This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

NeoCam: An edge-cloud platform for non-invasive real-time monitoring in neonatal intensive care units

Angel Ruiz-Zafra[1], Daniel Precioso[1], Blas Salvador, Simón P. Lubián-López, Javier Jiménez, Isabel Benavente-Fernández, Janet Pigueiras, David Gómez-Ullate[2], Lionel C. Gontard

***Abstract*—In this work we introduce NEOCAM, an open source hardware-software platform for video-based monitoring of preterms infants in Neonatal Intensive Care Units (NICUs). NEOCAM includes an edge computing device that performs video acquisition and processing in real-time. Compared to other proposed solutions, it has the advantage of handling data more efficiently by performing most of the processing on the device, including proper anonymisation for better compliance with privacy regulations. In addition, it allows to perform various video analysis tasks of clinical interest in parallel at speeds of between 20 and 30 frames-per-second. We introduce algorithms to measure without contact the breathing rate, motor activity, body pose and emotional status of the infants. For breathing rate, our system shows good agreement with existing methods provided there is sufficient light and proper imaging conditions. Models for motor activity and stress detection are new to the best of our knowledge. NEOCAM has been tested on preterms in the NICU of the University Hospital Puerta del Mar (Cádiz, Spain), and we report the lessons learned from this trial.**

***Index Terms*—NeoCam, preterm neonate, NICU, non-contact monitoring, edge computing, body pose, facial expression**

A. Ruiz-Zafra is with the Department of Software Engineering, University of Granada. 18071, Granada, Spain (e-mail: angelr@ugr.es)

D. Precioso, J. Jimenez, and B. Salvador are with the Department of Computer Science, Higher School of Engineering, University of Cádiz, E11519 Cádiz, Spain (e-mail:{daniel.precioso, javier.jimenez, blas.salvador}@uca.es)

I. Benavente-Fernández and S.P. Lubián-Lopez are with Division of Neonatology, Department of Paediatrics and with Biomedical Research and Innovation Institute of Cádiz (INiBICA) Research Unit, Puerta del Mar University, Cádiz, Spain (e-mail: simonplubian@gmail.com, isabel.benavente@gm.uca.es)

I. Benavente-Fernández is also with Area of Paediatrics, Department of Child and Mother Health and Radiology, Medical School, University of Cádiz, Cádiz, Spain

Janet Pigueiras is with the Department of Condensed Matter Physics, University of Cádiz, E11510 Cádiz, Spain (email: janet.pigueiras@uca.es)

D. Gómez-Ullate is with the School of Science and Teechnology, IE University, Madrid, Spain (e-mail:david.gomezullate@ie.edu)

L. C. Gontard is with the Department of Condensed Matter Physics, Applied Magnetism and Optics Research Group, and IMEYMAT, University of Cádiz, E11510 Cádiz, Spain (e-email: lionel.cervera@uca.es)

[1] A. Ruiz-Zafra, D. Precioso have contributed equally to the design and development of NeoCam platform.

[2] D. Gómez-Ullate has contributed to the NeoCam platform as coordinator of the node UCADatalab.

I. INTRODUCTION

Preterm infants born before the 37th gestational week have a high risk of developing complications during early extrauterine life, mainly respiratory, cardiovascular, infectious disorders and brain injury. Continuous monitoring in Neonatal Intensive Care Units (NICUs) is important for measuring the degree of healthy development, and for tuning neonatal care delivery [1]. The prevalent neurodevelopmental consequences of preterm birth, including cerebral palsy, epilepsy and cognitive impairment impact multiple domains of the future life of the infant [2]. The early identification of infants at a greater risk would allow to target early therapeutic intervention strategies [3], [4].

The vital signs of the neonates are measured continuously with sensors that are glued or attached to their body. These sensors can damage the immature epidermis of these infants, and because they are connected with several cables to electronic monitors located close to the incubators, they are also inconvenient for the medical staff who have to handle the infants frequently. Serial cranial neuroimaging [5] or EEGs are used for monitoring brain structures' maturation but only sporadically because the data requires extensive analysis carried out by trained personnel. Moreover, a non-invasive qualitative evaluation (mainly visual) may be used by medical staff to assess the overall comfort and health of the patients and may reveal an altered neurodevelopmental trajectory of the infant. These variables include assessing periods of pain or stress, lethargy, or abnormal movements. Nevertheless, a reliable identification of infants' emotions or specific patterns of motor activity is still a rather difficult, episodic and subjective task, which requires expertise of the clinician performing the evaluation. It is also time consuming and limited by skill, fatigue, and inter-rater reliability of the operators.

For all the above reasons, non-contact forms of smart monitoring based on video are being intensively investigated and they may well become ubiquitous in future NICUs [6], [7].

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

In this paper we present NEOCAM, an open-source software-hardware platform designed to support a non-invasive video-based monitoring of neonates at NICUs. It is composed by three main elements:

1) the NEOCAM Unit, an edge-computing device to collect and process video information in real-time using computer vision and AI algorithms,
2) the cloud infrastructure that supports most of the functionalities as services
3) an Android-based application for end-users (healthcare professionals and relatives) to consume services provided by NEOCAM.

The main contributions of the NEOCAM platform are that:

- it is a complete monitoring system, that comprises edge computing cameras, computer vision models, IoT network architecture, cloud services and a mobile app.
- all of the hardware elements are easily accessible in the market, and all of the software is open source and made available in a *Github* repository.

The paper is organized as follows. Section II shows a review of related projects on the literature. Section III presents an overview of the system architecture (further technical details can be found on the Supplementary Material). The main algorithms to perform the various monitoring tasks are described in Section IV together with their performance metrics. Experimental tests of the NEOCAM system at the NICU of Puerta del Mar University Hospital in Cádiz, Spain, are described in Section V). Finally, an overall summary and a discussion of future work are discussed in Section VI.

II. RELATED WORK

Video-based monitoring in NICUs has emerged strongly in the last years, with multiple potential applications as summarized in Table I. It has a bright future thanks to the developments in hardware (low-cost embedded platforms) and software (AI algorithms). Video is non-invasive, if appropriately designed videos can be used 24/7, and with the emergence of AI models it can now be used for multiple human-like analysis tasks.

It is being investigated in applications for non-contact monitoring of blood pressure and oxygenation, heart rate or respiration rate [8], [9], [10], with the aim of minimising the number of cables attached to the infants and also to reduce skin damage from the sensors. Another field of application is for the evaluation of lethargy/sleep periods [11], [12], for the emotional state identification [13], [14], detecting the presence of the infant [15] and for the general movements assessment (GMA) in neonates [16], [17].

General Movements (GMs), a type of spontaneous movement allows the integrity of the central nervous system to be assessed and to detect possible neurological alterations early in life [18]. For example, Prechtl's method for the assessment of GMs has attracted attention because it allows motor-function-based diagnosis without the need to apply stimuli. In clinical practice, GMs are visually assessed by qualifed licensees; however, this presents a difficulty in terms of quantitative evaluation. As a result there is an increasing interest in using video-analysis methods for the automated quantitative evaluation of GMs which include image segmentation using frame differences [19], or green screen background [20] to obtain the silhouette of the infants in order to identify the trajectories and velocities of the movement of the limbs, 3D joint extraction from 3D maps, and deep learning for pose analysis from features extracted on synthetic datasets like *SMIL* [21], [22]. More recently, learning-based human pose detectors such as *AlphaPose* have been used for markerless quantitative movement assesment [23].

