alerted. According to our experimental results, the proposed deep learning-based pain detection framework is more robust and reliable than the state-of-the-art (SOTA) methods.

2 Background

Unrelieved neonatal pain can lead to changes in the central nervous system and alter pain processing mechanisms [11, 12]. Prolonged pain during NICU stays can result in hyperalgesia (increased pain sensitivity) and allodynia (pain from non-painful stimuli), potentially evolving into chronic pain in adulthood. Timely neonatal pain management could reduce the risk of exposure to unrelieved pain. To enable the nurses, apply pain management, there has been increasingly more attention on neonatal pain detection and assessment. Based on the sensor type, the SOTA methods can be categorized into two classes:

1. Physical Contact Measurement-based methods. The key idea of this direction is to use the data provided by sensory-level input devices including one or more types of physical contact sensors. The framework proposed by Alzamzmi et al. in [13] combines the information from physical contact sensors, such as heart rate and heart rate variability, electromyography of the mouth, and facial grimacing, to predict whether a neonate is in pain or not. However, this method requires different types of data collected from multiple sensors, which can be hard to synchronize.
2. Remote Measurement-based methods. In this direction, most works choose to use the camera as the only input source of models because facial expressions can provide rich information related to a painful event. The work proposed by Brahnam et al. in [9] uses a single camera to collect the input data, and manually crops the face area of each image. The works proposed by Brahnam et al. in [14] and Celona et al. in [15] use the Dlib [16] face & landmark detection algorithms to detect and crop face area to form the input data for models, respectively. The works by Zamzmi et al. in [17] and Zhi et al. in [18] detect and crop face area by using the cascade of boosted Haar-like classifier and Active Appearance Model, respectively. The works proposed by Zamzmi et al. in [10, 19] take advantage of the Zface [20] to detect the facial area and crop the images to form the input data. Remote measurement-based methods are more practical and easier to implement.

Based on the model type, the SOTA methods can be categorized into two classes:

1) traditional hand-crafted feature-based models. In this direction, after obtaining the input data, various traditional hand-crafted feature extractors are applied to extract high-level features of input data. The work by Brahnam et al. in [9] applies Principal Component Analysis (PCA, [21]) and Linear Discriminant Analysis (LDA, [21]) to extract features from manually cropped input face images. The work by Brahnam

et al. in [22] uses Discrete Cosine Transform (DCT) feature extractors to generate high- level features for the model. The work by Brahnam et al. [14] applies an ensemble of feature descriptors to extract informative features from the input face image. The works in [23–25] proposes distance-based metrics to detect pain events. However, the issue in this direction is that the hand-crafted feature descriptors are sensitive to lighting/background/camera view variations. Therefore, the robustness of such models is relatively low; 2) deep learning feature-based methods. In this direction, deep neural networks are adopted for feature extraction. The work by Zamzmi et al. in [10] proposes a convolutional neural network (CNN) to extract high-level features from the face image and designs fully connected layers for decision-making. The work by Zamzmi et al. in [26] combines the hand-crafted features with deep learning features to train a Naive Bayes Classifier for the pain expression recognition task. Meanwhile, the work by Salekin et al. in [27] uses Visual Geometry Group (VGG16, [28]) to extract the face and body features for the model. Moreover, further work by Salekin et al. in [29] proposes a multimodal approach to tackle the neonatal postoperative pain assessment task, requiring more data for training. The key idea of these works is that the deep learning-based model can automatically learn hierarchical features (low to high levels) from data. With sufficient training data, the learned feature representation can be more robust than hand-crafted ones. However, existing deep learning works usually require a large amount of training data, adding another hurdle especially to medical studies like the one proposed. In this work, we focus on addressing the issues of data scarcity and public trust using a remote-measurement based approach with neonate face images only, and test its limit in neonate pain detection. Multimodality-based methods are beyond the scope of this paper as they require much more training data and are not suitable for limited data availability.

Transfer learning, a subset of machine learning, repurposes models trained for one task to initiate models for related tasks [30]. For example, a model trained to recognize animals might be adapted to detect cats, depicted in Fig. 1. Transfer learning has a better starting point, a higher slope, and higher asymptote [21]. Moreover, learning from scratch usually requires massive amounts of training data, while the dataset required by transfer learning is several orders of magnitude smaller.

3 Methods

To address the problem of timely neonatal pain detection and the concerns of data scarcity and public trust, we propose a deep learning-based pain detection frame- work that allows a trained neural network to collaborate with manual assessment. The framework consists of two main components: a transfer learning-based end-to-end pain detection neural network that can detect neonatal pain events by using neonatal facial information and alert the management team, and a manual assessment branch that allows a nurse to evaluate the pain state of a case if the neural network's output probability for the predicted class is lower than a threshold (usually, these are border- line/hard cases). The transfer learning technique enables our model to be well-trained on a relatively small dataset and to obtain a reasonably good generalizability. Addi- tionally, the inclusion of manual assessment on borderline/hard cases can increase public trust on this deep learning-based medical application.

