engineering proceedings

MDPI

Proceeding Paper

Neonatal Activity Monitoring by Camera-Based Multi-LSTM Network †

Imre Jánoki [1,*], Ádám Nagy [1], Péter Földesy [2], Ákos Zarándy [2], Máté Siket [3], Judit Varga [4] and Miklós Szabó [4]

[1] Institute for Computer Science and Control, Pázmány Péter Catholic University, 1088 Budapest, Hungary; nagyadam@sztaki.hu
[2] Institute for Computer Science and Control, 1111 Budapest, Hungary; foldesy.peter@sztaki.hu (P.F.); zarandy.akos@sztaki.hu (Á.Z.)
[3] Institute for Computer Science and Control, Óbuda University, 1084 Budapest, Hungary; siket.mate@sztaki.hu
[4] Budapest Division of Neonatology Istst Department of Pediatrics, Department of Obstetrics and Gynecology, Semmelweis University, 1082 Budapest, Hungary; varga.judit@med.semmelweis-univ.hu (J.V.); szabo.miklos@med.semmelweis-univ.hu (M.S.)
* Correspondence: janoki.imre.gergely@sztaki.hu
† Presented at the IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability, Tainan, Taiwan, 2–4 June 2023.

Abstract: The objective evaluation of an infant's activity and sleep pattern is critical in improving the comfort of the babies and ensuring the proper amount of quality sleep. The predefined behavioral states of an infant describe their consciousness and arousal level. The different states are characterized by different movements, body tone, eye movements and breath patterns. To recognize and adapt to these states is an essential part of development-friendly caring. It affects the neonate's sleep, influencing their brain development, while improving the bonding between mother and baby, and feeding is more successful during the state of quiet awakened. It can be a more difficult task to determine the level of arousal in premature neonates. In preterm clinics, the general practice is continuous observation, requiring the attention of the hospital staff. To create an automated, more objective system, helping the hospital staff and the parents, we developed a multi-RNN (multi-recurrent neural network) network-based solution to solve this classification problem, which works on a time-series-like feature set, extracted from cameras' video feeds. The set is composed of video actigraphy features, video-based respiration signal and additional descriptors. We separate infant caring from undisturbed presence based on our previous ensemble network solution. The network was trained and evaluated using our database of 402 h of footage, collected at the Neonatal Intensive Care Unit, Dept. of Neonatology of Pediatrics, Dept. of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary, with all-day recordings of 10 babies.

Keywords: actigraphy; sleep patter; motion estimation; breath rate; respiration rate; LSTM; non-contact; premature infant; neonatal

Citation: Jánoki, I.; Nagy, Á.; Földesy, P.; Zarándy, Á.; Siket, M.; Varga, J.; Szabó, M. Neonatal Activity Monitoring by Camera-Based Multi-LSTM Network. *Eng. Proc.* **2023**, *55*, 16. https://doi.org/10.3390/engproc2023055016

Academic Editors: Teen-Hang Meen, Kuei-Shu Hsu and Cheng-Fu Yang

Published: 28 November 2023

1. Introduction

The objective diagnosis of sleep patterns is essential to improve sleep comfort and to measure the effectiveness of interventions. Continuous observation and adjusting a caring schedule is important in development-friendly caring, but direct observations result in a burden on medical staff.

Polysomnography (PSG)—considered to be the "gold standard"—is complicated, needs hospital resources and causes discomfort for the infants. On the other hand, actigraphy is a cheap easy long-term measurement technique. Though its equivalency is up to debate, it is found to be an effective tool. Its video-based version is also an intensively investigated alternative [1]. This is a contact-free method, providing actigraphy using a camera, which quantifies body movements during long-term awake and sleep assessments,

even in difficult situations, and, at the same time, it does not require wiring or attached sensors [2]. The classification of sleep–wake and intermediate stages, based on video observation with different solutions, are presented in [1,3,4].

Our aim is to quantitatively evaluate sleep–wake–caring states of infants, based on video cameras and recurrent neural networks (RNNs). In order to separate infant caring and presence we used our previous ensemble network solution [5]. This mentioned care state analyzer solution is also an integral part of the behavioral state classification algorithm presented here, and we refer to it as "Top-Level-Classification-Block" in [5]. This network is important because the determination of a behavioral state is only possible—or worthwhile—if there is a baby in the picture and there is no feeding, caring or other intervention happening. In these cases, the baby's behavioral state is trivial and cannot be classified, as described in [6].

