**Pain classification using infant Facial expressions**

Prathibha H M

MSRIT  
prathibha@msrit.edu

Sriraam N  
Head R&D,MSRIT  
sriraam@msrit.edu

Tejswini   
MSRIT  
tejaswini.s@msrit.edu

*Abstract*—Infants with different health conditions are monitored at NICU. Different levels of pain are experienced by these neonates when neglected may lead to the death of the infant. The objective of this study is to monitor the neonates at NICU and detect the pain, discomfort experienced by the baby from facial expression analysis. In this context we aim at designing a monitoring system for infants which captures facial expressions of the infants and the algorithm detects the pain and its severity by analyzing the captured expressions. SVM classifier is used for both binary and Multiclass classifications. The given images are classified into Pain and no pain class under binary classification. Further images are classified into Mild, Moderate and Severe classes under multiclass classification. Algorithm gives an accuracy of 96% for binary class and 90% accuracy for multiclass classification.

Keywords—NICU, Pain Scale, SSIM, PSNR, Correlation Coefficient, Grid search, SVM classification

# Introduction

# New-born babies who need intensive medical care are often put in a special area of the hospital called the neonatal intensive care unit (NICU). Babies which are born preterm or with any health disorders are placed in NICU to provide special care and monitor continuously. New-born infants are exposed to painful experiences that might increase their short- and long-term morbidity and mortality, in addition to being associated with neurological developmental disorders [1].Pain is an unpleasant emotion associated with actual or potential tissue damage. Accurate assessment of pain is vital to ensure the optimal effectiveness and safety of pain management therapy in neonates who experience pain during the course of their NICU stay [2]. Accurate pain measurements in infants are difficult to achieve. Pain in infants are measured through behavioral changes such as crying, facial expressions, body posture, and movements. These behavior depends on infant’s gestational age, and maturity. Cry characteristics are also not good indicators in preterm or acutely ill infants, as it is difficult for them to produce a robust cry. Pain scales such as neonatal infant pain scale (NIPS); neonatal facial coding system (NFCS); neonatal pain, agitation, and sedation scale(N-PASS); cry, required oxygen, increased vital signs, expression, sleeplessness scale (CRIES); COMFORT Scale; and DouleurAigue Nouveau-ne (DAN) scoring system are used for pain validation.[3]. Infants cannot express pain verbally, so this impossibility has created the necessity of using other media for its evaluation and detection. In this way, pain scales based on vital signals and facial changes have been created to evaluate the pain of neonates. [4]

1. RELATED WORK

Ana Martínez et al, proposed a study that uses an automatic pain detection system by means of image analysis. The algorithm uses different texture descriptors like Local Binary Patterns, Local Ternary Patterns and Radon Barcodes along with SVM classifier for classification. SVM classifies input image into pain and non-pain classes.[4]

Cheng Li1 et al, proposes a real time monitoring system of young infants The system consists of two components: expression classification and expression state stabilization using Faster R-CNN and Hidden Markov Model for different expression classifications and discomfort detection[5]

Yosi Kristian et al, proposed a study uses a number of features based on action units for pain classification among infants. Active shape modeling (ASM) is used to extract geometrical features and facial boundaries, local binary pattern is used to extract texture features. Multi stage SVM classifier is used for severe pain classification.[6]

Yue Sun et al, proposed an automatic and continuous system to identify discomfort in infants. Motion trajectories of image are estimated using optical flow. For each video motion acceleration rate and extract 18 time- and frequency-domain features characterizing motion patterns are calculated. SVM classifier is used to classify comfort and discomfort among infants.[7]

Ruicong Zhi et al proposed a study for the assessment of infant pain based on dynamic pain facial expressions and fusion scheme for automatic pain assessment in infants by combining temporal appearance facial features and temporal geometric facial features. The effects of various factors that influence pain reactivity in infants, such as individual variables of gestational age, gender, and race are investigated. SVM classifier is used for infant pain recognition.[8]