Ethical Approval. Since the investigation involves human subjects, it adheres to the U.S. Food and Drug Administration (FDA) regulations. The research has been approved by the University of Alabama at Birmingham Institutional Review Board for human use.

Datasets and Details. Two open-access neonatal pain datasets are available: the Classification of Pain Expressions (COPE) from Missouri State University (MSU) by Brahnam et al. in [31] and the Multimodal Neonatal Procedural and Postoperative Pain dataset (MNPAD) from the University of South Florida (USF) by Salekin et al in [29]. The COPE dataset comprises 204 RGB images of 26 healthy Caucasian neonates, aged 18 hours to 3 days, captured during various stimulations. However, it is limited in data, lacks images of neonates with feeding tubes or oxygen delivery devices, and is imbalanced with a 7:3 non-pain to pain ratio. On the other hand, the MNPAD dataset, collected from 58 neonates (27-41 gestational age) in NICUs, offers videos of both facial expressions and body movements from 6 publicly available neonates, capturing two types of pain events: procedural pain and postoperative pain.

Ground Truth. To ensure that the ground truth of the MNPAD dataset is accurate, three neonatal nurses, who are experienced in neonatal pain assessment, developed an annotation protocol to annotate “pain” and “non-pain” events. The protocol was based on current neonatal pain scales consisting of various behavioral characteristics and stress cues. The nurses observed the videos and annotated the data independently then quantified the pain and non-pain events using the Neonatal Infant Pain Scale (NIPS, [32]). Once video annotation was completed, the nurses met to dis- cuss significant discrepancies and came to a consensus. The NIPS pain scores were averaged from the three nurse annotators. The pain score ranges from 0 to 7. As this paper focuses on detecting pain events using static images, we extracted keyframes from the MNPAD videos based on inter-frame differences in CIELUV color space [33]. A keyframe is identified if there is a local maximum of inter-frame difference, as described by Amanpreet Walia [34]. The keyframes with pain scores lower than 3 were labeled as “non-pain”, otherwise labeled as “pain”. The keyframes form the MNPAD image dataset, containing 71 “non-pain” and 116 “pain” images. The ratio of “pain” class to “non-pain” class in MNPAD image dataset is 4:6.

Model and Framework. In this work, we propose a deep learning-based pain detection framework that uses neonatal facial expression information to alert the pain management team. The framework consists of two components: a transfer learning- based deep neural network and a manual assessment branch. If the network classifies a case into the “pain” class with a sufficiently high probability (confidence score), it will alert the pain management team. If the neural network’s output probability for the predicted class is lower than a predefined threshold, the framework will notify the nurse to manually assess the subject in the case. Therefore, only when the assessment result is “pain”, the pain management team will be alerted. The entire process is shown in the Inference Process of Fig. 2.

Transfer learning-based Pain Detection Neural Network : To leverage facial infor- mation for neonatal pain detection, we utilize the YOLO model [44], pre-trained on the WIDER FACE dataset [45]. This model identifies and extracts facial regions, which are then resized to 224x224 pixels while preserving the original aspect ratio through adaptive padding. Given the constraints of limited data, we turn to transfer learning, adapting SOTAmodels to our specific needs. We select two pre-trained models: VGG16

[28] and ResNet50 [46], as the backbones of our pain detection neural networks. As shown in Fig. 2, once we obtain a pre-trained model, the classification layers used in the pre-training task, e.g., the facial expression recognition task (FER) or the face recognition task (e.g., VGGFace in Fig. 2), are removed. In our pain detection neural network, a pain classifier consisting of three fully connected layers is appended to the pre-trained model/backbone, e.g., VGG16 or ResNet50. Meanwhile, we use dropout and batch normalization to mitigate the potential overfitting issue and to speed up the convergence, respectively. The architecture of the pain classifier is shown in Fig. 3. The "Default" in Fig. 3 means the default setting of Keras [47] that is a high-level deep learning Python library [48] providing an interface of building and training neural networks.

The manual assessment branch: to add more assurance to the pain dection work- flow, we add a manual assessment branch in the proposed framework, which is based on the probability (confidence) score of the Softmax [49] output of our pain detection neural network. If the output probability score for the predicted class is lower than a predefined threshold, a nurse will be required to manually assess the pain state of the subject. A low probability usually indicates that the model has low confidence on the predicted class (borderline/hard case). Setting a higher threshold will result in a higher overall accuracy but more cases needing manual assessment, while a lower threshold will result in a lower accuracy but fewer cases will need manual assessment. A optimal threshold value can be determined by the preference of the nurses or through trial in field studies.

