Preterm Newborn Presence Detection in Incubator and Open Bed Using Deep Transfer Learning

Raphael Weber, Sandie Cabon, Antoine Simon, Fabienne Poree, Guy Carrault

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

Preterm newborn presence detection in incubator and open bed using deep transfer learning

Raphaël Weber, Sandie Cabon, Antoine Simon, Fabienne Porée, and Guy Carrault

***Abstract*— Video-based motion analysis recently appeared to be a promising approach in neonatal intensive care units for monitoring the state of preterm newborns since it is contact-less and noninvasive. However it is important to remove periods when the newborn is absent or an adult is present from the analysis. In this paper, we propose a method for automatic detection of preterm newborn presence in incubator and open bed. We learn a specific model for each bed type as the camera placement differs a lot and the encountered situations are different between both. We break the problem down into two binary classifications based on deep transfer learning that are fused afterwards: newborn presence detection on the one hand and adult presence detection on the other hand. Moreover, we adopt a strategy of decision intervals fusion in order to take advantage of temporal consistency. We test three deep neural network that were pre-trained on ImageNet: VGG16, MobileNetV2 and InceptionV3. Two classifiers are compared: support vector machine and a small neural network. Our experiments are conducted on a database of 120 newborns. The whole method is evaluated on a subset of 25 newborns including 66 days of video recordings. In incubator, we reach a balanced accuracy of 86%. In open bed, the performance is lower because of a much wider variety of situations whereas less data are available.**

***Index Terms*— Deep transfer learning, neonatal intensive care units, preterm newborn, video monitoring**

I. Introduction

NEWBORNS who are born before a gestational age of 37 weeks are considered premature and have not fully developed all of their vital functions. In order to ensure their optimal development, preterm newborns are hospitalized in neonatal intensive care units (NICUs), where they generally start their life in an incubator before gradually being settled in an open bed and require particular attention from medical staff. During their first days of hospitalization, extremely and very preterm newborns are in an incubator, which is a closed bed reproducing the in utero environment. They are then transferred in an open bed when they are mature enough. Regarding moderate and late preterm newborns, they are directly in an open bed after their birth. The open bed might be a cradle or equipped with a warm radiant. The monitoring of physiological signals allows to give information to the medical staff about the newborn state and the progress of maturation. Thus it helps them in the diagnosis of a disease and supports the treatment decision. Cardiac and respiratory signals are always monitored to detect, for instance, bradycardia and apnea. Brain activity may also be monitored for detecting neurological disorders such as neonatal seizures [1]. The drawback of monitoring such signals is the invasiveness. Indeed, it requires to stick adhesive electrodes or transducers on the very fragile skin of the newborn, which may damage it and increase the risk of infection.

Several technologies have been proposed in order to monitor cardiac activity and/or respiratory signal in a non-invasive manner (see [2], [3] for reviews), such as the use of capacitive sensors [4], piezoelectric sensors [5], ultrasonic systems [6], radar systems [7], thermal imaging [8] or cameras [9].

Moreover, video modality can be used for high-level analysis and proved to be relevant in several clinical applications in pediatrics, particularly with motion analysis [10]. Regarding clinical applications involving preterm newborns, video-based motion analysis has been investigated for instance for early cerebral palsy detection [11], estimation of sleep stages [12], or maturation characterization [13]. The latter is motivated by the fact that the motor activity evolves along with the age of the newborns [14]. Recently, analysis of motion estimated from ECG has been investigated for the detection of late-onset sepsis [15]. Indeed, it has been shown that sepsis is related to lethargy, which can be observed as the absence of motion [16]. So it could be interesting to tackle this problem with video-based motion analysis.

Nevertheless, video-based motion analysis faces several challenges due to the real-life acquisition conditions in NICUs, particularly in the context of long-term monitoring. As it has been already highlighted in [9], [10], before analyzing newborn motion, it is crucial to detect the periods when the newborn is absent and when an adult is present in the image frame. Indeed, in NICU, the newborn is regularly taken out of the bed in order to change the bedding or to have skin-to-skin contact with the parents. These periods must be discarded from the analysis because otherwise it may be wrongly interpreted as an absence of motion. Moreover, nurses regularly handle the newborn inside the bed, which generates irrelevant motion patterns in the image, so they must be discarded as well. In

Paper submission: 29th of May 2020. Major revision submission: 9th of October 2020. Results incorporated in this publication received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689260 (Digi-NewB project).

All authors are with Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France. Email: {firstname}.{name}@univ-rennes1.fr

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

the perspective of a fully automatic video-based monitoring system, it is essential to automatically detect the periods of interest, i.e. when the newborn is solely present in the image. Only a few papers have proposed to address this problematic in NICUs. In [13] only adult presence detection in open bed is considered, whereas in [9] both newborn absence and adult presence are considered in incubator.

In this paper, we propose a video-based method for automatic detection of newborn presence in incubator or open bed. The goal is to label images of a video stream with three classes: "adult & newborn present," "newborn absent" and "newborn solely present." Our method is based on deep transfer learning. We use two binary classifiers: on the one hand, we detect the presence of the newborn, on the other hand, we detect the presence of adults. We add a strategy of decision intervals fusion in order to take advantage of temporal consistency. Finally, we fuse the decision of the two classifiers in order to get one of the three classes. One originality of this work is to address the problem of newborn presence detection both in incubator and open bed, thus the whole stay of the newborn in NICU can be analyzed with a common system.

This work is part of the Digi-NewB project, whose aim is to develop a monitoring system of preterm newborns, with a focus on late-onset sepsis detection and maturation characterization. It is the extension of a previously published work [17] where we used the same workflow, but instead of learning two binary classifiers, we trained one single 3-class classifier. Moreover, in [17] the method was only tested in incubator.

The remaining of this paper is organized as follows. After having introduced the related work, we describe our method for automatic detection of preterm newborn presence. Then, we describe the database00 that we used in our experiments. Next, we report the results of our experiments. Then, we discuss the limitations of our method. Lastly, we conclude the paper and give some perspectives of this work.

II. Related work

In the context of the Digi-NewB project, our team aims at developing a new non-invasive monitoring system using video analysis. Previous video-based works were mainly related to motion analysis using frame differencing [18], [19] or optical flow [20], [21], followed by extraction of features to characterize motor activity [10]. However, a major limitation to deploy these methods as a continuous monitoring technique is the presence of adults in the field of view of the camera or the absence of the newborn from the bed.