Lykele Hazelhoff et al proposed an automated video survey system for the detection of discomfort in newborn babies by analysis of their facial expression. Proposed algorithm automatically segments the face from the background and localizes the eye, eye-brow and mouth regions a hierarchical classifier is employed to distinguish between the states sleep, awake and discomfort.[9]

Yue Sun et al proposed a system for automatic discomfort detection in infants by analyzing infant’s facial expression. Deep convolution Neural Network based on DenseNet is employed. The performance of the deep-learning model is improved when using our proposed strategic fine-tuning involving pre-training with generic people pictures and dataset balancing combined with two-fold cross-validation.[10]

.

1. Materials of methods
   * 1. Dataset
2. The Infant COPE Database, contains 204 facial images of 26 neonates experiencing the pain of a heel lance and three non-pain stressors: transport from one crib to another (a stressor that triggers crying that is not in response to pain), an air stimulus on the nose (a stressor that provokes eye squeeze), and friction on the surface of the heel (a stressor that produces facial expressions of distress that are similar to the expressions of pain). In addition to these four facial displays, the database includes images of the neonates in the neutral state of rest. All photographs were taken using a Nikon D100 digital camera under ambient lighting conditions in a room separated from other newborns.[11]. The protocols and ethical directives for research involving human subjects at St. John's Health System, Inc. Informed consent was obtained from a parent, usually the mother in consultation with the father. Most parents were recruited in the neonatal unit of a St. John's Hospital sometime after delivery. Only mothers who had experienced uncomplicated deliveries were approached.
3. The PT-FT NICU selected for analysis are recorded from NICU at MSRTH, Bengaluru using Wi-Fi camera and mobile camera. NICU includes both full term and preterm neonates with gestation age ranging from 28-37 weeks and with varied health conditions. Different activities of the baby are recorded which includes cries and sleep of the baby using Fingers 1080 Hi-Res Webcam. 200 videos of neonates are recorded out of which 130 videos are discarded due to repeated subjects and unfocused baby face due to disturbance of camera stand by the clinical staff for maintenance purposes .70 videos are available for analysis. Videos are recorded at a frame rate of 30 fps. Camera was placed beside the infant warmer focusing the baby face area with the help of a customized stand. The stand is undisturbed during the process of recording as shown in Figure 1.



**Figure 1 NICU camera set-up**

**Table 1 Details of PT-FT NICU database**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Patient Id | Gender | Gestation week(weeks) | Duration  (Min) | Frame rate | Camera used | Diagnosis |
| PTV001 | M | 34 | 04.09 | 30fps | Wifi cam | PT |
| PTV017 | M | 37 | 01.10 | 30fps | Mobile | PT/AGA |
| PTV018 | M | 28 | 03.39 | 30fps | Mobile | PT/AGA/RDS |
| PTV016 | M | 32 | 21.55 | 30fps | Wifi cam | PT |
| PTV015 | M | 32 | 03.00 | 30fps | Mobile | PT |
| FTV003 | M | 40 | 03.00 | 30fps | Mobile | SGA |
| FTV001 | M | 37 | 21.96 | 30fps | Wifi cam | AGA/MAS |
| FTV002 | M | 40 | 21.55 | 30fps | Wifi cam | AGA/ MSAF |
| PTV006 | F | 33 | 21.55 | 30fps | Wifi cam | PT/AGA |
| PTV007 | M | 33 | .21.55 | 30fps | Wifi cam | PT/AGA/Rh-ve mother |
| PTV008 | M | 33 | 21.55 | 30fps | Wifi cam | PT/AGA/Male |
| PTV009 | F | 34 | 21.55 | 30fps | Wifi cam | PT / AGA/ M |
| PTV010 | F | 34 | 21.55 | 30fps | Wifi cam | PT / AGA/ M |
| PTV011 | M | 36 | 21.55 | 30fps | Wifi cam | Late PT |
| PTV012 | M | 36 | 21.55 | 30fps | Wifi cam | PT/Perinatal depression |
| PTV013 | F | 29 | 21.55 | 30fps | Wifi cam | Early PT/AGA |
| PTV002 | M | 34 | 21.96 | 30fps | Wifi cam | PT/AGA/IDM |

Table 1 gives the details of neonates recorded NICU from MSRMH, Bengaluru. Dataset details include infant’s Gender, Gestational age, Duration of recorded video, Frame rate, Camera used for recording and diagnosis details which tells the reason for admitting infant’s to NICU.