4 Experiments and Results

Two experiments were conducted in this study, including a pre-trained model selection experiment and a comparative study experiment.

In the pre-trained model selection experiment, VGG16 and ResNet50 were chosen as the network backbone candidates. Those network backbones can be pre-trained on various tasks and datasets, e.g., building detection task in [50]. In this study, we chose the facial expression recognition (FER) task and the face recognition (FR) task as the pre-training tasks, because FER and FR tasks are relevant to the pain detection task that uses facial information. In this experiment, the VGG16 and ResNet50 are pre- trained on two datasets: VGGFace2 [51] dataset that is the SOTA face recognition dataset, and FER2013 [52] facial expression recognition dataset provided by Kaggle. The backbones pre-trained on these tasks and large datasets are expected to capture facial features with efficacy. As shown in Fig. 2, during the training process, the pre- trained backbone is connected to a new set of three fully connected layers that perform the classification at the end of the pipeline for pain detection. The performance of each model on the pain detection task is evaluated by using the leave-one-subject-out cross-validation (LOSOCV) method on the COPE dataset containing 26 subjects. The entire COPE dataset is divided into 26 subsets (one for each subject.) Then, a model is trained on 25 subsets while the validation (early stopping) and testing (accuracy estimation) are performed on the remaining subset [53]. This operation is repeated 26 times for each model, and each time we use one subset as testing data and the remaining 25 subsets as training data and re-train the model (not cumulative training so each test subset is unseen during training the corresponding model), resulting in 26 accuracy scores. The average of the 26 accuracy scores is used as the evaluation metric. Additionally, since the COPE dataset contains very few images, data augmentation techniques are used during the training process to mitigate the potential overfitting issue. We perform the following perturbations on the training data, including image rotation (0°-15°) and horizontal flip while keeping the labels the same as the original. Some augmented image samples are shown in Fig. 4. The models are trained on a single NVIDIA Tesla P100 16GB GPU with a learning rate of 0.00001.

As shown in Table 1, we can see that the neural network with ResNet50 backbone pre-trained on a facial expression recognition (FER) task, achieved the highest average accuracy score of 96.39% on the COPE dataset. According to the results in Table 1, a neural network, which was pre-trained on the facial expression recognition task, outperforms its counterpart pre-trained on the face recognition task (VGGFace), in addressing the neonatal pain detection problem. It may be because that the "pain" or "non-pain" facial activities can be considered as facial expressions to some extent. Therefore, we choose the FER pre-trained ResNet50 as the backbone of our pain detection neural network for further experiments.

Table 1 The results of the pre-trained model selection experiment

Model	LOSOCV Accuracy on the COPE Dataset
VGG16 (VGGFace)	90.52%
ResNet50 (VGGFace)	93.39%
VGG16 (FER)	93.30%
ResNet50 (FER)	96.39%

In the comparative study experiment, we compared our transfer learning-based pain detection neural network with the SOTA models including WGAN [15] and N- CNN [10], on the MNPAD image dataset. All the models were trained on the COPE dataset and tested on the unseen MNPAD image dataset (187 images). In this exper- iment, the FER pre-trained ResNet50 is used as the backbone of our pain detection neural network. Our neural network was trained on the COPE dataset cumulatively. Since there is no publicly accessible code of WGAN and N-CNN, we rebuilt the WGAN and N-CNN models and trained them on the COPE dataset following the instructions in the works ([10, 15]). Accuracy is used to measure the performance of each model, because the class distribution in the test dataset is near-balanced. As shown in Table 2, the accuracy of the proposed pain detection neural network is at least 25% higher than that of the SOTA algorithms. Furthermore, we tested the entire deep learning- based pain detection framework on the MNPAD image dataset, consisting of the pain detection network and a manual assessment branch. Using a threshold of 0.55 in the manual assessment branch, the proposed pain detection framework achieved a accu- racy of 82.35%, in which only 10 of the 187 images require manual evaluation by the nurse. As the threshold increases, the accuracy of the framework improves. However, the number of images requiring manual evaluation also increases. For example, at a threshold of 0.6, the framework achieved an even higher accuracy of 85.03%. However, 29 out of the 187 cases required manual assessment. The threshold can be optimized by nurses based on their preference for the balance between the accuracy and manual assessment workload.