To overcome this difficulty, we proposed a first ad-hoc method to deal with adult presence detection in open bed [13]. The arrival and the departure of the adult in the image frame are detected by analyzing the change of motion in the image edge. A manual initialization is needed as the algorithm must start at a frame where the newborn is solely present in the image. The method has been tested on a database of nine newborns with a total of 149 hours of recording.

A similar approach has been proposed in [22] for infant presence detection in cradle. It has not been applied to recordings in NICUs but at home. Motion is estimated both inside and outside a region of interest surrounding the cradle. A set of temporal rules based on both motion estimates are designed in order to decide whether the infant is in or out of the bed. The method has been tested on a database of five infants with 77 recordings of 24 hours each.

The drawbacks of these two methods are related to the acquisition context in NICUs. If the camera is not optimally placed, the newborn may generate motion in the image edge, that would lead to a false adult detection. Noisy motion can also be generated in the image edge by abrupt changes of lighting. Moreover, the detection of adult presence may be missed when the adult is present but is not moving for a while. The same phenomenon may happen in the case of infant detection when he or she is not moving. These methods are not viable in the context of long-term monitoring in NICUs because the approach has to be robust to camera placement and lighting variations without any manual intervention.

In this context, image classification based on deep learning appears as a way to overcome these limitations. In particular, deep transfer learning is a powerful approach when the database has a lot of variety but the number of images is not large enough to train a neural network from scratch. This allows to benefit from the generalization capacities of a deep neural network. Transfer learning methods aim at transferring the knowledge of a source task to a target task. In the case of deep learning, a neural network is pre-trained for the source task. Then, the output layer performing classification is replaced by another classifier that is trained for the target task. The training dataset of the source task has usually a huge number of samples, which leads to a model with a good ability to generalize.

This technique has been used for a wide variety of target tasks [9], [23]–[26], where the source task is usually ImageNet classification [27]. The implementation choices lie mostly in the deep neural network and the output classifier, and are motivated by the target task. The most commonly used deep neural networks are Alexnet [28] and VGG16 [29]. The output classifier can be the output layer of the deep neural network [24], another neural network with typically one or two layers [9], [23], [25], or classical machine learning algorithms such as support vector machine, logistic regression or random forests [25], [26].

In particular, deep transfer learning with VGG16 [29] is used in [9] for newborn presence detection. Regarding adult presence detection, a two-stream network with ResNet50 [30] is trained. It has been tested in incubator on a database of 15 newborns with a total of 214 hours of video recordings.

III. Method

In this section, we introduce our method of automatic video-based detection of newborn presence. First, we give an overview. Then, we detail each step of the method.

A. Overview

Fig. 1 gives an overview of the method. The idea is to break the problem down into two binary classifications that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

TABLE I
DEFINITION OF THE CLASSES

	Newborn absent	Newborn present
Adult absent	"newborn absent"	"newborn solely present"
Adult present		"adult & newborn present"

are fused afterwards. On the one hand, there is the newborn presence detection regardless of the adult presence, referred to as "newborn" classification. On the other hand, there is the adult presence detection, referred to as "adult" classification.

Binary classification is based on deep transfer learning. The images of the video stream pass through a pre-trained deep neural network in order to extract bottleneck features. These features are then used as the input of the binary classifier, which gives a sequence of decisions. In order to remove short peaks of detection that are assumed to be irrelevant, we adopt a strategy of decision intervals fusion.

Finally, the binary decisions are fused in order to get one of the following classes, as illustrated in Table I:

- "adult & newborn present": both newborn and adult are present,
- "newborn absent": newborn is absent, regardless of the adult presence,
- "newborn solely present": newborn is present and adult is absent.

We choose to define only one class when the newborn is absent because there are only a few images with newborn absent and adult present compared to the other classes.

B. Bottleneck features extraction

In the context of deep transfer learning, bottleneck features correspond to the knowledge of the source task. They are extracted from a deep neural network that has already been trained for a specific classification problem.

Fig. 2 illustrates bottleneck features extraction. First, the image must be pre-processed according to the input layer of the pre-trained network. Then it passes through the network but stops before the output layer, which is usually a fully-connected layer performing the classification. Thus, we obtain the most high-level features of the pre-trained network and they are ready to be used for another classification problem.

In this paper, the selected source task is ImageNet [27] as it proved to be powerful in deep transfer learning. We test three different deep neural networks:

- VGG16 [29],
- MobileNetV2 [31],
- InceptionV3 [32].

The pre-processing step consists in resizing the image to the shape 250 pixels x 250 pixels. MobileNetV2 requires a specific shape, so in this case it is 224 pixels x 224 pixels.

C. Classification

Two binary classifications are performed: "newborn" classification and "adult" classification (see Subsection III-A). The binary classifiers are trained for the target task and take as input the bottleneck features extracted from a pre-trained deep neural network.

In this paper, two kinds of classifiers are tested: support vector machine (SVM) with a linear kernel and small neural network (NN). The small neural network consists of one fully-connected layer with the rectified linear unit as the activation function, followed by another fully-connected layer with softmax as the activation function. Droupout regularization is applied before this layer in order to prevent overfitting.

D. Decision intervals fusion

A binary classifier returns a sequence of decisions for a video. Let i be a class. We call "interval of class i" a subsequence where all the decisions are equal to class i. If the duration of the temporal interval that separates two successive intervals of class i is less than a threshold δ_i, then the latter are fused by changing the decisions of the separating interval to class i. This process is done sequentially with a different threshold for each class.

Fig. 3 illustrates the step of decision intervals fusion for the positive class. There are three intervals for the positive class: $[t_1, t_2]$, $[t_3, t_4]$ and $[t_5, t_6]$. So, the separating intervals are $[t_2, t_3]$ and $[t_4, t_5]$. Since $t_3 - t_2 > \delta_1$ and $t_5 - t_4 \leq \delta_1$, only the interval $[t_4, t_5]$ is changed to class 1.