**Diagnosis details**

* PT- Preterm
* AGA-Appropriate for Gestational Age
* RDS-Respiratory distress syndrome
* SGA-Small for gestational age
* MSAF-Meconium-stained amniotic fluid
* IDM -Infants of diabetic mothers

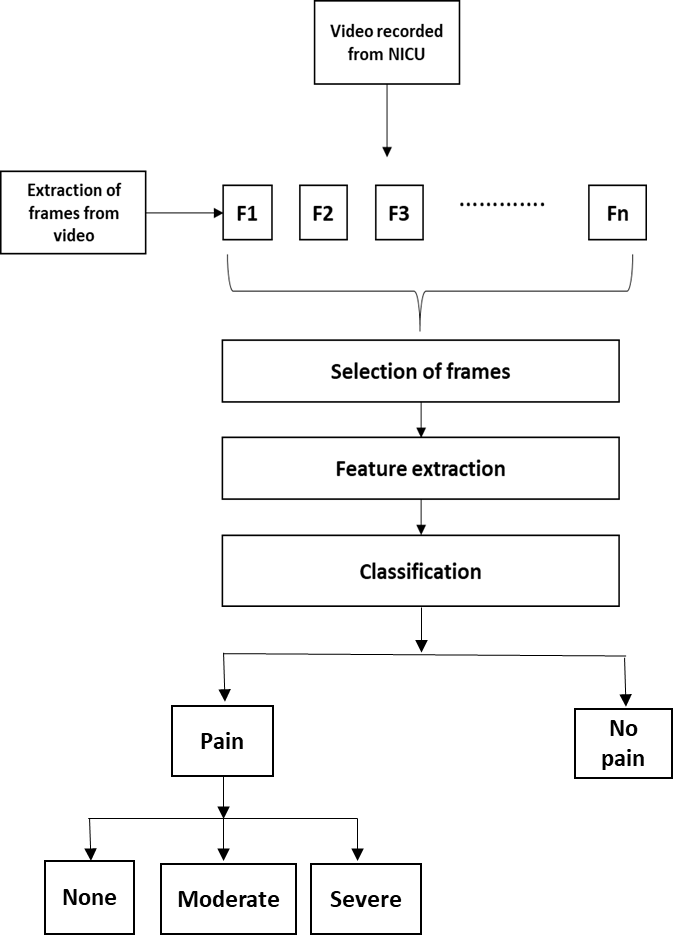


**Figure 2 Frames extracted from recorded videos**

Figure 2 shows the frames extracted from recorded videos. Still picture of the baby is clicked at the before starting recording which is considered as reference frame ‘I’.

* + 1. Overall framework

In the proposed work frames are extracted from recorded videos recorded. Selection of frames is based on SSIM, PSNR, Correlation and randomized selection criteria. Grid search cv is used for feature extraction and SVM classifier is used for classifying the given images into two classes i.e. binary class- Pain and No pain, multi class- None, Moderate and Severe classes.



**Figure 3 Proposed framework**

* + 1. Preprocessing and frame selection

The videos are recorded from the preterm and full term neonates with varied health conditions at NICU of for facial expression analysis. These videos are annotated by clinical staff or doctors based on standard clinically accepted pain scale. Videos are first converted into frames. 20% of frames are taken for analysis. These frames are then selected based on following criteria

(a)Case 1 - Selecting Random frames -20% of frames are taken for analysis by considering first, mid and last frame.

(b) Case 2- Selecting frames based on SSIM (Structural Similarity Index) ‘I’ is taken as reference frame, SSIM is performed by based on given threshold value.