Table 2 The results of comparative study

Model	**Test Accuracy on the Unseen MNPAD Image Dataset**
WGAN ([15])	61.50%
N-CNN ([10])	62.03%
ResNet50 (FER)	77.54%
ResNet50 (FER) + Manual Assessment	**82.35% (10 out of 187 are manual assessed)**

The results reported in Table 2 clearly show that our proposed neural network outperforms the SOTA methods on the unseen MNPAD image dataset, achieving the highest accuracy score of 77.54% (without using manual assessment.) The accuracy scores of N-CNN and WGAN models are 62.03% and 61.50%, respectively. The dif- ferent performances of the models in this experiment may be caused by the design of the models. N-CNN model uses random weights for model initialization and is trained from scratch. This design can lead to a severe overfitting issue on a relatively small training dataset, e.g., the COPE dataset containing only 204 images. N-CNN model classifies all the unseen images as “non-pain”, which indicates that it overfits to the COPE dataset. The performance of WGAN model is even worse than that of N-CNN, which indicates WGAN’s limited generalizability. As mentioned in the Model and Framework section, our pain detection neural network is based on transfer learning, which can reduce the dependence on the amount of training data. Additionally, the dropout layers in our classifier can mitigate the potential overfitting issue and improve generalizability.

When comparing the MNPAD dataset with the COPE dataset, it becomes evi- dent that the COPE dataset lacks samples that present challenging scenarios, such as faces obscured by nasal tubes, oxygen tubes, or significant occlusions. These obstruc- tions can make facial recognition and pain detection particularly challenging. Despite these disparities in data distribution and the fact that our transfer learning-based deep neural network has not been exposed to such challenging samples during training, our model still demonstrated remarkable proficiency. It correctly classified a signifi- cant portion of these samples, achieving an overall accuracy of 77.54% on the unseen MNPAD dataset. This performance underscores our model’s superiority over SOTA methods. We believe that, with the increasing availability of neonatal training datasets, the cases with tubes or occlusions will become less of a problem for the proposed network.

5 Conclusion and Future Work

In this work, we propose a novel deep learning-based pain detection framework for detecting neonatal pain events using neonatal facial information. The proposed framework consists of two main components, including a transfer learning-based pain detection neural network and a manual assessment branch. Compared to the SOTA methods, the performance of the proposed pain detection neural network is at least 25% higher. The transfer learning technique and dropout training strategy enable our neural network to be well-trained on a relatively small dataset, thereby addressing the data scarcity issue in the neonatal pain detection task. Additionally, the man- ual assessment of borderline/hard cases mitigates the public trust concerns about medical applications. The future work can be categorized into three directions: (1) further developing a training data-efficient multimodality classifier (for datasets with multi-sensor information) by leveraging existing pre-trained models for each modality involved, (2) collecting a much larger multimodality dataset for neonatal pain detec- tion and assessment, and (3) developing more portable and low-computation models that can be deployed on the user-end edge devices.

Declarations

Funding Declaration:

this work was generously funded by the University of Alabama at Birmingham, School of Nursing, Dean's Scholar Award.

Conflict of Interest:

all authors certify that they have no affiliations with or involve- ment in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Funding:

the authors did not receive support from any organization for the submitted work.

Author Contribution

FZ, CZ, and KD made substantial contributions to the conception and design of the work. FZ and CZ made contributions to acquisition, analysis and the creation of the new software the work. KD, AS, PLC, and KL made significant contributions to the interpretation of the data for coding. FZ, CZ, and KD drafted the work and revised the work. FZ, CZ, KD, AS, PLC, and KL approved the version to be published. FZ, CZ, and KD agree to be accountable for all aspects of the work ensuring that questions related to the accuracy of any part of the work are appropriately investigated and resolved.