E. Binary fusion

Let d_n be the decision for "newborn" classification and d_a be the decision for "adult" classification. The final decision d_f is computed as follows:

$$d_f = \begin{cases} \text{"adult \& newborn present,"} & \text{if } d_n = 1 \ \& \ d_a = 1, \\ \text{"newborn absent,"} & \text{if } d_n = 0, \\ \text{"newborn solely present,"} & \text{if } d_n = 1 \ \& \ d_a = 0. \end{cases} \quad (1)$$

For an objective evaluation, the proposed approach was compared to two other methods [13], [17] in Section V.

IV. DATABASE DESCRIPTION

This work is part of the Digi-NewB project in which a large database of video recordings has been acquired. For our experiments, we created two subsets from this database: the database of still images and the database of videos. The former will be used to train the binary classifiers and the latter to set the parameters of decision intervals fusion. Both of them are split with respect to the bed type: incubator and open bed.

In this section, we introduce these two databases. In the first subsection we focus on the data acquisition setup. In the following subsections we describe the content of the database of still images and then of the database of videos.

This study received ethics approval from the Ouest IV Ethics Committee (reference number 34/16) and one parent of each newborn gave its signed agreement to take part in it.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

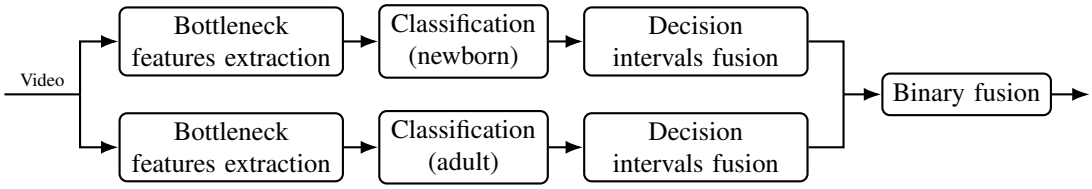


Fig. 1. Overview of our method for automatic video-based detection of newborn presence

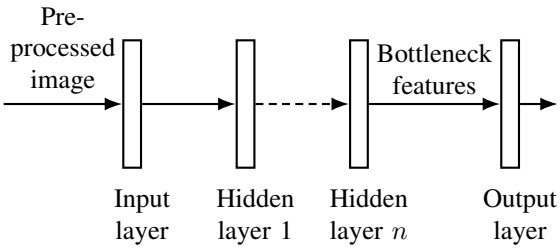


Fig. 2. Illustration of bottleneck features extraction

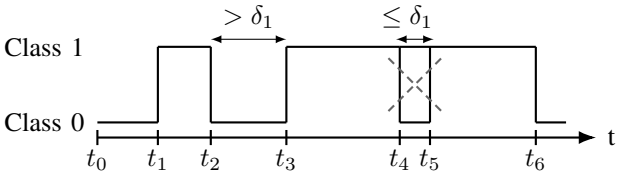


Fig. 3. Illustration of decision intervals fusion for the positive class

(a) Incubator

(b) Open bed

Fig. 4. Example images of the class "adult & newborn present"

A. Data acquisition setup

In the scope of the Digi-NewB project, a dedicated audio and video acquisition device was designed [33]. In particular, it embedded two black & white infrared cameras with a resolution of 752x480 pixels (FMVU-03MTMCS), each one associated with an infrared illumination by four LEDs (VSMY3850 from Vishay). LEDs operate in the near infrared at 850 nm and have a spectral bandwidth of 30 nm. Having two cameras allows two different viewpoints of the newborn. Video streams were encoded with MPEG-4 encoding, under AVC container, at 25 frames per second. The protocol for camera positioning has been approved by the Biomedical Engineering Department of CHU Rennes. Furthermore, the Hygiene department of CHU Rennes helped us to elaborate a sterilization protocol.

When the newborn is in an incubator, it has been decided, in coordination with the doctors and nurses, to place the camera inside the incubator at the foot of the newborn, in order to be the least disturbing for the medical staff. Thus it enables to acquire images of good quality, with a narrow image view. In this context, the newborns are continuously monitored during one day up to 11 days.

When the newborn is in an open bed, there is more freedom in the camera placement, so the view of the newborn might be quite narrow or wide. Compared to the incubator, this leads to very different viewpoints of the newborn and the surrounding environment, which explains why we split the database according to the bed type. In the context of open bed, there are up to five recordings per newborn, each one lasting few hours up to 24 hours and spaced out of ten days.

Data was collected in six different NICUs during routine care with no constraints apart from very few instructions on camera placement, so there is a wide variety of background and lighting. The variety of the images is illustrated in Fig. 4 to 6, according to the three target classes.

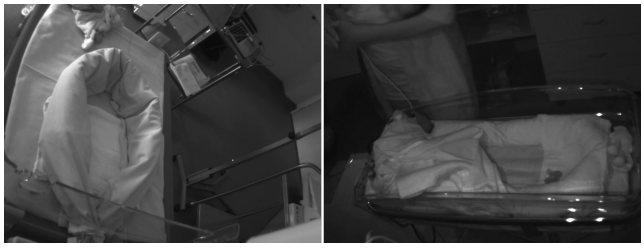
Fig. 4 shows example images of the class "adult & newborn present." The adult presence might be temporally short, for example when the nurse measures temperature, or long, for example during an intervention. In incubator, it is characterized by the emergence of the adult's arm. The hand may be hardly visible when emerging in a shaded area (see the top right image, inside the white circle). Sometimes it can be characterized by the emergence of the adult's body as well. In open bed, the encountered situations are much more varied than in incubator since the surrounding environment might occupy a large part of the image. For example in the image in the bottom right, the adult is standing up in the background. We can see that the cradle is near the bed of the parents, who might lay on it for several hours.

Fig. 5 shows example images of the class "newborn absent." We encounter images with a blank mattress or the bed body. The adult may be present for changing the bedding (see the top right and bottom right images).

Fig. 6 shows example images of the class "newborn solely present." In incubator, the newborn may be totally covered by the blanket or there may be a stuffed toy near the newborn (see the top right image in the left bottom corner). In the case

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

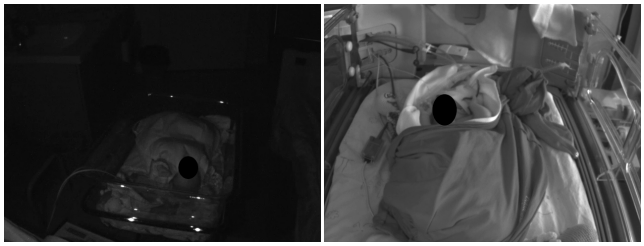
(a) Incubator



(b) Open bed

Fig. 5. Example images of the class "newborn absent"

(a) Incubator



(b) Open bed

Fig. 6. Example images of the class "newborn solely present"

of the open bed, the bed might be in any orientation. During the night, the surrounding environment is not visible.