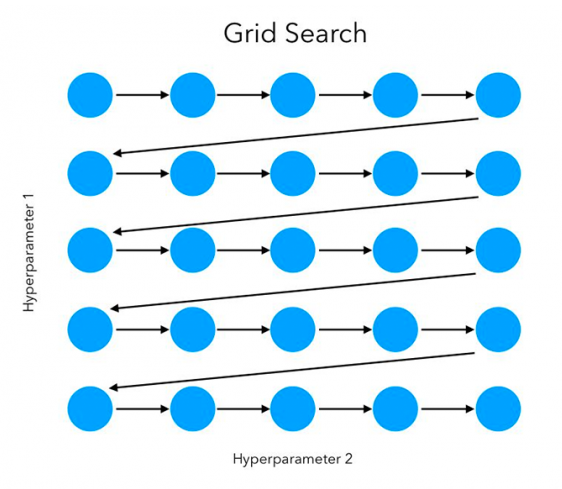
(c)Case 3- Selection based on PSNR value.

(d) Case 4- Selection based on Correlation.

(e) Case 5- Selecting all the frames.

* + 1. Feature extraction

Features are extracted from the deformations that occurs in the infant face. Grid search Cross-Validation is used to extract the features. It is a technique to select the best of the machine learning model, parameterized by a grid of hyper parameters. tries all combinations of parameters grid for a model and returns with the best set of parameters having the best performance score. It is a technique to select the best of the machine learning model, parameterized by a grid of hyper parameters. tries all combinations of parameters grid for a model and returns with the best set of parameters having the best performance score. The hyper parameters are set up in a discrete grid and then it uses every combination of the values in the grid, evaluating the performance using cross-validation. The point of the grid that maximizes the average value in cross-validation, is the optimum combination of values for the hyper parameters. Grid search is used along with SVM classifier.



**Figure 4 Grid search representation**

* + 1. Classification

In order to classify the features effectively, Support Vector Machine (SVM) classifier is used. algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane called as support vectors can be attributed to different classes. A kernel is a function that assigns to each pair of elements a real value corresponding to the scalar product of the transformed version of that element in a new space. The selection of the kernel depends on the application. Two SVM model parameters are

**Gamma:** if low value of gamma is used then model will consider that data points also are used which are far from hyperplane and if high gamma value is used then nearest point is considered with more weight. gamma defines how much influence a single training example has. The larger gamma is; the closer other examples must be to be affected.

**C**: controls tradeoff between smooth decision boundary and classifying training points correctly. The parameter C, common to all SVM kernels, trades off misclassification of training examples against simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly.

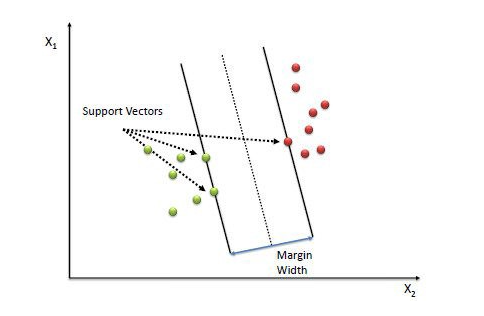
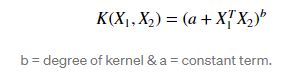
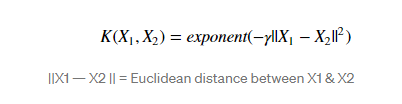


Figure 5 Support vector points

* Polynomial:  In general, the polynomial kernel is defined as, 

Polynomial kernel is calculated by increasing the dot product by power of the kernel.

* **Radial basis function kernel (RBF)/ Gaussian Kernel:** RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format

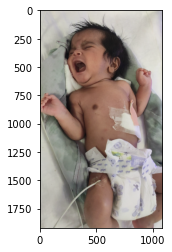


where γ is specified by parameter gamma, must be greater than 0.