References

1. Anand KJ, Hickey PR (1987) Pain and its effects in the human neonate and fetus. N Engl J Med 317(21):1321–1329
2. Chawanpaiboon S, Vogel JP, Moller A-B, Lumbiganon P, Petzold M, Hogan D, Landoulsi S, Jampathong N, Kongwattanakul K, Laopaiboon M et al (2019) Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. Lancet global health 7(1):37–46
3. Cruz MD, Fernandes A, Oliveira C (2016) Epidemiology of painful procedures performed in neonates: a systematic review of observational studies. Eur J Pain 20(4):489–498
4. Montgomery S (2023) : The Average Time Babies Spend In The NICU & Why It Mat- ters. https://www.romper.com/p/the-average-time-babies-spend-in-the-nicu-w hy-it-matters-18202975. Accessed: Sep. 4
5. Harrison WN, Wasserman JR, Goodman DC (2018) Regional variation in neonatal intensive care admissions and the relationship to bed supply. J Pediatr 192:73–79
6. Dudding KM (2018) : Recognition of pain in the neonate to increase effective neonate to nurse communication. PhD thesis, The University of Arizona
7. Carrington JM, Effken JA (2011) Strengths and limitations of the electronic health record for documenting clinical events. CIN: Computers Inf Nurs 29(6):360–367
8. Dudding KM, Gephart SM, Carrington JM (2018) Neonatal nurses experience unintended consequences and risks to patient safety with electronic health records. CIN: Computers Inf Nurs 36(4):167–176
9. Brahnam S, Chuang C-F, Shih FY, Slack MR (2006) : Svm classification of neona- tal facial images of pain. In: Fuzzy Logic and Applications: 6th International Workshop, WILF 2005, Crema, Italy, September 15–17, 2005, Revised Selected Papers 6, pp. 121–128 Springer
10. Zamzmi G, Paul R, Goldgof D, Kasturi R, Sun Y (2019) : Pain assessment from facial expression: Neonatal convolutional neural network (n-cnn). In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–7 IEEE
11. Williams MD, Lascelles BDX (2020) Early neonatal pain—a review of clinical and experimental implications on painful conditions later in life. Front Pead 8:30
12. McPherson C, Miller SP, El-Dib M, Massaro AN, Inder TE (2020) The influ- ence of pain, agitation, and their management on the immature brain. Pediatr Res 88(2):168–175
13. Alzamzmi GA, Goldgof D, Sun Y, Kasturi R, Ashmeade T (2020) Machine-based infants pain assessment tool. Google Pat US Patent 10:827,973

14. Brahnam S, Nanni L, McMurtrey S, Lumini A, Brattin R, Slack M, Bar- rier T (2023) Neonatal pain detection in videos using the icopevid dataset and an ensemble of descriptors extracted from gaussian of local descriptors. Appl Comput Inf 19(1/2):122–143
15. Celona L, Brahnam S, Bianco S (2019) : Getting the most of few data for neonatal pain assessment. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 298–301
16. King DE (2009) Dlib-ml: A machine learning toolkit. J Mach Learn Res 10:1755–1758
17. Zamzmi G, Pai C-Y, Goldgof D, Kasturi R, Ashmeade T, Sun Y (2016) : An approach for automated multimodal analysis of infants' pain. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 4148–4153 IEEE
18. Zhi R, Goldgof GZD, Ashmeade T, Sun Y (2018) Automatic infants' pain assess- ment by dynamic facial representation: effects of profile view, gestational age, gender, and race. J Clin Med 7(7):173
19. Zamzmi G, Pai C-Y, Goldgof D, Kasturi R, Ashmeade T, Sun Y (2019) A 12 comprehensive and context-sensitive neonatal pain assessment using computer vision. IEEE Trans Affect Comput 13(1):28–45
20. Jeni LA, Cohn JF, Kanade T (2015) : Dense 3d face alignment from 2d videos in real- time. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 1, pp. 1–8 IEEE
21. Olivas ES, Guerrero JDM, Martinez-Sober M, Magdalena-Benedito JR, Serrano L et al Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques. IGI global
22. Brahnam S, Nanni L, Sexton RS (2008) : Neonatal facial pain detection using nnsoa and lsvm. In: IPCV, pp. 352–357
23. Schiavenato M, Baeyer CL (2012) : A quantitative examination of extreme facial pain expression in neonates: The primal face of pain across time. Pain Res Treat 2012
24. Schiavenato M, Butler-O'Hara M, Scovanner P et al (2011) Exploring the associa- tion between pain intensity and facial display in term newborns. Pain Res Manage 16:10–12
25. Zamzami G, Ruiz G, Goldgof D, Kasturi R, Sun Y, Ashmeade T (2015) : Pain assessment in infants: Towards spotting pain expression based on infants' facial strain. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 5, pp. 1–5 IEEE
26. Zamzmi G, Goldgof D, Kasturi R, Sun Y (2018) : Neonatal pain expression recognition using transfer learning. arXiv preprint arXiv :180701631
27. Salekin MS, Zamzmi G, Goldgof D, Kasturi R, Ho T, Sun Y (2019) : Multi- channel neural network for assessing neonatal pain from videos. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 1551– 1556 IEEE
28. Simonyan K, Zisserman A (2014) : Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv :14091556