In parallel, for a practical use in real NICUs a current field of research concerns the design of specific architectures with user-oriented applications, capable of combining information from several sensors [9], [24], [25], and possibly including dedicated data annotation tools, like *Avim* [26] or *Movidea* [27]. Probably, the most advanced hardware-software platform today is *Voxyvi* [28], consisting on hardware for recording sound and multi-spectral video in the infrarred and visible ranges, and software for managing and automating the recordings. This software includes AI features for automated annotation of the presence of the infants in the field-of-view. NEOCAM can be included in this field of research, with the novel application of edge computing to the problem of monitoring NICUs with a network of video sources.

III. SYSTEM ARCHITECTURE

The design, development and deployment of NEOCAM platform is a process where several technologies and techniques are involved (i.e. AI algorithms, cloud computing, computer vision, Internet of Things, software development, etc). Figure 1 illustrates the global system architecture of NEOCAM platform, along with the different technologies, stakeholders, components and interaction between them.

The NEOCAM Unit is the core element of the NEOCAM platform, which performs the following tasks: sensing of images, recording of information, image processing with AI algorithms and information exchange with the cloud through a hybrid access network (WiFi + Celullar Data). The NEOCAM Unit operates with an OAK-D camera from Luxonis[1]. It has two monochrome and one RGB image sensors and integrates the Myriad X coprocessor, which is a Vision Processing Unit (VPU) from Intel®. The two

[1]www.luxonis.com/

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

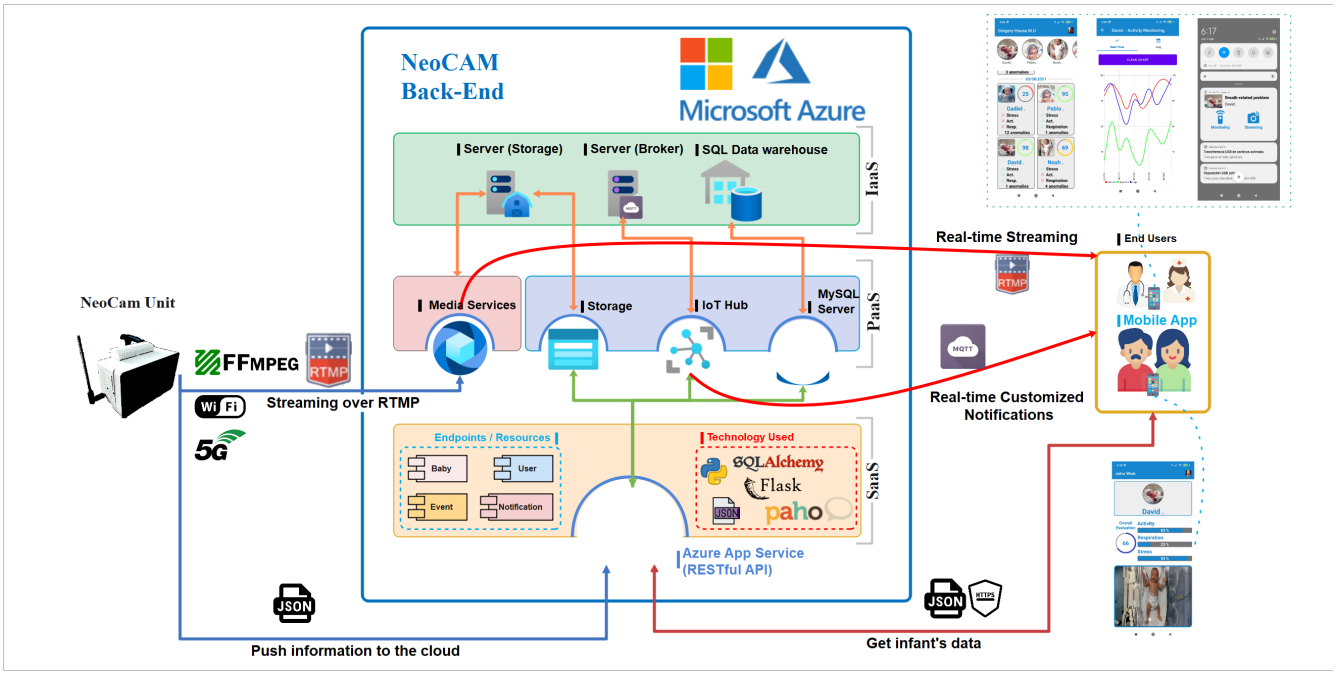

Figure 1: NeoCam's ecosystem includes the NEOCAM Unit, the cloud infrastructure, computer vision algorithms and a mobile App.

stereo monochrome cameras allow object depth perception, which is used for the breathing rate measuring algorithm (Section IV). The VPU module has a computing capacity of 4 trillion Ops/sec which allows to process frames with several deep learning ML models (CNNs) running in parallel at frame speeds of 20-30 fps.

The NEOCAM Unit pushes information asynchronously to the services supported by the cloud-based backend. These services focus on data and CRUD (Create, Read, Update and Delete) operations, real-time streaming media services and real-time customized notifications. These services are consumed by end-users through the functionalities provided by the mobile application.

An entire technical report about the NEOCAM platform is available as Supplementary Material. This technical report describes in depth the system architecture through the *4+1 architectural view model*, with the aim to help software engineers in the understanding and replication of NeoCam.

It is worth stressing that all code and libraries are open-source, which enables any research group to build upon the work we release, and replicate the development, assembling and deployment of the entire platform.

IV. ALGORITHMS FOR NON-CONTACT MONITORING

This section explains in detail the machine learning and computer vision techniques implemented in the NEOCAM platform which include the use of body and face detection and classification, passive depth mapping and pose estimation. These techniques can be potentially applied to solve different clinical applications as described in Table I, and they use solely video frames as input data.

To process a frame, the NEOCAM system first applies two CNNs that recognize and locate the body and face of the neonate (Section IV-A). The outputs of these models are then fed to three parallel modules. The first module uses the face position to estimate the torso location, and by analysing its oscillatory motion, the system is capable of measuring the breathing rate of the neonate (Section IV-B). The second module computes a motor activity index, by estimating the neonate's pose using both body and face detection (Section IV-C). The last module crops the neonate's face and sends it to a third CNN that performs emotion analysis (section IV-D).

A. Body and face recognition

The first step of the preprocessing pipeline is to locate the head and body of the infant in the image. Two CNN were used for this task. For simplicity, we named these models *CNN-face* and *CNN-body*. The models we used are known as residual networks (ResNet), whose architecture is aimed at avoiding the problem of vanishing gradients that appears in models with a high number of layers [32]. With edge

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

Table I: Computer vision algorithms implemented in real-time with NEOCAM and potential clinical applications. GMA stands for General Movements Assessment and PCK means Percent of Correct Points.