Table II describes the population of the Digi-NewB project that has been used in this study, including 120 newborns from the six different NICUs. Since there can be several recordings for one newborn, for each newborn we computed the cumulative recording duration and then we computed the mean and standard deviation along newborns. The distribution of total recording duration per NICU is as follows: 35.6%, 3.2%, 17.5%, 3.6%, 20.2%, 19.8,%. This reflects the respective ability of each NICU to include newborns and the differences of acquisition conditions in each NICU.

B. Database of still images

In order to select the deep neural networks for bottleneck features extraction and then train the binary classifiers (see Fig. 1), we created a database of still images with 57 newborns in incubator for a total of 4961 images and 60 newborns in open bed for a total of 3859 images. The following procedure was adopted for extracting the images: a set of images were randomly picked from the Digi-NewB database and has been manually annotated by one annotator with regards to the three classes (see Table I). Table III details the distribution of images in each class.

TABLE II
SUBSET OF THE POPULATION OF DIGI-NEWB USED FOR THIS STUDY. GA STANDS FOR GESTATIONAL AGE. PMA STANDS FOR POST-MENSTRUAL AGE. STD STANDS FOR STANDARD DEVIATION AND IS GIVEN IN NUMBER OF DAYS (FOR GA, PMA, RECORDING DURATION) OR GRAMS (BIRTH WEIGHT).

Number of newborns (% male)	120 (60%)
Mean birth weight in grams (STD)	1319.71 (681.17)
Mean GA (STD)	29+3 (28.94)
Mean PMA (STD)	33+0 (35.09)
Mean total recording duration per newborn in days (STD)	7.81 (8.01)

TABLE III
DISTRIBUTION OF THE CLASSES IN THE DATABASE OF STILL IMAGES

Class	Number of images	
	Incubator	Open bed
"adult & newborn present"	2150	1361
"newborn absent"	1346	1370
"newborn solely present"	1465	1128
Total	4961	3859

C. Database of videos

For validating and testing the whole process on videos, we created a database of videos split into a validation set and a test set for both bed types. All the video recordings have been manually annotated by one annotator with regards to the three classes (see Table I). The validation set is composed of newborns that are included in the validation set used for training the final binary classifiers, so they are included in the database of still images. The test set is composed of newborns that are not included in the database of still images.

Table IV details the content of the database of videos with the number of newborns, the total quantity of data in number of days and the distribution of the three classes. We can see that in incubator this distribution is similar between the validation set and the test set, whereas in open bed there is a lot of variety. The difference of data quantity between incubator and open bed is explained by the fact that, according to the protocol of the Digi-NewB project, for a single newborn a recording can last up to ten days in incubator, whereas in open bed there are up to five recordings, each lasting up to 24 hours and spaced out of ten days.

V. RESULTS

In this section, we report the results of our experiments. They are divided in two parts: the classification of still images

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

TABLE IV
DISTRIBUTION OF THE CLASSES IN THE DATABASE OF VIDEOS. ANP STANDS FOR THE CLASS "ADULT & NEWBORN PRESENT." NA STANDS FOR THE CLASS "NEWBORN ABSENT." NSP STANDS FOR THE CLASS "NEWBORN SOLELY PRESENT." STD STANDS FOR STANDARD DEVIATION.

		Validation	Test
Incubator	Number of newborns	4	7
	Data quantity (days)	28.02	28.48
	Mean % ANP (STD)	8.69 (1.60)	8.24 (3.76)
	Mean % NA (STD)	11.58 (8.76)	9.42 (4.34)
	Mean % NSP (STD)	79.73 (9.77)	82.34 (6.81)
Open bed	Number of newborns	5	9
	Data quantity (days)	3.17	7.05
	Mean % ANP (STD)	22.24 (21.87)	22.34 (16.54)
	Mean % NA (STD)	14.14 (6.01)	31.33 (15.83)
	Mean % NSP (STD)	63.62 (20.23)	46.33 (19.00)

and the classification of videos. In the former we focus on the selection of the pre-trained network for bottleneck features extraction and the selection of the classifier. The latter allows us to focus on the parameters setting of decision intervals fusion and to assess the performance of the whole method.

Experiments are conducted with scikit-learn [34] and keras with tensorflow backend.

A. Classification of still images

1) Protocol: As illustrated in Fig. 1, two binary classifications are considered. First, "newborn" classification focuses on newborn presence detection, regardless of adult presence: the negative class is "newborn absent" and the positive class is the union of "adult & newborn present" and "newborn solely present." We used class balancing during the training as the dataset is heavily unbalanced for this problem. Secondly, "adult" classification focuses on adult presence detection: the negative class is "newborn solely present" and the positive class is "adult & newborn present."

For each classification problem, we have to select the pre-trained deep neural network for bottleneck features extraction on the one hand (see Subsection III-B) and the classifier on the other hand (see Subsection III-C). We used the database of still images (see Subsection IV-B) for the experiments.

We adopted a 5-fold cross-validation strategy with preservation of classes proportion. For each fold, the database is split in three sets: train (72.25%), validation (12.75%) and test (15%). We make sure that a newborn is included only in one of the three sets. The performance is assessed with the following metrics expressed as a percentage: balanced accuracy, recall and specificity. We computed the mean metrics over the test sets of the cross-validation folds. The advantage of balanced accuracy is to provide an overall performance, while taking into account unbalance between the classes. Recall and specificity are not impacted by unbalance between the classes.

Regarding SVM, regularization parameter is set on validation set so that it maximizes balanced accuracy. Regarding NN, as a first step we chose the stochastic gradient descent (SGD) as optimizer with decay and momentum set to 0. We conducted a grid search on the following parameters:

- Size of the fully-connected layer: $\{64, 128, 256\}$,
- Learning rate of SGD: $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$.