29. Salekin MS, Zamzmi G, Hausmann J, Goldgof D, Kasturi R, Kneusel M, Ashmeade T, Ho T, Sun Y (2021) Multimodal neonatal procedural and postoperative pain assessment dataset. Data Brief 35:106796
30. Thrun S, Pratt L Learning to Learn. Springer
31. Brahnam S, Nanni L, Sexton R Introduction to neonatal facial pain detection using common and advanced face classification techniques 13
32. Desai A, Aucott S, Frank K, Silbert-Flagg J (2018) Comparing n-pass and nips: improving pain measurement in the neonate. Adv Neonatal Care 18(4):260–266
33. Mahy M, Van Eycken L, Oosterlinck A (1994) Evaluation of uniform color spaces developed after the adoption of cielab and cieluv. Color Res Application 19(2):105–121
34. Walia A Key Frames Extraction. https://github.com/amanwalia123/KeyFra mesExtraction
35. Pexels (2023) : Neutral Expression. https://www.pexels.com/photo/yawning-newborn-i n-close-up-view-18070123/. Accessed: August 15
36. Pexels (2023) https://www.pexels.com/photo/a-sleeping-newborn-baby-14219543/. Accessed: August 15
37. Pexels (2024) : Close-up Photo of Crying Baby. https://www.pexels.com/photo/close-u p-photo-of-crying-baby-47090/. Accessed: February 26
38. Pexels (2024) : Smile Man in Front of White Wall. https://www.pexels.com/photo/por trait-photo-of-smiling-man-with-his-arms-crossed-standing-in-front-of-white-w all-2379004/. Accessed: February 26
39. Pexels (2024) : Woman in Collared Shirt. https://www.pexels.com/photo/woman-in-col lared-shirt-774909/. Accessed: February 26
40. Pexels (2024) : Screaming Woman in Gray Tank Top. https://www.pexels.com/photo/w oman-in-gray-tank-top-3812746/. Accessed: February 26
41. Pexels (2024) : Smile Man. https://www.pexels.com/photo/man-in-brown-button-up-s hirt-3785079/. Accessed: February 26
42. Pexels (2024) : Sad Female Crying and Looking at Camera. https://www.pexels.com/photo/sad-female-crying-and-looking-at-camera-6382719/. Accessed: February 26
43. Pexels (2024) : Woman with Band-aid on Her Face. https://www.pexels.com/photo/w oman-with-band-aid-on-her-face-5207298/. Accessed: February 26
44. Redmon J, Farhadi A (2018) : Yolov3: An incremental improvement. arXiv preprint arXiv :180402767
45. Yang S, Luo P, Loy C-C, Tang X (2016) : Wider face: A face detection bench- mark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5525–5533 14
46. He K, Zhang X, Ren S, Sun J (2016) : Deep residual learning for image recogni- tion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778
47. Chollet F (2023) : keras. https://github.com/fchollet/keras. Accessed: Sep. 4
48. Van Rossum G, Drake Jr FL (1995) : Python tutorial: Centrum voor wiskunde en informatica amsterdam. The Netherlands

49. Pearce T, Brintrup A, Zhu J (2021) : Understanding softmax confidence and uncer- tainty. arXiv preprint arXiv :210604972
50. Zhao F, Zhang C (2020) : Building damage evaluation from satellite imagery using deep learning. In: 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), pp. 82–89 IEEE
51. Cao Q, Shen L, Xie W, Parkhi OM, Zisserman A (2018) : Vggface2: A dataset for recognising faces across pose and age. In: 2018 13th IEEE International Con- ference on Automatic Face & Gesture Recognition (FG 2018), pp. 67–74 IEEE
52. Goodfellow IJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Cukierski W, Tang Y, Thaler D, Lee D-H et al (2013) : Challenges in rep- resentation learning: A report on three machine learning contests. In: Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3–7, 2013. Proceedings, Part III 20, pp. 117–124 Springer
53. Bentoumi M, Daoud M, Benaouali M, Taleb Ahmed A (2022) Improvement of emotion recognition from facial images using deep learning and early stopping cross validation. Multimedia Tools Appl 81(21):29887–29917