Video-based analysis	**Method**	**Measurand**	**Clinical use**
Body detection	CNN, model *MobileNetV2SSD* (lite) (see footnote 1) MParams: 4.475, GFlops: 1.525, coco-Precision: 22.2449% Output: Coords. bounding box @ 20-25 fps	Presence of infant Motor activity Height of infant	Recording classification [15] Lethargy-sepsis [11] Sleep-wake state [12] Stress, pain [13], [14] Growth trajectory GMA
Face detection	CNN, model *SqueezeNetSDD* (see footnote 2) MParams: 0.588, GFlops: 1.067, Average Precision(AP): 83% Output: Crop (300 x 300 x 3) of face @ 25 fps	Presence of infant Input for emotion classification Reference for breathing measurement Reference for heart rate monitoring	Recording classification [15] Video anonimization Breathing rate [10] Heart rate [8]
Facial expression classification	CNN, model *MobileNetV2* MParams: 2.483, GFlops: 0.126, F1: 0.84 Output: Binary classification @ 25 fps	Crying-not crying	Stress, Pain [13] Lethargy-sepsis [11]
Pose estimation	CNN, model *Blazepose* body [29] (full) MParams: 3.5, MFlops: 6.9 Average PCK @ 0.2 tolerance: 97.2% on AR dataset Output: (x, y , z) of 33 keypoints @ 20 fps and Segmented body's silhouette	Movement of limbs over time Body parts' length	GMA [19], [30], [22], [23] Lethargy-sepsis [11] Sleep-wake state [12] Stress, pain [13], [14] Recording classification [15] Growth trajectory Heart rate, breadth rate [9]
3D depth mapping	Passive Stereo-pair matching Output: z = z(t) averaged in a small region	Body volume change Pose estimation Presence of infant	Breathing rate GMA [31], [21] Recording classification

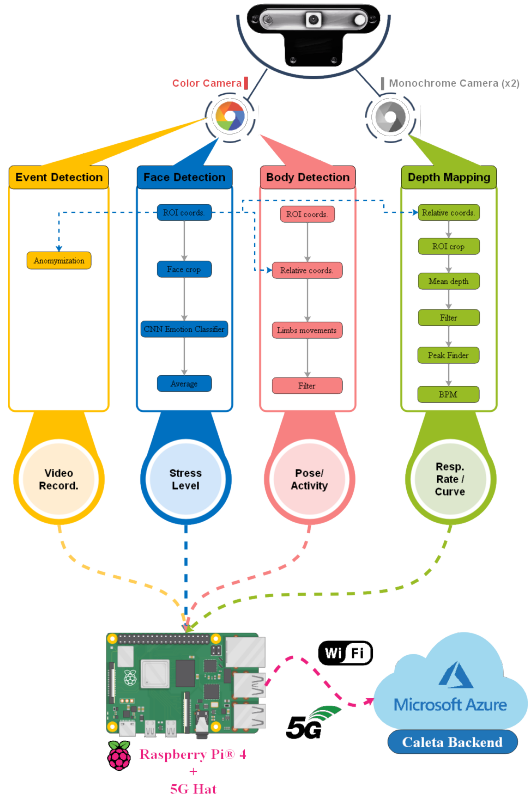

Figure 2: Pipeline of information acquisition and processing with parallel running algorithms inside NEOCAM Unit.

computing we need to strike a balance between reaching good performance while keeping a reasonable size of the models, so that they can run inside the OAK-D hardware and perform inference in real-time. The CNN-face model is a pretrained OpenVino Face Detection Retail model[2]. This network is based on the SqueezeNet architecture [33], which is a small CNN that can be compressed to less than 0.5MB. The model receives as input an RGB image with size 300×300 pixels and outputs two points, that describe the corners of a bounding box around the detected face, see Figure 3.

$$\text{CNN-face:} \qquad \left(P_{1,x}, P'_{1,x}\right) \times \left(P_{1,y}, P'_{1,y}\right)$$

The CNN-body model is based on MobileNet [34], a lightweight deep neural network, based on depthwise separable convolutions. For this work, we downloaded the pretrained weights from Luxonis[3]. CNN-body receives the same input as CNN-face, and outputs another pair of points, that describe a bounding box around the body in the image, as shown in Figure 3.

$$\text{CNN-body:} \qquad \left(P_{2,x}, P'_{2,x}\right) \times \left(P_{2,y}, P'_{2,y}\right)$$

B. Breathing rate

Breathing rate measurement is achieved with the help of the two Mono Cams of the OAK-D to detect the variations in the volume of the thoracic region of the infants while breathing.

[2]https://docs.openvino.ai/nightly/omz_models_model_face_detection_retail_0004.html

[3]https://docs.luxonis.com/projects/api/en/v2.2.1.0/samples/17_video_mobilenet/

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

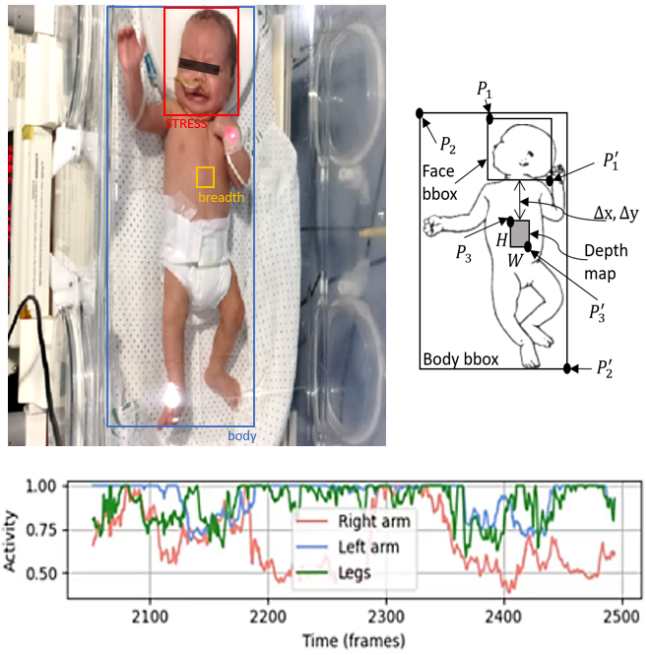

Figure 3: Top, representative video frame with bounding boxes detected in real-time marking the the face and body of a neonate. It includes an annotation with the region used for measuring the breadth whose dimensions and its relative location to the face can be preset. Below, plots showing the activity of limbs and legs as measured from the variable dimensions of the bounding box of the body.

First, CNN-face is used to detect a bounding box around the infant's head (see Figure 3). A region of interest (ROI) in the infant's chest is fixed in relation to the head, assuming constant proportions. This is the region that will be used to measure the breathing rhythm. More specifically, with the coordinates P_1 and $P_1^{'}$ of the box containing the infant's face, we define the position of a new box given by the coordinates $P_3 = (P_{3,x}, P_{3,y})$ and $P_3^{'} = (P_{3,x}^{'}, P_{3,y}^{'})$, defined by some preset (adjustable) ROI distances Δx and Δy with respect to the face box, so that:

$$\begin{aligned} P_{3,x} &= P_{1,x} + \Delta x \\ P_{3,y} &= P_{1,y} + \Delta y \end{aligned} \tag{1}$$

Since the body of a infant has a known distribution and proportions, these distances Δx and Δy can be adjusted in order to position the new box in the thoracic region. The depth mapping feature of the OAK-D camera is able to detect the depth of each pixel in an image by triangulation using the two mono cams. We compute the average depth of all the pixels in the ROI window with (adjustable) height (H) and width dimensions (W), placed on the infant's chest, to reach an average depth for each instant (frame), that we call $z(t)$. Variations on the periodic signal $z(t)$ allow us to compute the breathing rate, smoothing and normalizing the signal whenever needed. More specifically, we apply a rolling mean average to the signal $z(t)$ over a sliding window $\Delta t << 60s$, or alternatively, over N frames. We found that for an average speed of the camera of 30fps a value of $N = 10$ gave good results, i.e. we chose $\Delta t = 1/3s$. Breathing rate amounts then to finding the number of significant peaks of the signal $z(t)$ in 1 minute. For this purpose, we filtered the temporal signal combining a piecewise linear detrend and a band pass filter, and then the function `findpeaks` from the `scipy` library [35].