Then, we kept the couple of parameters leading to the best balanced accuracy and we conducted another grid search on the following parameters:

- Optimizer: {SGD, rmsprop, adagrad, adam}
- Decay of the optimizer: $\{0, 10^{-2}, 10^{-3}\}$,
- Momentum of the optimizer (when applicable): $\{0, 0.5, 0.9\}$.

In the end we kept the set of parameters leading to the best balanced accuracy.

2) Results: Table V reports the performance of classification in incubator on the database of still images. Regarding "newborn" classification, the highest balanced accuracy is obtained with the combination of MobileNetV2 and SVM. Regarding "adult" classification, the highest balanced accuracy is obtained with the combination of InceptionV3 and NN, followed by the combination of InceptionV3 and SVM. The combination of MobileNetV2 and NN gives a better balanced accuracy than InceptionV3 and SVM but it does not offer a good trade-off between recall and specificity.

Table VI reports the performance of classification in open bed on the database of still images. Regarding "newborn" classification, the best balanced accuracy is obtained with the combination of VGG16 and NN, followed by the combination of VGG16 and SVM. Regarding "adult" classification, the best balanced accuracy is obtained with the combination of VGG16 and SVM. The performances are significantly lower than the other binary problems, which was to expect since the variety of situations is much wider.

For comparison purpose, both in incubator and open bed, we reported the performance of classification with logistic regression (LR) applied directly on the pixels of the image (resized to 250 pixels x 250 pixels). We can see that the performance is really poor with LR compared to deep transfer learning, which witnesses the fact that the good performance of our method is not due to a highly biased database, but due to the pre-trained neural network.

B. Classification of videos

1) Protocol: In incubator, for "newborn" classification we selected the combination of MobileNetV2 and SVM, for "adult" classification we tested the combination of InceptionV3 and both SVM and NN. In open bed, for "newborn" classification we tested the combination of VGG16 and both SVM and NN, for "adult" classification we selected the combination of VGG16 and SVM. The binary classifiers have been trained on the database of still images (see Subsection IV-B) with 80% of images for training and 20% for validation.

We used the database of videos (see Subsection IV-C) for assessing the performance of the whole process. We down-sampled the videos to one image per second because of heavy processing. It does not cause loss of information as we do not seek to detect events that are shorter than one second.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

TABLE V
PERFORMANCE OF BINARY CLASSIFICATION IN INCUBATOR ON THE DATABASE OF STILL IMAGES

Bottleneck	Classifier	Set	"newborn" classification			"adult" classification		
			Balanced accuracy	Recall	Specificity	Balanced accuracy	Recall	Specificity
None	LR	*Val*	*55.71*	*65.82*	*45.60*	*52.37*	*60.39*	*44.35*
		Test	56.00	69.88	42.11	54.09	63.28	44.91
VGG16	SVM	*Val*	*87.63*	*92.85*	*82.41*	*82.37*	*83.83*	*80.90*
		Test	89.60	92.81	86.38	82.39	84.76	80.01
	NN	*Val*	*85.02*	*92.74*	*77.30*	*77.90*	*81.35*	*74.45*
		Test	86.15	92.64	79.67	76.06	83.02	69.10
Mobile-NetV2	SVM	*Val*	*89.91*	*91.96*	*87.86*	*82.35*	*86.17*	*78.52*
		Test	**91.21**	92.85	**89.57**	81.22	87.38	75.06
	NN	*Val*	*89.41*	*93.62*	*85.19*	*82.94*	*96.13*	*79.75*
		Test	91.03	**94.48**	87.58	84.40	**87.59**	77.20
Incep-tionV3	SVM	*Val*	*88.37*	*90.25*	*86.48*	*82.74*	*85.08*	*80.39*
		Test	88.16	90.03	86.29	83.50	84.04	82.96
	NN	*Val*	*88.64*	*94.47*	*82.81*	*84.06*	*86.64*	*81.48*
		Test	89.68	93.79	85.57	**85.36**	85.36	**85.13**

TABLE VI
PERFORMANCE OF BINARY CLASSIFICATION IN OPEN BED ON THE DATABASE OF STILL IMAGES

Bottleneck	Classifier	Set	"newborn" classification			"adult" classification		
			Balanced accuracy	Recall	Specificity	Balanced accuracy	Recall	Specificity
None	LR	*Val*	*49.55*	*60.41*	*38.69*	*60.39*	*63.75*	*57.04*
		Test	51.10	55.48	46.72	62.36	63.65	61.08
VGG16	SVM	*Val*	*87.21*	*86.96*	*87.46*	*73.01*	*72.37*	*73.64*
		Test	82.97	83.54	82.39	**74.72**	**78.81**	70.62
	NN	*Val*	*88.07*	*89.55*	*86.59*	*70.37*	*73.03*	*67.70*
		Test	**83.98**	83.36	84.60	69.63	77.19	62.08
Mobile-NetV2	SVM	*Val*	*84.69*	*82.14*	*87.24*	*74.92*	*81.03*	*68.81*
		Test	79.46	76.33	82.59	74.25	75.22	**73.29**
	NN	*Val*	*84.57*	*86.99*	*82.16*	*73.77*	*78.43*	*69.10*
		Test	80.46	82.61	78.30	72.72	73.56	71.88
Incep-tionV3	SVM	*Val*	*86.16*	*83.36*	*88.97*	*68.39*	*77.65*	*59.13*
		Test	81.94	76.76	**87.12**	69.00	72.13	65.86
	NN	*Val*	*86.24*	*89.04*	*83.45*	*70.87*	*72.04*	*69.70*
		Test	81.68	**83.87**	79.49	72.99	71.30	74.69

Regarding decision intervals fusion, there are three parameters to set (see Subsection V-A.1 for the definition of the classification problems):

- δ_0^n for negative class of "newborn" classification,
- δ_1^n for positive class of "newborn" classification,
- δ_1^a for positive class of "adult" classification.

In our database, we observed several situations where the presence of an adult lasts for only one second. So, we decided not to perform decision intervals fusion on the negative class of "adult" classification ("newborn solely present") in order to avoid changing the class of images with an adult present to "newborn solely present."

The thresholds were set separately for incubator and open bed thanks to a grid search with the following values: δ_0^n and δ_1^n ranging from 0 second to 30 seconds with a step of 2 seconds, δ_1^a ranging from 0 second to 120 seconds with a step of 5 seconds. We selected the triplet $(\delta_0^n, \delta_1^n, \delta_1^a)$ that leads to the best balanced accuracy on the validation set.