Figures

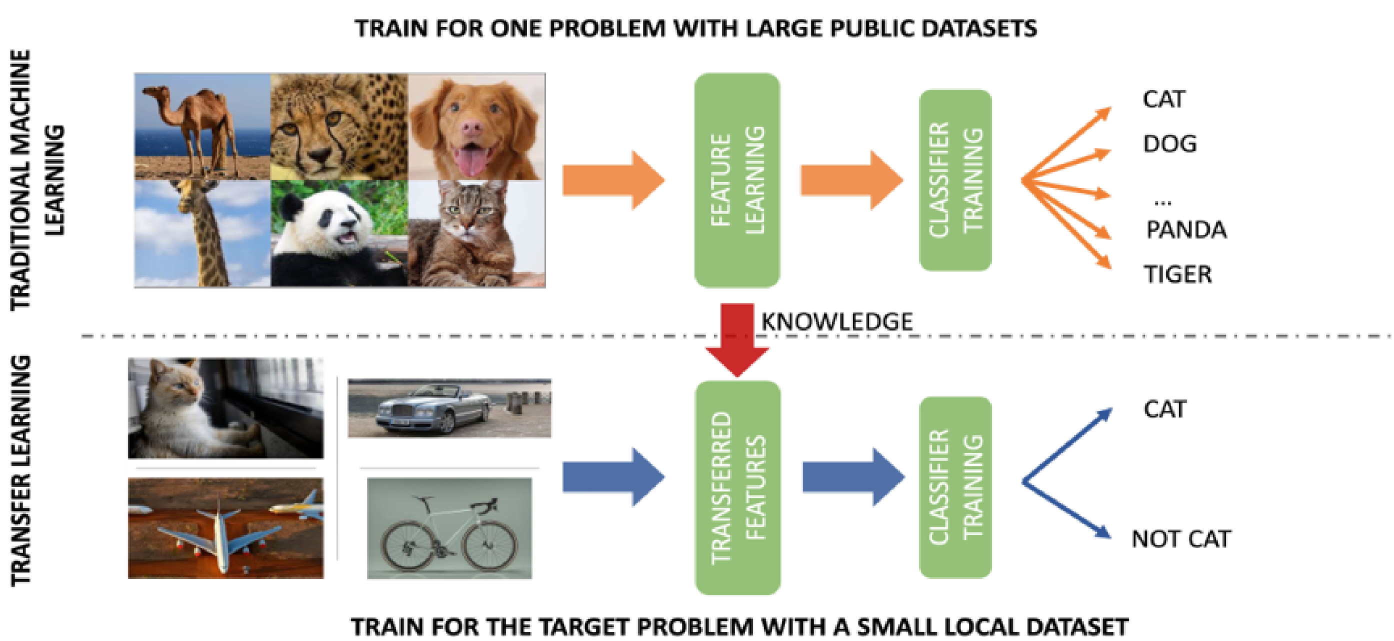

Figure 1

The illustration of transfer learning

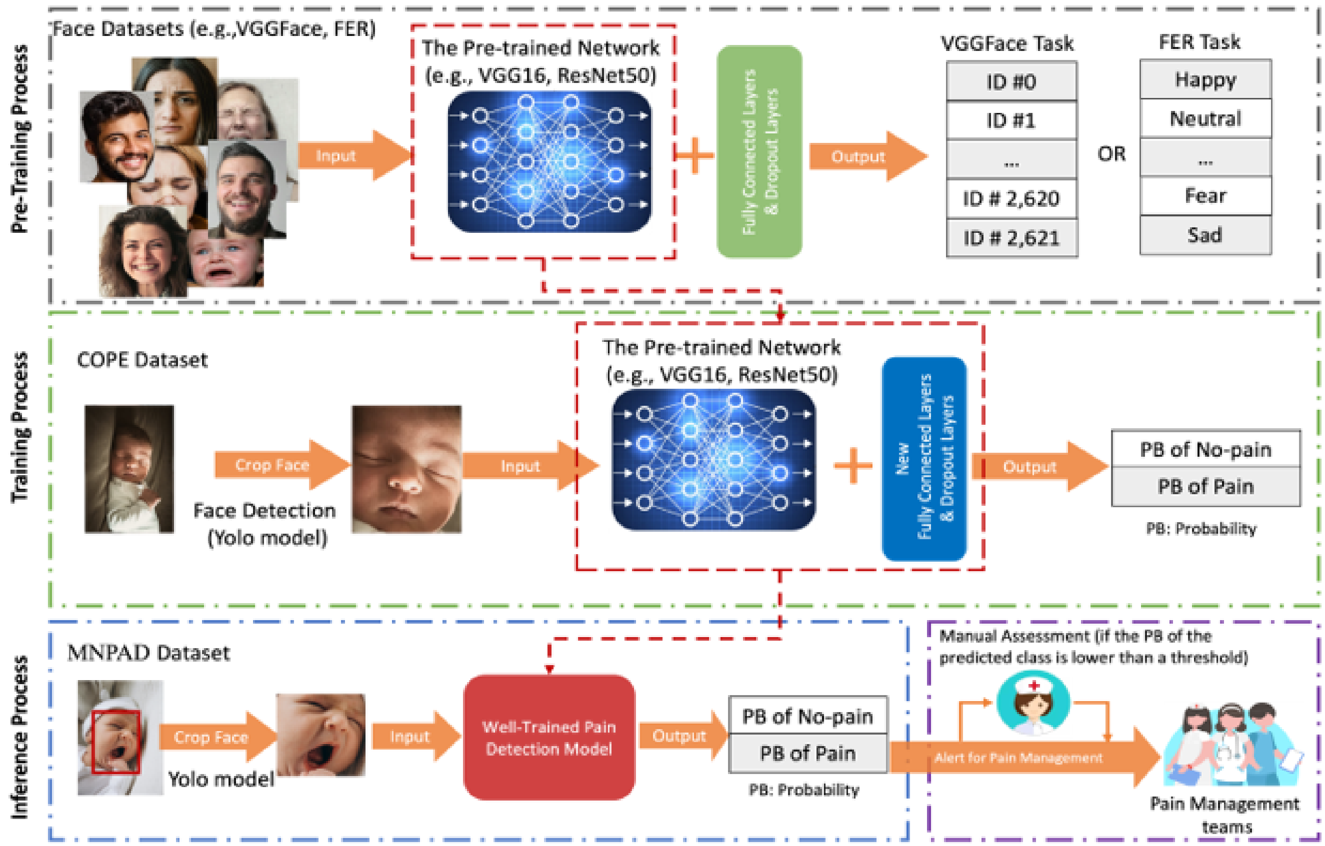

Figure 2

The framework architecture (human images are provided by online sources [35–43].)

Layer Name	Layer Details
FC #1	Fully Connected Layer, 512, Relu Activation
Dropout #1	Dropout Rate: 0.5
BN #1	Batch Normalization Layer, Default
FC #2	Fully Connected Layer, 128, Relu Activation