C. Motor activity index

The goal of the motor activity module is to detect when the infant is moving his arms or legs, and to use these movements to define an aggregated activity index. This module could also be used to alert on abnormal sequences of movements (e.g., GMA), to report lack of movement, or for measuring sleep/wake cycles.

The motion model uses the two bounding boxes detected by CNN-face and CNN-body. From the relative position of these two boxes, applying a set of rules we can infer the motion of arms and legs. To this end, we define continuous quantities $\hat{s}$ that take a value of 1 if the limb is extended and 0 if it is contracted.

Using the coordinates of the CNN-face bounding box, we can estimate the size (in pixels) of the infant's head as:

$$h_{\text{head}} = |P_{1,y} - P_{1,y}^{'}|, \tag{2}$$

and use this value to normalize all distances. Thus, we define the following stretching variables

$$\begin{aligned} \hat{s}_L^{\text{arm}} &= \frac{|P_{1,x} - P_{2,x}|}{h_{\text{head}}} \\ \hat{s}_R^{\text{arm}} &= \frac{|P_{1,x}^{'} - P_{2,x}^{'}|}{h_{\text{head}}} \\ \hat{s}^{\text{legs}} &= \frac{|P_{1,y}^{'} - P_{2,y}^{'}|}{h_{\text{head}}} - 2.5 \end{aligned} \tag{3}$$

The average length of the torso and contracted legs, relative to the head is 2.5 for most infants. Therefore, with the above definitions stretching variables range typically between 0 (contracted) and 1 (extended). We have clipped this outcome to the $(0, 1)$ interval for visualization purposes in the mobile app.

An example of this algorithm in operation can be seen Figure 3 and the associated video below[4].

[4]https://youtu.be/CGLl9O9GtEg

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

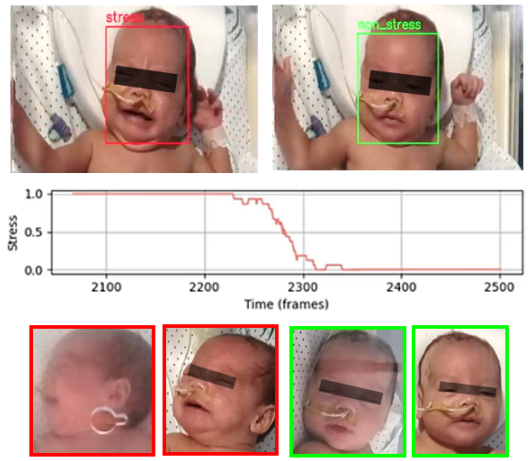

Figure 4: Example of emotion classification between two states, stressed-non stressed, using the output of the CNN-face detector. The curve shows a transition of the infant between the two states in a period of 20 seconds. Below, three representative images of images of stressed and non-stressed faces used for training the *MobilenetV2* classifier.

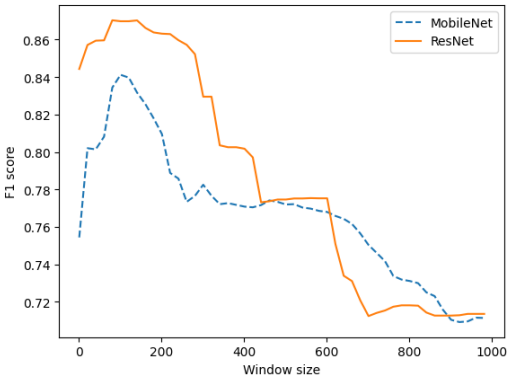

Figure 5: Smoothing study with a moving average window applied on *ResNet* and *MobileNet*.

D. Emotional status

Although they are only a few days old, newborn infants are able to express basic emotions (pain, stress, calm, etc.) with their facial expressions. This is indeed one of the most informative traits that adults use to learn about the infant's needs. Having thus a computer vision algorithm which is trained to interpret these face expressions is a key element towards monitoring stress levels and emotional status of the newborn infant.

In order to develop the emotional status model, we had to build an ML pipeline to acquire data, clean and preprocess them, annotate the data set, and finally train and validate the classifier.

The raw data acquired to train the classifier is a video footage recorded with NEOCAM at the NICU of the Puerta del Mar University Hospital, (Cádiz, Spain). More specifically, we have recorded several hours of videos from five different neonates. This footage includes different points of view, with a preference towards videos where the infant is facing upwards and the camera is placed above the incubator. We have developed a simple annotation tool that allows to label the videos by an expert on the fly. First, videos have been segmented in chunks of at most 2 minutes. The annotator can then determine when a stress period begins and when it ends by pressing different keys. Every frame in the raw video passes through the CNN-face model, that crops the face and attaches the corresponding label (stress, calm) determined by the expert. In this way we obtain a labelled dataset with about 27121 images of infant faces, 35% of which belong to the stress class. Since these images come from a video capture, it is important not to perform a random train-test split, as this would likely result in data leakage because the train and test sets would contain very similar images. To tackle this issue we have first split the videos into training and test preserving the overall class distribution, and then performed the frame extraction randomly in each set.

The classifier models are obtained by a downstream fine-tuning of the usual networks pre-trained on the ImageNet database: *ResNet50* [32] and *Mobilenet* [34]. For hyperparameter tuning we have used *Bayesian Optimization* with the aid of the library *Weights and biases* [36].

The first model trained was *ResNet50*, which gives the results shown in Table II. After 10 runs, the best set of hyperparameters obtained an F_1-score of 0.83 during the fourth epoch, using *RMSprop* as optimizer algorithm, with a learning rate close to $5e{-}5$ and a batch size of 16. Even after distilling and other tricks to reduce memory size, *ResNet50* is too large to run inside the edge-computing device. For this reason we use *Mobilenet*, a lighter model which after 20 runs achieves an F_1-score of 0.75, slightly lower than *ResNet50*. These F_1-scores are based on performance over independent frames. Persistence of the state over frames that are close in time allows to reduce noise by discarding isolated or rapidly varying changes. In order to smoothen the final output, we have performed a study to select the optimal parameters for the size of the moving average window. Figure 5 shows that the optimal value for *ResNet* is 80 frames (1.3 sec.) and 100 frames (1.6 sec) for *MobileNet*. Table II shows how *ResNet*, after smoothing, reaches a 0.87 F_1-score, but the model weighs 294 MB. On the other hand, *MobileNet* reaches 0.84 with a lighter model of size 41M.