2) Results: For "newborn" classification, the grid search showed that δ_0^n allows to decrease the recall and increase the specificity, whereas δ_1^n has the opposite effect with less impact. A combination of the two can lead to an increase in the balanced accuracy. First, δ_0^n is used to significantly increase the specificity, then the induced decrease of the recall is compensated with δ_1^n. In our experiments, the impact of δ_0^n stops being significant above 10 seconds. The optimal values are $\delta_0^n = 4$ seconds and $\delta_1^n = 30$ seconds in incubator, $\delta_0^n = 8$ seconds and $\delta_1^n = 28$ seconds in open bed. The choice of limiting the value of δ_1^n below 30 seconds is arbitrary. In open bed, SVM outperformed NN when applied on videos, so we chose to report the results with SVM.

Regarding "adult" classification, the grid search showed that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

TABLE VII

PERFORMANCE OF CLASSIFICATION ON VIDEOS IN INCUBATOR. FUSION STANDS FOR DECISION INTERVALS FUSION. BA STANDS FOR BALANCED ACCURACY. R1, R2 AND R3 STANDS RESPECTIVELY FOR RECALL OF "ADULT & NEWBORN PRESENT," "NEWBORN ABSENT" AND "NEWBORN SOLELY PRESENT" CLASSES.

Method	Fusion	Set	BA	R1	R2	R3
Motion [13]	No	Val	-	87.78	-	88.98
		Test	-	80.99	-	**90.98**
3-class [17]	Yes	Val	85.09	90.48	78.86	85.93
		Test	80.50	82.47	75.42	83.60
Propo-sed method	No	Val	86.09	79.98	88.83	**89.45**
		Test	80.91	65.68	88.27	88.77
	Yes	Val	**88.77**	**93.13**	**89.23**	83.94
		Test	**86.02**	**84.37**	**89.18**	84.50

δ_1^a allows to increase the recall and decrease the specificity while increasing the balanced accuracy. The optimal values are $\delta_1^a = 40$ seconds in incubator and $\delta_1^a = 120$ seconds in open bed. In incubator, SVM outperformed NN when applied on videos, so we chose to report the results with SVM.

We compared the performance of our method with our previous work [17] and the method of [13]. In our previous work, we used the same technique of deep transfer learning, but instead of training two binary classifiers, we trained one 3-class classifier. The method of [13] only deals with adult detection, so the detection of the newborn is the manual annotation and we do not report balanced accuracy and recall of "newborn absent". The detection of the adult is performed by analyzing the motion in the edge of the image and requires a manual initialization.

Table VII reports the performance of classification in incubator. The decision intervals fusion allows to increase the balanced accuracy, particularly on the test set, while increasing recall of "adult & newborn present" and decreasing recall of "newborn solely present." Since our priority in our application context is to minimize the number of false positives with regards to "newborn solely present," it is desirable to increase the recall of "adult & newborn present" and "newborn absent." Compared to the 3-class model [17], the combination of binary classifiers allows to significantly increase the recall of "newborn absent," while keeping a similar recall on the two other classes. Compared to [13], our method reaches a greater recall of "adult & newborn present" and a lower recall of "newborn solely present," so we have a slightly better ability to detect adult presence.

Table VIII gives the mean confusion matrix computed on the test set in incubator. We can see that most of the errors occur mainly when incorrectly classifying "adult & newborn present" as "newborn solely present" and vice versa.

Table IX reports the performance of classification in open bed. The decision intervals fusion allows to increase balanced accuracy both on the validation set and the test set by significantly increasing the recall of "adult & newborn present" and decreasing the recall of "newborn solely present," which is desirable in our application context. Compared to the incubator, the performance is very low and there is a poor

TABLE VIII

MEAN CONFUSION MATRIX COMPUTED ON TEST SET OF THE DATABASE OF VIDEOS IN INCUBATOR. ANP STANDS FOR THE CLASS "ADULT & NEWBORN PRESENT." NA STANDS FOR THE CLASS "NEWBORN ABSENT." NSP STANDS FOR THE CLASS "NEWBORN SOLELY PRESENT."

		Prediction		
		ANP	NA	NSP
Ground truth	ANP	84.37%	2.64%	12.99%
	NA	4.21%	89.18%	6.60%
	NSP	12.48%	3.02%	84.50%

TABLE IX

PERFORMANCE OF CLASSIFICATION ON VIDEOS IN OPEN BED. FUSION STANDS FOR DECISION INTERVALS FUSION. BA STANDS FOR BALANCED ACCURACY. R1, R2 AND R3 STANDS RESPECTIVELY FOR RECALL OF "ADULT & NEWBORN PRESENT," "NEWBORN ABSENT" AND "NEWBORN SOLELY PRESENT" CLASSES.

Method	Fusion	Set	BA	R1	R2	R3
Motion [13]	No	Val	-	68.21	-	**78.90**
		Test	-	**68.94**	-	54.72
3-class [17]	Yes	Val	72.91	62.15	77.69	78.90
		Test	57.44	58.95	64.98	55.62
Propo-sed method	No	Val	73.85	51.81	86.26	83.46
		Test	56.49	42.84	68.47	**65.78**
	Yes	Val	**79.48**	**76.16**	**89.67**	72.61
		Test	**58.46**	56.46	**69.16**	57.45

ability to generalize since the balanced accuracy drops by 21.02% between the validation set and the test set. The same conclusion applies to the 3-class method [17]. In this context, the method of [13] based on motion analysis in the image edge could be more suitable for adult detection, but it would need to cope with two difficulties. First, it may happen that the light is often turned on and off, which generates motion in the image edge and leads to an incorrect adult detection. Thus this affects the recall of "newborn solely present." Secondly, the parent may lie motionless on the bed for a long time, which leads to adult presence not detected. Thus this affects the recall of "adult & newborn present."

Table X gives the confusion matrix computed on the test set in open bed. We can see that regarding "adult & newborn present" and "newborn solely present," there are more images misclassified as "newborn absent" compared to the incubator (see Table VIII). Moreover, there is a more significant percentage of "newborn absent" images that are misclassified as "newborn solely present" compared to the incubator (see Table VIII).