Dropout #2	Dropout Rate: 0.5
BN #2	Batch Normalization Layer, Default
Output	Fully Connected Layer, 2, Softmax Activation

Figure 3

The architecture of the pain detection classifier

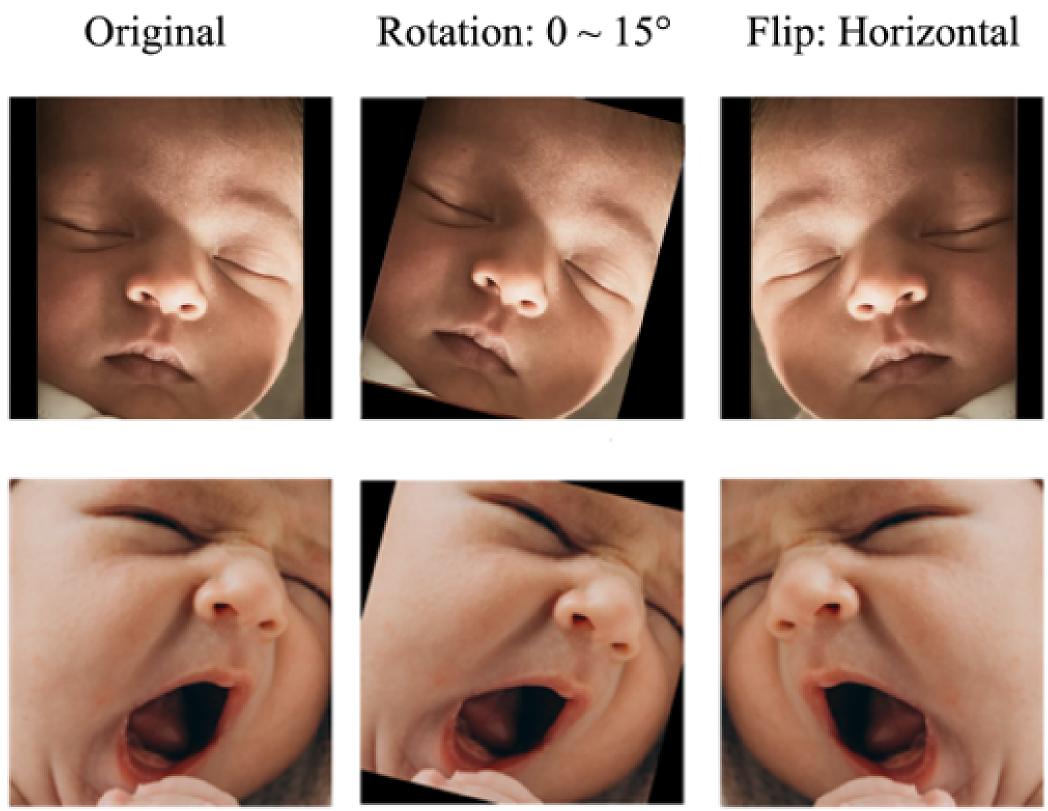

Figure 4

Minor rotations and flips are used to increase the amount of training data and to prevent our model from overfitting on the small dataset. (Neonatal images are provided by online sources [35, 36].)