E. Body pose estimation

Accurate body pose estimation is very relevant for the monitoring of infants, and can be applied to many different

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

Table II: Comparison of *ResNet* and *Mobilenet* results

Model	F_1-*score* (no smoothing)	F_1-*score* (smoothing)	Size [MB]
ResNet	0.836	0.87	294
MobileNet	0.751	0.84	41

clinical applications such as GMA, body presence, lethargy identification or sleep stage (see Table I).

For this study we have tested *Mediapipe Blazepose*, a lightweight pretrained CNN architecture for human pose estimation from *Google* that is tailored for real-time inference on mobile devices with the use of inter-frame regression [29]. Three versions of the model (lite, full and heavy) are available. During inference, the network produces 33 sets of the projected spatial coordinates corresponding to keypoints of the body: nose, eyes, mouse and the bilateral shoulders, elbows, wrists, fingers, hips, knees, ankles, heels and toes . Moreover, in combination with depth mapping, that is available with the NEOCAM Unit, *Blazepose* can output the keypoints in three dimensions (x, y, z). Figure 8 shows examples of pose annotation of 24 landmarks on video frames of different neonates in the real conditions of a NICU. In our tests using NEOCAM the poses were measured at 20 fps and the coordinates were output as time-series of data. The visualization in Figure 8b corresponds to the 142 poses of the infant acquired in a period of one minute in steps of 0.42 seconds.

V. EXPERIMENTAL TEST AT PUERTA DEL MAR UNIVERSITY HOSPITAL

During a period of several weeks, from March 2021 to July 2022, we conducted tests of the NEOCAM platform at the NICU of the Puerta del Mar University Hospital[5] (Cadiz, Spain). Informed consent was obtained from parents or legal guardians for all infants according to the institutional guidelines. The field tests included evaluating the working environment, checking the best position to locate the NEOCAM Units, evaluating the imaging conditions that may limit the quality of the computer vision models, and obtaining metrics to compare the measurements of our method with those obtained by standard equipment. We also tested the set up of NEOCAM Units to monitor different incubators in parallel and the functionalities provided through the mobile application. In this section we describe the results obtained in each field test, incorporating feedback from clinicians.

A. Non-contact Respiration tests

Figure 7 and the associated video[6] show some tests of the breathing rate algorithm. The two curves (raw data and after filtering) display peaks and valleys corresponding to the changes in depth of a region of the chest of a newborn while breathing. The scale in the left y-axis shows that the body was at about 80 cm from the NEOCAM camera, and the scale in the right y-axis shows that the height of the peaks due to respiration were of the order of 1 cm or less. The breathing rate is irregular, as it often happens with neonates of low gestational age (GA). Note that as the infant moves, the baseline distance from the chest to the camera changes, but this has no effect in the breathing rate computation, as the peaks are correctly identified. It could happen, however, that a larger motion of the infant might cause the camera to stop measuring if the ROI window falls outside the thoracic region. In order to make the system more robust to varying lighting levels and to children with different body poses and

[5]https://hospitalpuertadelmar.com/
[6]https://youtu.be/ZsHf2NaaHW8

Table III: Population used for this study. PMA stands for post menstrual age

Number of newborns (% male)	5 (80)
Mean PMA (std)	32.2 (2.6)
Mean height (std) [cms]	42.6 (4.5)
Mean weight (std) [grs]	1760 (508)
Mean total recording (std) [hrs]	3(2)
Continuous recording duration [hrs]	0.1-0.5

Figure 6: Above, network of two NEOCAM Units at NICU hold with an articulated arm to a mast usually used for intravenous infusions of fluids. Below, detail of the casing located above the incubator and touchscreen displaying video annotated with bounding boxes and data plots.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

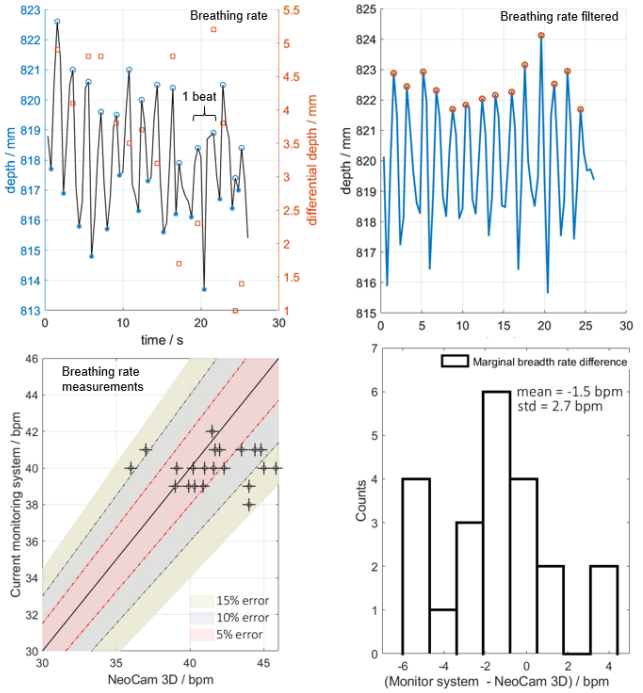

Figure 7: Above, example of one representative breathing rate signal (before and after filtering) measured using stereo depth mapping. The right y-axis indicates differential depth changes (distance between peak and trough) of one beat. Below, difference between NEOCAM measurements and those obtained with a conventional monitor system.

proportions, the algorithm implemented in NEOCAM allows changing with a keyboard the dimensions (W and H) and the distances (Δx, Δy) to the face of the ROI window (see Eqs. 1 and Figure 3). Adjusting these parameters, we were able during the tests to detect and measure breathing rate in about 83% of the cases. We compared the breathing rate measured using NEOCAM Unit with the values measured with the standard breathing sensor used at the NICU. The test was conducted on two neonates (girl and boy) during approximately one hour with the custom parameters of the method kept constant throughout this time. Figure 7 shows a histogram of the marginal error, in beats per minutes, bpm, of the measurements captured by both systems, with a mean/standard deviation (std) of -1.5/2.7 bpm. The figure also shows the analysis with error band of the measurements obtained using of both methods. The mean of the relative error of the NEOCAM Unit is 5.9% (std = 4.9%). Every measurement corresponds to a period of 30 seconds to estimate the breathing rate.

B. AI algorithms evaluation

We have tested the human pose detector *Mediapipe Blazepose* that has been conceived for real-time detection in edge devices. Figure 8 shows a selection of frames of videos processed successfully at 20 fps from 4 neonates in incubators. The overlay in color corresponds to the skeletal body poses measured. We discarded the landmarks associated to the eyes and nose. The images show that the pose detector is very robust when imaging newborns of different sizes and with different postures, with images acquired with strong perspectives, and under strongly varying lighting conditions (artificial, natural, dim, bright...). For example, the pose of the infant in Figure 8e is detected correctly even though he is imaged horizontally, through the top surface of an incubator that was blurry due to surface wear and scratches, and with a hole for cable insertion in the middle of the field of vision. The infant with low gestational age in Figure 8g sleeps face down and arms folded but the pose is also detected with only a small error on the right leg.