VI. DISCUSSION

Overall, this work shows that a deep transfer learning approach is relevant for the detection of periods of sole presence of the newborn. Indeed, a good generalization of the model is obtained in incubator with more than 86% of balanced accuracy. The results are weaker in open bed but are still encouraging since it is the first time that the problems of adult presence detection and newborn presence detection have been considered together.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3062617, IEEE Journal of Biomedical and Health Informatics

TABLE X
MEAN CONFUSION MATRIX COMPUTED ON TEST SET OF THE DATABASE OF VIDEOS IN OPEN BED. ANP STANDS FOR THE CLASS "ADULT & NEWBORN PRESENT." NA STANDS FOR THE CLASS "NEWBORN ABSENT." NSP STANDS FOR THE CLASS "NEWBORN SOLELY PRESENT."

		Prediction		
		ANP	NA	NSP
Ground truth	ANP	56.46%	17.12%	26.41%
	NA	8.42%	69.16%	22.41%
	NSP	19.86%	22.69%	57.45%

The development of a method for automatic preterm newborn presence in NICU is a relatively new field of research. To date, the closest work to our study is [9]. In this work, two systems are developed: one for newborn detection (equivalent to our "newborn" classification) and one for intervention detection (equivalent to our "adult" classification). Their study took place in one NICU. Regarding "newborn" classification, they reached an accuracy of 98.8%, a recall of 100% and a specificity of 96.8%. Regarding "adult" classification, they annotated 214 hours of video recordings and reached an accuracy of 94.5%, a recall of 94.7% and a specificity of 94.4%. For both classification problems, their performances are higher than in our study. But it is worth to keep in mind that in [9], only video recordings in daylight illumination were considered, whereas we include both daytime and nighttime in our study. Moreover, in our study, there are six different NICUs and the method has been validated and tested on a larger database of videos with more than 1300 hours of annotated video recordings in incubator. So there is a broader variety of situations in our study which might explain the weaker performance.

However, some limitations can be identified in our study and offer prospects for improvement. For instance, potential sources of bias such as capture bias, negative bias and category bias can be discussed [35], [36].

First of all, although capture bias had been limited in our study thanks to the high diversity in collected images, we noticed a drop between the validation and the test set with regards to the class "adult & newborn present" in incubator (see Table VII). This may denote a lack of generalization of our model for this particular class. To cope with this difficulty, we could enhance the database of still images, which is used to train the binary classifiers.

Secondly, the negative bias induced by the two binary classifiers is restricted due to the clinical context of our application. Indeed, the situation where both newborn and adult are absent mostly concerns empty rooms with no activity, so that the diversity of this class is mainly based on the disparity of the viewpoints and the background of the rooms. However, the possibility of obstruction of the video camera was not taken into account and this may be a significant negative class bias. One way to solve this problem would be to integrate a class with obstructed camera into the learning process.

Thirdly, there may be a category bias in open bed as witnesses the poor performance in this case. We investigated this aspect further by splitting the results according to the type of open bed: warm radiant (see the right image in Fig. 4(b)) and cradle (see the left image in Fig. 4(b)). For the latter, the newborn is almost full-term and the cradle is near the parents' bed, which is partly visible in the image. These recordings are particularly difficult to analyze because of a wider variety of situations compared to recordings in warm radiant. We have noticed that the model has a better ability to generalize in warm radiant than in cradle. In future work, it may be relevant to split the data in open bed into two parts (warm radiant and cradle) and train two separate models. But this would require a much larger database of still images in order to cope with the wide variety of encountered situations.

Another limitation is that we take into account the full image which may cause the discard of relevant periods in clinical applications such as newborn motion analysis. Indeed, in open bed, images with an adult in the background would be discarded, despite the fact that the area of the newborn is not impacted by this presence. In order to be able to detect the adults only when they are manipulating the newborn, we could perform bed segmentation to extract the region of interest. Thus, this region would be the input for adult's detection and newborn motion estimation. The difficulty with this approach is that when the adult is in the background, he still may be visible next to the bed because of the open bed transparency.

VII. CONCLUSION

In the context of long-term video-based monitoring of preterm newborn activity in NICU, it is of primary importance to be able to accurately detect the presence of the newborn prior to automatic analysis such as motion estimation. In this paper, we introduced a new method of video-based detection of preterm newborn presence in incubator and open bed. To our knowledge, this is the first method exploiting videos acquired during standard clinical care of several NICUs. This problem was addressed as a classification problem and we proposed to fuse the decisions of two binary classifiers: one for newborn presence detection, the other one for adult presence detection. They both follow the same workflow based on deep transfer learning: bottleneck features, extracted from a pre-trained deep neural network, are used as input of a classifier. We then adopted a strategy of decision intervals fusion in order to take advantage of temporal consistency. A comparative study between several bottleneck features (extracted from VGG16, MobileNetV2 and InceptionV3 deep neural networks) and classifiers (SVM, NN) was conducted.

As a result, we obtained a good performance in incubator with a balanced accuracy of more than 86% on both the validation set and test set. This is a satisfying result and our method will be used in a clinical trial [37]. Regarding open bed, the balanced accuracy on the validation set reached almost 80% but dropped to 58% on the test set, which witnesses the fact that there are a lot more of variability in open bed whereas less data are available. This result is not satisfying and is not sufficient for a use in a clinical application involving open bed.

Further works could focus on methodological aspects and clinical applications. For the first point, the use of classifiers

or neural networks that take into account temporal information -for example, to model arrival and the departure times of the adults-, is one perspective direction. Additionally, in open bed, our method could benefit from a greater amount of data in order to model the wide variety of situations. For the second point, the proposed approach will be used in order to automatically analyse newborn motion, which is of primary importance for different clinical applications, including late-onset sepsis detection and evaluation of neurobehavioral development.

Acknowledgment

Results incorporated in this publication received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689260 (Digi-NewB project).

References

[1] F. Pisani and C. Spagnoli, "Monitoring of newborns at high risk for brain injury," *Italian journal of pediatrics*, vol. 42, no. 1, p. 48, 2016.

[2] W. Daw, R. Kingshott, R. Saatchi, D. Burke, A. Holloway, J. Travis, R. Evans, A. Jones, B. Hughes, and H. Elphick, "Medical devices for measuring respiratory rate in children: a review," *Journal of Advances in Biomedical Engineering and Technology*, vol. 3, pp. 21–27, 2016.