1. **Results**

Classification is categorized into Binary and Multiclass classification. Videos are annotated by clinicians based on the clinical standard pain scale. Frames extracted from these videos are used for supervised training. Videos recorded from both Preterm and full-term infants with different health conditions are used for analysis. Infant Cope dataset are also used along with preparatory dataset. 80% of images are used for training and 20% of images are used for testing.

1. Binary classification- Categorized into pain and no pain classes. Model is trained with Infant COPE database images and tested with proprietary database images for cross validation.
2. Case 1- Training with open source data and testing with proprietary database.



Pain = 56.094791350366926%

No pain = 43.90520864963307%

The predicted image is : Pain

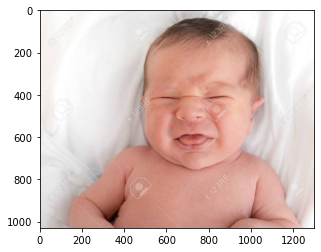
1. Case 2 – Training with proprietary data and testing with unknown data.



No pain = 99.0426916147363%

Pain = 0.9573083852637106%

The predicted image is: No pain

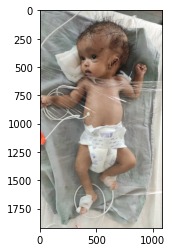


Pain = 76.10128328122552%

No pain = 23.898716718774498%

The predicted image is: Pain

1. Multi class classification- Categorized into None, Moderate and Severe classes. Classifier is trained with 104 ‘None’ class, 100 ‘Moderate’ class and 104 ‘Severe’ class images. Testing is done with images taken from YouTube videos and open source database images. The unbalanced number of images is due to the number of pictures of each class available in the database.

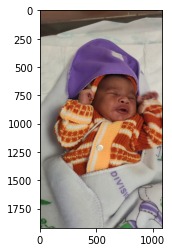


**moderate = 25.095100341583905%**

**Mild = 60.30653489330289%**

**severe = 14.598364765113189%**

**The predicted image is: None**

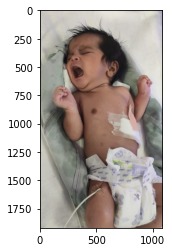


**moderate = 95.3253408163106%**

**Mild = 2.420899909551253%**

**severe = 2.2537592741381443%**

**The predicted image is: moderate**



**moderate = 8.648209596205698%**

**Mild = 3.2216067411114846%**

**severe = 88.13018366268282%**

**The predicted image is: Severe**



**moderate = 40.38214589165237%**

**Mild = 20.109537893228474%**

**severe = 39.50831621511915%**

**The predicted image is: moderate**



**moderate = 7.648209596205698%**

**Mild= 4.5216067411114846%**

**severe = 88.27018366268282%**

**The predicted image is: Severe**

* + 1. Evaluation matrix
* Accuracy-It is the ratio of number of correct predictions to the total number of input samples.

Accuracy=

* Sensitivity -True Positive Rate is defined as TP/ (FN+TP). True Positive Rate corresponds to the proportion of positive data points that are correctly considered as positive, with respect to all positive data points.

Sensitivity=

* Specificity -True Negative Rate is defined as TN / (FP+TN). False Positive Rate corresponds to the proportion of negative data points that are correctly considered as negative, with respect to all negative data points.

Specificity =

* Precision- It is the number of correct positive results divided by the number of positive results predicted by the classifier.

Precision=

* Recall: It is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

Recall=

* F1 Score - It is the Harmonic Mean between precision and recall. The range for F1 Score is [0, 1]

F1 score= 2\*

**Table 2 Binary classification Case1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
| Accuracy | 98.87±0.15 | 97.35±0.19 | 98.6±0.13 | 97.57±0.11 | 98.21±0.23 |
| Sensitivity | 93.75±0.23 | 93.6±0.25 | 94.88±0.18 | 95.48±0.15 | 93±0.04 |
| Specificity | 98.96±0.01 | 98.96±0.01 | 98.23±0.14 | 99±0.17 | 98.95±0.14 |
| F1 Score | 95±0.23 | 96±0.13 | 95.3±0.07 | 95.5±0.03 | 97.2±0.012 |