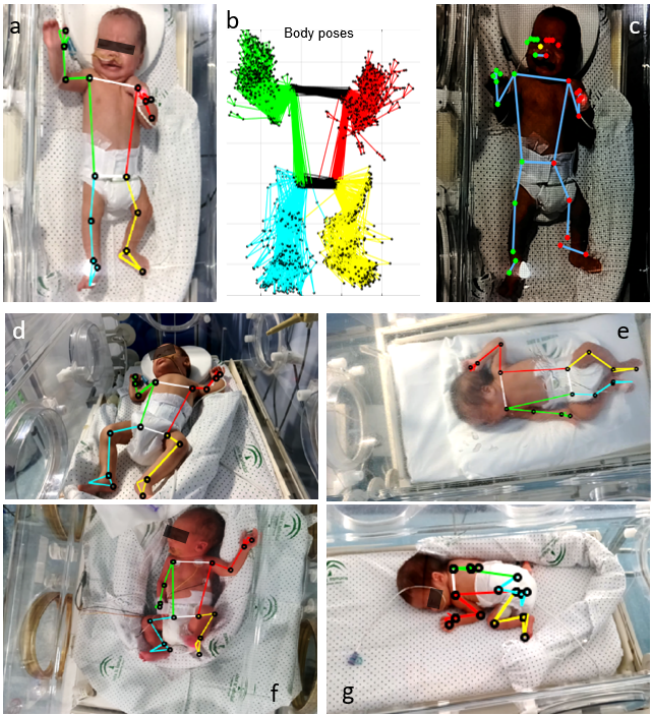

Figure 8: Selection of video frames showing the robustness of the human pose estimation at speeds of $\approx$ 20 fps, for infants with different postures, camera locations and lighting conditions.

As explained in Section IV, we devised a novel algorithm that uses the bounding boxes of two object detectors to provide a fast and robust estimation of the motor activity of preterm infants (see Figure 3). This has the advantage

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

that it can be used for other applications (see Table I), such as detecting the presence of the infant or measuring his body length. We have tested also the image classifier for emotional status monitoring (stressed or calm). The robustness of the algorithm relies on the correct detection of a face in the image. The algorithms of face detection and body detection were functioning correctly in 90% cases for normal conditions (infants with frontal pose, not covered with blankets and under a normal lighting level). The detection was erratic, when the infants were facing down or were covered by blankets, when the light level was too low or too high (e.g., light reflections on the surfaces of the incubator).

C. Clinician Feedback

NEOCAM was shown and explained to the medical staff in the NICU. A survey was carried out to get the opinion of doctors and nurses working in the unit of the Hospital. The survey was answered by 8 nurses and 6 neonatologists and the replies are found in Figure 9. Nine of the respondents had more than 5 years of experience, 30% of them work 4 or 5 days per week, and 50% more than 7 hours per day at the NICU. The survey contained questions about the ease of installation and use of the camera and mobile application, integration of the NEOCAM system within their normal tasks, and the usefulness of the system. Allowed replies were *Agree, Disagree, Do not Know/No Answer (DK/NA)*. In addition, there was free space for respondents to write suggestions.

From the responses to the survey, we deduce that neonatologists feel safe with the system and the majority consider that its installation is totally safe for the newborn. In contrast, most nurses cast some doubts about the convenience of the system. A large number (60%) of clinicians did not give an answer about the easiness of use and installation, because they did not have the chance to actually use and install the system, and their opinion is based only on the information provided with videos. Overall, most clinicians wanted to have more time to test the system, and considered that the system is appropriate. Among the negative answers, most concerns or worries were related to privacy and control issues.

Summarizing, from the tests we can extract the following considerations that are key for improving proper imaging and for understanding the requirements for any video-based monitoring system under the real changing situations of a NICU:

- Infants with different GAs or PMAs may have very different body sizes.
- Infants in the incubators are usually facing upwards or downwards.
- They have often nasal probes and different sensors attached to the skin and cables.
- The most common type of incubator in our study is a closed transparent box made of PVC whose surface might show signs of wear and scratches.
- Infants are often placed in open beds, generally when they are about to be discharged.
- Sometimes the infants are partially occluded by blankets, pillows or bed sheets.
- Lighting conditions can vary strongly, including artificial and natural light sources from ceiling lights and windows. Sometimes light shines on the surface of the incubator or the incubator can be covered with a blanket.

Regarding the comparison of NEOCAM measurements with those obtained by standard medical equipment (Dräger Infinity Delta), we observe that

- There are no current medical devices that measure emotional status and motor activity, so our methods cannot be compared with standard practice.
- Breathing rate measured by NEOCAM is within 6% of the standard measurements obtained by probes attached to the skin, provided there are no impediments to proper imaging.

VI. SUMMARY AND OUTLOOK

We have implemented and tested a system of video cameras and related software in NICUs, which is designed to operate with minimal impact on normal healthcare staff operations, and very simple installation requirements. Most of its features draw their advantages from edge computing (running AI models directly on the device), which allow for more efficient data usage, enhanced security and multiple applications running in parallel. More specifically, we discuss some of these novel features below:

- Existing solutions for video-based monitoring are designed as single units, able to monitor one bed or incubator. NEOCAM displays cameras in a network architecture so that several units can run in many incubators efficiently at the same time.
- Table I shows recent research in different clinical applications of video monitoring of preterms, and that focus on specific applications. In contrast, since the NEOCAM unit is able to run several AI models in parallel in real-time, it can cover several applications in the same device. In this work, 3 CNNs and depth estimation models were run in parallel at 20-30 fps, and a SOTA pose detector can run robustly at 20 fps.
- Each NEOCAM Unit can transmit only high-level information, thus minimizing the bandwidth required to transmit and process video streams in a network of

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

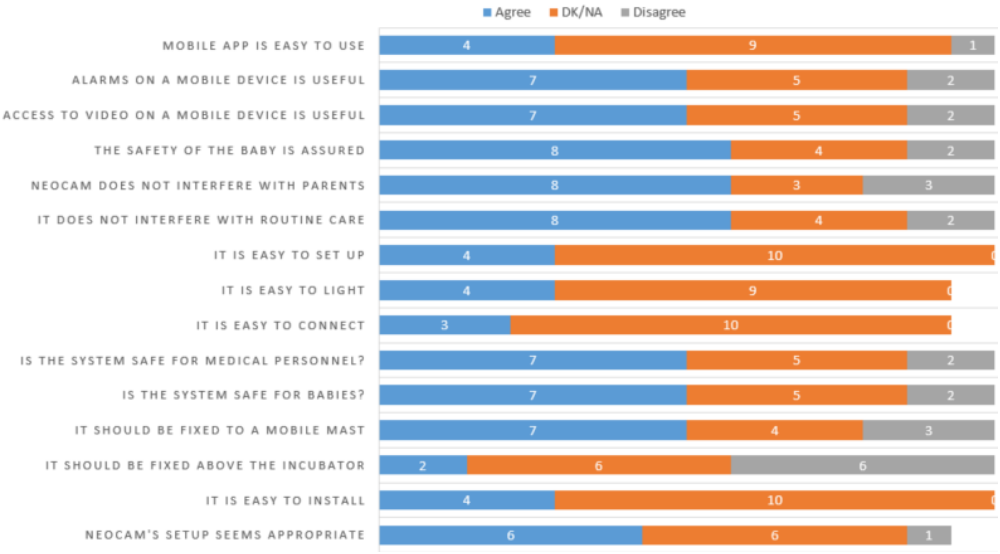

Figure 9: Survey on NEOCAM among medical staff of Puerta del Mar University Hospital (Cádiz, Spain). DK/NA stands for Do not Know/No Answer.

cameras. It means a much more efficient use of data, and by avoiding the sending of images it can help to comply with ethical and privacy issues.