[3] A. C. Kevat, D. V. Bullen, P. G. Davis, and C. O. F. Kamlin, "A systematic review of novel technology for monitoring infant and newborn heart rate," *Acta Paediatrica*, vol. 106, no. 5, pp. 710–720, 2017.

[4] L. Atallah, A. Serteyn, M. Meftah, M. Schellekens, R. Vullings, J. Bergmans, A. Osagiator, and S. B. Oetomo, "Unobtrusive ecg monitoring in the nicu using a capacitive sensing array," *Physiological measurement*, vol. 35, no. 5, p. 895, 2014.

[5] S. Nukaya, M. Sugie, Y. Kurihara, T. Hiroyasu, K. Watanabe, and H. Tanaka, "A noninvasive heartbeat, respiration, and body movement monitoring system for neonates," *Artificial Life and Robotics*, vol. 19, no. 4, pp. 414–419, 2014.

[6] P. Arlotto, M. Grimaldi, R. Naeck, and J.-M. Ginoux, "An ultrasonic contactless sensor for breathing monitoring," *Sensors*, vol. 14, no. 8, pp. 15 371–15 386, 2014.

[7] J. D. Kim, W. H. Lee, Y. Lee, H. J. Lee, T. Cha, S. H. Kim, K.-M. Song, Y.-H. Lim, S. H. Cho, S. H. Cho, *et al.*, "Non-contact respiration monitoring using impulse radio ultrawideband radar in neonates," *Royal Society open science*, vol. 6, no. 6, p. 190149, 2019.

[8] C. B. Pereira, X. Yu, T. Goos, I. Reiss, T. Orlikowsky, K. Heimann, B. Venema, V. Blazek, S. Leonhardt, and D. Teichmann, "Noncontact monitoring of respiratory rate in newborn infants using thermal imaging," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 4, pp. 1105–1114, 2018.

[9] M. Villarroel, S. Chaichulee, J. Jorge, S. Davis, G. Green, C. Arteta, A. Zisserman, K. McCormick, P. Watkinson, and L. Tarassenko, "Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit," *npj Digital Medicine*, vol. 2, no. 1, pp. 1–18, 2019.

[10] S. Cabon, F. Porée, A. Simon, O. Rosec, P. Pladys, and G. Carrault, "Video and audio processing in paediatrics: a review," *Physiological Measurement*, vol. 40, no. 2, p. 02TR02, 2019.

[11] H. Rahmati, O. M. Aamo, Ø. Stavdahl, R. Dragon, and L. Adde, "Video-based early cerebral palsy prediction using motion segmentation," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 3779–3783.

[12] S. Cabon, F. Porée, A. Simon, B. Met-Montot, P. Pladys, O. Rosec, N. Nardi, and G. Carrault, "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.

[13] S. Cabon, F. Porée, A. Simon, M. Ugolin, O. Rosec, G. Carrault, and P. Pladys, "Motion estimation and characterization in premature newborns using long duration video recordings," *IRBM*, vol. 38, no. 4, pp. 207–213, 2017.

[14] L. Curzi-Dascalova, "Physiological correlates of sleep development in premature and full-term neonates," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 22, no. 2, pp. 151–166, 1992.

[15] R. Joshi, D. Kommers, L. Oosterwijk, L. Feijs, C. Van Pul, and P. Andriessen, "Predicting neonatal sepsis using features of heart rate variability, respiratory characteristics and ecg-derived estimates of infant motion," *IEEE journal of biomedical and health informatics*, 2019.

[16] E. H. Verstraete, K. Blot, L. Mahieu, D. Vogelaers, and S. Blot, "Prediction models for neonatal health care–associated sepsis: A meta-analysis," *Pediatrics*, vol. 135, no. 4, pp. e1002–e1014, 2015.

[17] R. Weber, A. Simon, F. Porée, and G. Carrault, "Deep transfer learning for video-based detection of newborn presence in incubator," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 2147–2150.

[18] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, K. H. Grunewaldt, and R. Støen, "Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study," *Developmental Medicine & Child Neurology*, vol. 52, no. 8, pp. 773–778, 2010.

[19] S. Orlandi, K. Raghuram, C. R. Smith, D. Mansueto, P. Church, V. Shah, M. Luther, and T. Chau, "Detection of atypical and typical infant movements using computer-based video analysis," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 3598–3601.

[20] E. A. Ihlen, R. Støen, L. Boswell, R.-A. de Regnier, T. Fjørtoft, D. Gaebler-Spira, C. Labori, M. C. Loennecken, M. E. Msall, U. I. Möinichen, *et al.*, "Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study," *Journal of Clinical Medicine*, vol. 9, no. 1, p. 5, 2020.

[21] Y. S. Dosso, S. Aziz, S. Nizami, K. Greenwood, J. Harrold, and J. R. Green, "Video-based neonatal motion detection," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 6135–6138.

[22] X. Long, E. van der Sanden, Y. Prevoo, L. ten Hoor, S. den Boer, J. Gelissen, R. Otte, and E. Zwartkruis-Pelgrim, "An efficient heuristic method for infant in/out of bed detection using video-derived motion estimates," *Biomedical Physics & Engineering Express*, vol. 4, no. 3, p. 035035, 2018.

[23] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1717–1724.

[24] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[25] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials*, vol. 157, pp. 322–330, 2017.

[26] E. Rezende, G. Ruppert, T. Carvalho, A. Theophilo, F. Ramos, and P. de Geus, "Malicious software classification using VGG16 deep neural network's bottleneck features," in *Information Technology-New Generations*. Springer, 2018, pp. 51–59.

[27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[31] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[33] 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.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[35] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *CVPR 2011*. IEEE, 2011, pp. 1521–1528.
[36] T. Tommasi, N. Patricia, B. Caputo, and T. dataset bias," in *Domain adaptation in computer vision applications*. Springer, 2017, pp. 37–55.
[37] NEOVIDEO : Impact of monitoring motor activity by video analysis on the sleep of very preterm infants. ClinicalTrials.gov, Rennes University Hospital, Identifier NCT04624347, 2020. [Online]. Available: https://www.clinicaltrials.gov/show/NCT04624347