**Table 3 Binary classification Case 2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Case 1** | **Case 2** | **Case 3** | **Case 4** | **Case 5** |
| **Accuracy** | 96.87±0.15 | 95.35±0.19 | 95.6±0.13 | 96.57±0.11 | 96.21±0.23 |
| **Sensitivity** | 93.75±0.23 | 92.6±0.25 | 92.88±0.18 | 93.48±0.15 | 93±0.04 |
| **Specificity** | 99.96±0.01 | 99.96±0.01 | 98±0.14 | 99±0.17 | 99.95±0.14 |
| **F1 Score** | 97±0.23 | 96±0.13 | 96.3±0.07 | 97.5±0.03 | 97.2±0.012 |

Table 2 and 3 describes evaluation parameters such accuracy, sensitivity, specificity and F1 score for all the case of binary classification

**Table 4 Multiclass Classification**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Case 1** | | | **Case 2** | | | **Case 3** | | | **Case 4** | | | **Case 5** | | |
| **Mild** | **Moderate** | **Severe** | **Mild** | **Moderate** | **Severe** | **Mild** | **Moderate** | **Severe** | **Mild** | **Moderate** | **Severe** | **Mild** | **Moderate** | **Severe** |
| **Sensitivity** | 95±0.15 | 85.71±0.23 | 85.71±0.05 | 93±0.23 | 87.14±0.22 | 87.14±0.25 | 94±0.12 | 83.71±0.17 | 83.71±0.23 | 96±0.45 | 86.71±0.24 | 86.71±0.03 | 96±0.24 | 86.71±0.08 | 86.71±0.09 |
| **Specificity** | 88.09±0.17 | 97.56±0.41 | 97.56±0.16 | 86.09±0.12 | 95.62±0.31 | 95.62±0.35 | 84.09±0.21 | 95.63±0.13 | 95.63±0.25 | 89.46±0.28 | 95.6±0.16 | 95.6±0.18 | 88.09±0.25 | 95.6±0.24 | 95.6±0.48 |
| **Precision** | 79.16±0.23 | 94.73±.036 | 94.73±0.31 | 71.66±0.22 | 93.73±0.27 | 93.73±0.24 | 75.16±0.14 | 91.73±0.15 | 91.73±0.28 | 80.16±0.34 | 95.73±0.18 | 95.73±0.23 | 78.16±0.26 | 92.73±0.36 | 92.73±0.16 |
| **Accuracy** | 90.32±0.05 | 93.54±0.51 | 93.54±0.41 | 89.32±0.35 | 90.54±0.26 | 90.54±0.14 | 93.22±0.16 | 90.54±0.26 | 90.54±0.14 | 93.22±0.15 | 93.54±0.19 | 93.54±0.15 | 92.32±0.34 | 91.54±0.25 | 91.54±0.27 |

Table 3 represents the values of accuracy, sensitivity, specificity and precision values for Multiclass classification. Case 4 and Case 5 performs better compared to other cases for the given data. Over all Case

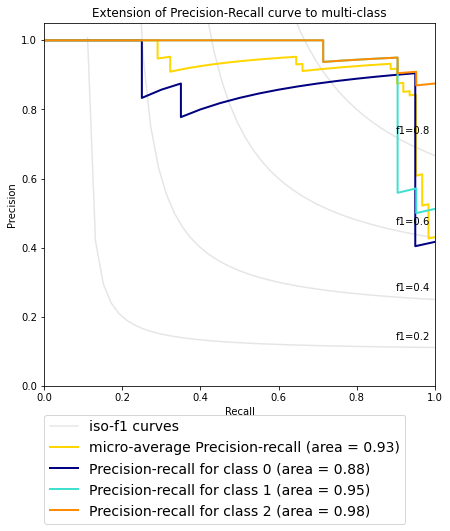
5, performs better for both binary and multi class classification.