- Each NEOCAM unit can perform real-time anonymization of the videos for example by blurring or pixelating the bounding box of the detected face (see Figure 3). This is an important security feature that can be used before transmitting videos outside the NEOCAM unit.
- NEOCAM is a network of edge-computing devices that can communicate with third-party devices/software through a hybrid (WiFi/5G) access network.
- The streaming service as well as the notification system supported by NEOCAM may not only provide real-time monitoring for healthcare professionals, but also enable relatives to keep a closer contact with their infants.

Compared to the current SOTA in this field, which is probably *Voxyvi* [28], the advantages of NEOCAM come mostly from its hardware/software architecture with edge computing. *Voxyvi* is a recording device that can detects the presence/non-presence of the infant in the field-of-view. It has still some advantages in its use of infrared cameras and sound, and it has been tested in a larger number of hospitals. A summary of this comparison can be found in Table IV.

NEOCAM can be improved in several aspects. The first one is to ensure robustness under difficult imaging conditions. Incubators in NICUs may often be under dim lighting or even in the dark (and infants often covered with blankets) to help them rest and keep them warm. In such situations, video-analysis, for example for GMA with the pose or the motor activity estimator, can fail if normal daylight cameras are used. NEOCAM may be improved, as in [28], adding thermal or near-infrared (NIR) sensitive cameras that can be used with daylight or in dark conditions with the help of IR illumination. Thermal cameras provide relevant clinical information (detailed body temperature maps) but they must be located inside the incubator because the PVC used to fabricate the wall of the incubators is no longer transparent to the mid- or long IR wavelengths.

The AI models also admit improvement, which would essentially come from training over larger datasets. Emotion classification has been fine tuned with transfer learning and optimised with an optimal smoothing. A better generalization would be obtained by training with faces in any orientation and with hybrid algorithms that combine image (and facial landmarks) with sound. The simple pose detector and motor activity estimation can also be improved by fine tuning the face and body detectors on neonates, whose body proportions differ from children or adults. Moreover, infants with low GA are often curled up with arms and legs drawn, and may have large diapers, blankets and sensors attached, all of which are a potential source of inaccuracies.

Measuring breathing rate is especially difficult on preterms with low PMA because 3D movements are imperceptible with stereo depth mapping when the infant is poorly illuminated. Stereo depth mapping depends on feature matching, which is hard under a low light environment as features are not clearly visible. Low lighting may be solved using active illumination or structured light (e.g., a dot projector) as in [10]. Even better, for infants with low PMA, the high-accuracy of the Time-of-Flight (ToF) approach will

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

Table IV: NEOCAM versus SOTA

	Specification	***Voxyvi*** [28]	***NeoCam***
Technology	Connectivity	Hardwired	Wireless
	Edge computing	No	Yes
	AI inference	Yes, 1 model offline	Yes, 3 models in parallel, in real-time (at ≈ 30 fps)
	AI models tested	Video classification	Body detection, Face detection-classification / Pose estimation
	Multi-modal	RGB / IR / Thermal	RGB / BW left / BW right
	3D	No	Yes, stereo matching
	Sound	Yes	No
	IoT Net	No	Yes
	Computing	Local PC	Edge / Cloud
Install	Mount	Fixed / Ballasted mast	Ballasted mast
	Size	Compact	Compact
	Location	Inside incubator	Outside incubator
	Wiring	Power/Ethernet	Power
Data	Access	Local PC	Local touchscreen and Mobile App
	Processing	Offline	Real-time
	Transfer	Heavy-Video	Heavy-Video / Light-measurements
	Storage	Heavy	Video and/or only measurements
	Streams	Raw videos / Raw sound	Raw video / High-level data
	Measurands	Baby presence (offline)	Baby presence, Breathing rate, Body poses, Facial stress (online)
Validation	technical	6 hospitals (≈ 96 newborns in incubators)	1 hospital (5 newborns in incubators)
Ethical	Anonymization / Privacy	Only after the data is transferred to the server	Video with face blurring in real-time Only high-level relevant data can be transferred

probably be the best solution. Another potential improvement to better cope with changes in baby's posture and movements is to dynamically find and adjust the optimal window location defined in our algorithm.

While several tests in real operating conditions have been performed, the NEOCAM platform still needs further validation. Close collaboration with medical staff is essential to understand the real needs and concerns for an innovative technology that could be regarded in conflict with privacy, and for further development of clinical applications. Nevertheless, the ability to process sensitive image data directly on the device through edge computing allows compliance with personal data protection regulations and reduces the risk of data breaches from cloud servers.

The main purpose of preparing a detailed description of the system (see the Supplementary Material) and rendering all our code accessible for other groups is to enable further research and testing based on the current platform. As an open-source solution, developers and researchers could reassemble the NEOCAM Unit with additional or alternate devices to support another features, re-use some back-end services as part of other functionalities, re-use the notification and streaming service to support other features and use the trained AI models in other applications or systems. In this respect, NEOCAM architecture may be of interest in other medical care applications, such as monitoring neonates in cradles, adult patients in hospital' beds, or elderly people in residences.

ACKNOWLEDGMENT

LCG, IBF and SPLL ackowledge support from EU program H2020-MSCA-ITN-2020 through project PARENT (GA Nº: 956394) and from Fundacion Progreso y Salud (FPS-SSPA) through EU-FEDER project ITI-0019-2019-INIBICA. LCG wants to acknowledge funding from the Spanish MCIU/AEI under grant PGC2018-101538-A-I00. The research of DGU is supported in part by the spanish AEI under grants PID2021-122154NB-I00 and TED2021-129455B-I00, and by a 2021 BBVA Foundation project for research in Mathematics. He also acknowledges support from the EU under the 2014-2020 ERDF Operational Programme and the Department of Economy, Knowledge, Business and University of the Regional Government of Andalusia (FEDER-UCA18-108393). All the authors thank to the babies and their parents for their help.

Finally, all authors would like to thank OpenCV for awarding this project in the OpenCV AI Competition 2021 [7] with the 2nd place in the *Global Grand Prize* and 1st place in the *Regional Prize* category.

ETHICS APPROVAL

The study was approved by the Neonatology unit of the Puerta del Mar University Hospital[8] (Cadiz, Spain) and after obtaining the informed consent of the parents. All research data are de-identified and securely stored. Data access is limited to approved study personnel.

SUPPLEMENTARY MATERIAL

A Technical Report on NEOCAM is available as supplementary material, including a detailed description of the architecture, specifications for the hardware components of the NEOCAM unit, a description of the functionalities of the mobile app, recordings form the platform's performance and links to the software repositories.

[7]https://opencv.org/opencv-ai-competition-2021/
[8]https://hospitalpuertadelmar.com/

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3240245

REFERENCES

[1] C. V. Hofsten, "An action perspective on motor development," *Trends in cognitive sciences*, vol. 8, no. 6, pp. 266–272, 2004.

[2] B. Kolb *et al.*, "Principles of plasticity in the developing brain," *Developmental Medicine & Child Neurology*, vol. 59, no. 12, pp. 1218–1223, 2017.

[3] M. G. others, "Neonatal stress reactivity: Predictions to later emotional temperament," *Child development*, vol. 66, no. 1, pp. 1–13, 1995.

[4] I. Novak *et al.*, "Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment," *JAMA pediatrics*, vol. 171, no. 9, pp. 897–907, 2017.

[5] L. C. Gontard *et al.*, "Automatic segmentation of ventricular volume by 3d ultrasonography in post haemorrhagic ventricular dilatation among preterm infants," *Scientific Reports*, vol. 11, no. 1, p. 567, Jan 2021. [Online]. Available: https://doi.org/10.1038/s41598-020-80783-3

[6] O. Bonner *et al.*, "'there were more wires than him': the potential for wireless patient monitoring in neonatal intensive care," *BMJ innovations*, vol. 3, no. 1, 2017.

[7] N. Taufik *et al.*, "Development of infant telemonitoring system," *Evolution in Electrical and Electronic Engineering*, vol. 1, no. 1, pp. 153–160, 2020.

[8] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE journal of biomedical and health informatics*, vol. 18, no. 4, pp. 1153–1160, 2013.

[9] M. Villaroel *et al.*, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," vol. 1, no. 3, pp. 87–91, 2014.

[10] V. Ottaviane *et al.*, "Contactless monitoring of breathing pattern and thoracoabdominal asynchronies in preterm infants using depth cameras: A feasibility study," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–8, 2022.

[11] G. Zamzmi *et al.*, "A comprehensive and context-sensitive neonatal pain assessment using computer vision," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 28–45, 2022.

[12] L. Cailleau *et al.*, "Quiet sleep organization of very preterm infants is correlated with postnatal maturation," *Frontiers in Pediatrics*, vol. 8, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fped.2020.559658

[13] G. Zamzmi *et al.*, "Pain assessment from facial expression: Neonatal convolutional neural network (n-cnn)," 2019, pp. 1–7.

[14] S. Lysenko *et al.*, "Towards automated emotion classification of atypically and typically developing infants," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2020, pp. 503–508.

[15] R. Weber *et al.*, "Preterm newborn presence detection in incubator and open bed using deep transfer learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1419–1428, 2021.

[16] N. Silva *et al.*, "The future of General Movement Assessment: The role of computer vision and machine learning – A scoping review," *Research in Developmental Disabilities*, vol. 110, no. November 2020, 2021.

[17] K. Raghuram *et al.*, "Automated movement recognition to predict motor impairment in high-risk infants: a systematic review of diagnostic test accuracy and meta-analysis," *Developmental Medicine & Child Neurology*, vol. 63, no. 6, pp. 637–648, 5 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/dmcn.14800http://files/183/Raghurametal.-2021-Automatedmovementrecognitiontopredictmotorim.pdf

[18] H. Kanazawa, Y. Yamada, K. Tanaka, M. Kawai, F. Niwa, K. Iwanaga, and Y. Kuniyoshi, "Open-ended movements structure sensorimotor information in early human development," *Proceedings of the National Academy of Sciences*, vol. 120, no. 1, p. e2209953120, 2023.

[19] T. Tsuji *et al.*, "Markerless measurement and evaluation of general movements in infants," *Scientific reports*, vol. 10, no. 1, pp. 1–13, 2020.

[20] C. Tacchino *et al.*, "Spontaneous movements in the newborns: a tool of quantitative video analysis of preterm babies," *Computer methods and programs in biomedicine*, vol. 199, p. 105838, February 2021.

[21] S. Moccia, L. Migliorelli, V. Carnielli, and E. Frontoni, "Preterm infants' pose estimation with spatio-temporal features," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 8, pp. 2370–2380, 2020.

[22] A. S. Schroeder, N. Hesse, R. Weinberger, U. Tacke, L. Gerstl, A. Hilgendorff, F. Heinen, M. Arens, L. J. Dijkstra, S. P. Rocamora *et al.*, "General movement assessment from videos of computed 3d infant body models is equally effective compared to conventional rgb video rating," *Early Human Development*, vol. 144, p. 104967, 2020.

[23] H. I. Shin, H.-I. Shin, M. S. Bang, D.-K. Kim, S. H. Shin, E.-K. Kim, Y.-J. Kim, E. S. Lee, S. G. Park, H. M. Ji *et al.*, "Deep learning-based quantitative analyses of spontaneous movements and their association with early neurological development in preterm infants," *Scientific Reports*, vol. 12, no. 1, pp. 1–9, 2022.

[24] A. AlZubaidi *et al.*, "Review of biomedical applications of contactless imaging of neonates using infrared thermography and beyond," *Methods and Protocols*, vol. 1, no. 4, 2018. [Online]. Available: https://www.mdpi.com/2409-9279/1/4/39

[25] S. Cabon *et al.*, "Audio- and video-based estimation of the sleep stages of newborns in neonatal intensive care unit," *Biomedical Signal Processing and Control*, vol. 52, pp. 362–370, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809419301089

[26] S. Orlandi *et al.*, "Avim—a contactless system for infant data acquisition and analysis: Software architecture and first results," *Biomedical Signal Processing and Control*, vol. 20, pp. 85–99, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809415000749

[27] W. Baccinelli, M. Bulgheroni, V. Simonetti, F. Fulceri, A. Caruso, L. Gila, and M. L. Scattoni, "Movidea: A software package for automatic video analysis of movements in infants at risk for neurodevelopmental disorders," *Brain sciences*, vol. 10, no. 4, p. 203, 2020.

[28] S. Cabon, F. Porée, G. Cuffel, O. Rosec, F. Geslin, P. Pladys, A. Simon, and G. Carrault, "Voxyvi: A system for long-term audio and video acquisitions in neonatal intensive care units," *Early Human Development*, vol. 153, p. 105303, 2021.

[29] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking. arxiv," *arXiv preprint arXiv:2006.10204*, 2020.

[30] C. Tacchino *et al.*, "Spontaneous movements in the newborns: a tool of quantitative video analysis of preterm babies," *Computer methods and programs in biomedicine*, vol. 199, 2 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33421664/

[31] K. D. McCay, E. S. L. Ho, H. P. H. Shum, G. Fehringer, C. Marcroft, and N. D. Embleton, "Abnormal infant movements classification with deep learning on pose-based features," *IEEE Access*, vol. 8, pp. 51 582–51 592, 2020.

[32] K. He *et al.*, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 770–778, 2016.

[33] F. Iandola *et al.*, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv preprint arXiv:1602.07360*, pp. 1–13, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[34] A. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, no. April 2017, 4 2017. [Online]. Available: https://arxiv.org/abs/1704.04861v1http://arxiv.org/abs/1704.04861

[35] P. Virtanen *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[36] L. Biewald, "Experiment tracking with weights and biases," 2020, software available from wandb.com. [Online]. Available: https://www.wandb.com/

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.