1. **Performance evaluation**
2. Binary class classification
3. Case 1
4. Case 2

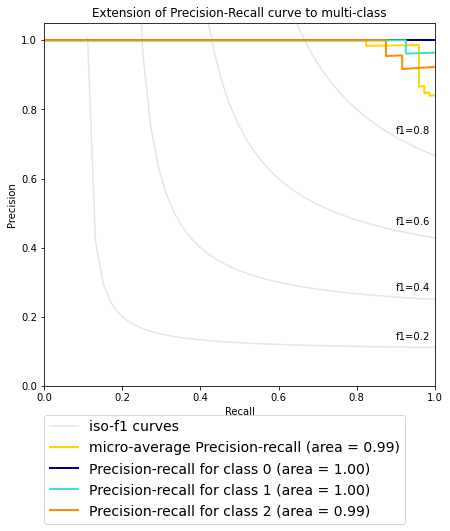
Algorithm gives an average accuracy of 96% in all cases with 99±0.17 specificity and 93.75±0.23 for binary class for proprietary vs unknown data. 98% average accuracy for open source vs proprietary data.

1. Multi class classification

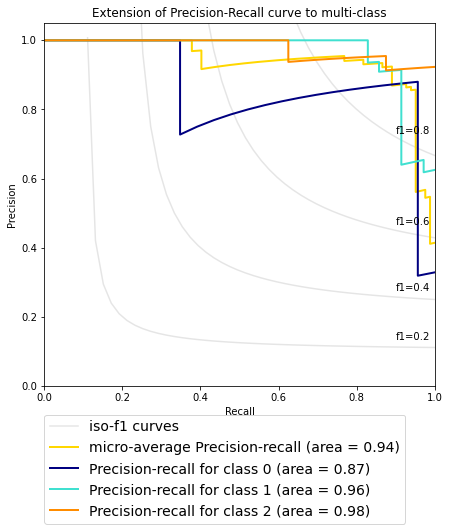
None, Moderate & Severe



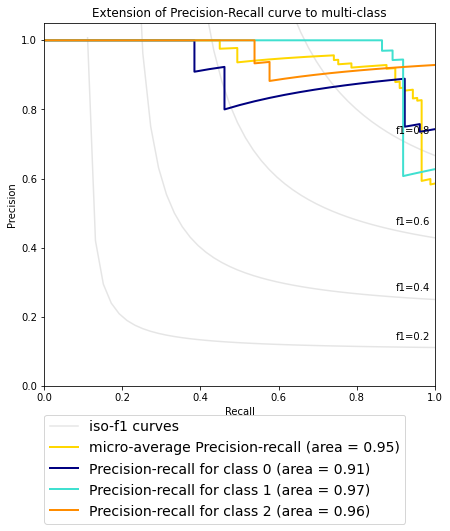
**Figure 6 PR graph for Multiclass classification for case1**



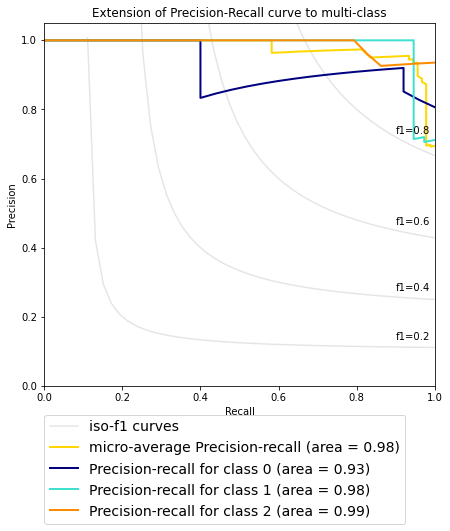
**Figure 7 PR graph for case2**



**Figure 8 PR graph for case 3**



**Figure 9 PR graph for case 4**



**Figure 10 PR graph for case5**

Precision-Recall curves summarize the trade-off between the true positive rate and the positive predictive value for a predictive model using different probability thresholds. Figures 6-10 shows Precision Recall graphs for multi class classification of None, moderate and Severe pain classes. From the above figures we can infer that Case 2: Selecting frames based on SSIM index with I’ is taken as reference frame, SSIM is performed by based on given threshold value gives the best precision recall value.

1. Discussion

The videos of the neonates are recorded from NICU. Videos are converted to still frames. The best frames are selected using PSNR, SSIM and Correlation techniques. Infant COPE database are used for analysis. Features are extracted using Grid search CV along with SVM classifier. Classifier are trained with PTFT NICU (proprietary) database and infant COPE database and tested with completely unknown data. Infant face images are classified into Pain and No Pain classes initially. Algorithm gives an average accuracy of 96% in all cases with 99±0.17 specificity and 93.75±0.23 for proprietary vs unknown data and 98% average accuracy for open source vs proprietary data. Further images are classified into None, Moderate and Severe classes. Algorithm gives an accuracy of 90% for multiclass for both proprietary and open source databases.

1. Conclusion

The proposed algorithm gives an accuracy of 95% when trained with open source as well as proprietary database. The size of the preparatory database needs to be improvised. This algorithm needs to be further tested with different kernels for better performance. The robustness of the algorithm can be improvised by using other open source database for testing. Our future study aims at continuous monitoring of infants at NICU which detects the severity of pain experienced by neonates in real time scenarios and sleep / wake cycle detection among the neonates.

##### References

[1] T. M. Heiderich, A. Teresa, F. Stochero, and R. Guinsburg, “Neonatal procedural pain can be assessed by computer software that has good sensitivity and specificity to detect facial movements,” pp. 63–69, 2015, doi: 10.1111/apa.12861.

[2] L. G. Maxwell, M. V Fraga, and C. P. Malavolta, “Assessment of Pain in the N e w b o r n : An Update Neonate Pain Assessment Pain scales,” vol. 46, pp. 693–707, 2019, doi: 10.1016/j.clp.2019.08.005.

[3] R. Srouji, S. Ratnapalan, and S. Schneeweiss, “Pain in Children : Assessment and Nonpharmacological Management,” vol. 2010, 2010, doi: 10.1155/2010/474838.

[4] A. Martínez, F. A. Pujol, and H. Mora, “Application of texture descriptors to facial emotion recognition in infants,” *Appl. Sci.*, vol. 10, no. 3, 2020, doi: 10.3390/app10031115.

[5] C. Li, A. Pourtaherian, L. Van Onzenoort, and P. H. N. De With, “Automated infant monitoring based on R-CNN and HMM,” *VISIGRAPP 2021 - Proc. 16th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl.*, vol. 5, no. Visigrapp, pp. 553–560, 2021, doi: 10.5220/0010299605530560.

[6] Y. Kristian *et al.*, “A Novel Approach on Infant Facial Pain Classification using Multi Stage Classifier and Geometrical-Textural Features Combination,” no. February, 2017.

[7] Y. Sun *et al.*, “Automatic and Continuous Discomfort Detection for Premature Infants in a NICU Using Video-Based Motion Analysis,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 5995–5999, 2019, doi: 10.1109/EMBC.2019.8857597.

[8] R. Zhi, G. Zamzmi, D. Goldgof, T. Ashmeade, and Y. Sun, “Automatic Infants ’ Pain Assessment by Dynamic Facial Representation : Effects of Profile View , Gestational Age , Gender , and Race,” pp. 1–16, doi: 10.3390/jcm7070173.

[9] L. Hazelhoff and J. Han, “Behavioral State Detection of Newborns Based,” pp. 698–709, 2009.

[10] A. Manuscript, “Ac ce d M us pt,” 2019.

[11] S. Brahnam, C. Chuang, R. S. Sexton, and F. Y. Shih, “Machine assessment of neonatal facial expressions of acute pain,” vol. 43, pp. 1242–1254, 2007, doi: 10.1016/j.dss.2006.02.004.