It is important to note that, in this work, the output classes of the applied model are not the same as the sleep stages (e.g., NREM, REM) that you can read about in the mainstream literature. Instead, the output classes used here are based on the manual observations made by doctors and nurses, and they represent the activity phases of the observed infant [7]. This is a routinely employed classification in hospitals that doctors and nurses are trained for (e.g., in "Family and Infant Nerve Development Education" [6]). It also includes the state of sleep, but it is more about activity phases; henceforth we shall call them that. The literature defines five or six phases: quiet sleep, active sleep, intermediate, quiet alert, and active alert. In our case, medical advisors were most interested in how much the babies sleep, therefore, we merged some classes and simplified the classifications into the following three categories: sleep, intermediate, and alert. These three classes were satisfactory to estimate how much the babies sleep and to help move towards development-friendly caring.

2. Experimental Setup

The dataset used in the project was collected in cooperation with Semmelweis Medical University. We built a complete data collection system that monitored the infant with a conventional color camera (Basler acA2040-55uc, Basler AG, Ahrensburg, Germany) at 20 FPS, with a resolution of 500 × 500, that was able to record the physiological data of the babies in parallel and in sync. The recordings were made from several angles with zoom optics.

Later, it became necessary to observe the infants in closed incubators, even at night. We made a 3D printed capsule, which included both the camera and an infrared illuminator (Figure 1). Infrared illumination was only used for the night recordings. This solution always recorded the baby from the same fixed position on top of the incubator.

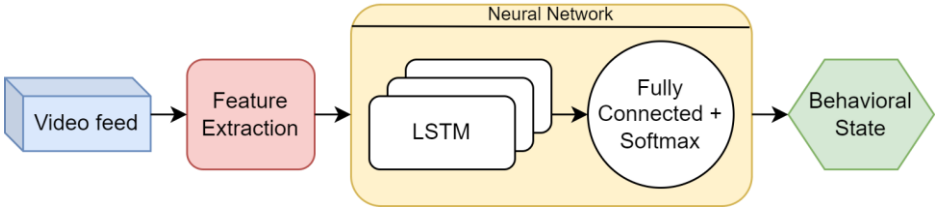

Figure 1. The data acquisition system that can be attached to the top of the incubator with suction cups. It includes a camera and an infrared illuminator.

Data storage, managing and annotation were carried out by custom software that we wrote. You can read more about this data collection and data managing system in [5]. A total of 402 h of data were collected in the neonatal intensive care unit from infants aged between 32 and 44 weeks. More details regarding the population of the recorded babies and their characteristics can be found in Table 1.

Table 1. Characteristics of the participants.

Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Recording time (hours)	96.7	5.5	39.4	27.4	51	105.5	50.1	36.4
Gender	F	M	M	F	F	M	F	F
Gestational age (weeks)	32	32 + 3	31 + 4	35 + 4	39	32	33	38 + 6
Birth weight (g)	2020	1840	1850	1870	3150	2120	2080	2840
Postnatal age (days)	4	4	10	8	4	7	2	7
Actual weight (g)	1900	1850	1680	1820	2905	2040	1960	3150
Length (cm)	46	44	-	45	57	45	44	48
Head circumference (cm)	32	29.5	-	32	34	30	32	33
Respiratory support	no	no	no	no	no	no	no	yes
Any drugs	no	no	no	no	yes	no	no	yes
Fitzpatrick scale	2	3	2	2	2	2	2	2

3. Methodology

We assumed that, if we extracted features from appropriate locations, in the absence of interventions or other disturbances, an RNN stack, which is designed for processing dynamic data, could solve the previously defined problem of activity phase classification.

Our solution for the activity phase classification algorithm can be divided into two main modules: **feature extraction** and **classification** (Figure 2). In addition, we extended the algorithm with a **scene analysis** module, which runs before the behavioral stage analysis and determines whether the current scene is suitable for the analysis. In the following subsections we introduce all modules separately.

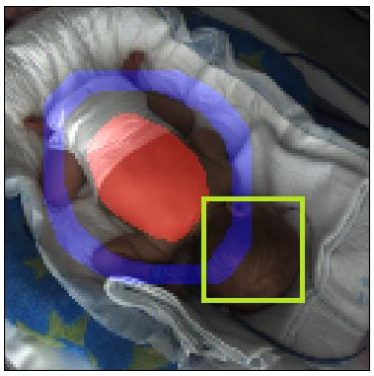

Figure 2. Summary of the presented behavioral state classification algorithm. The colors showing the different regions of interests: the head, the torso and the area around the torso.

3.1. Scene Analysis

First, we needed to determine that there was a baby in the picture, i.e., they had not been taken out of the incubator and there was no feeding or other care occurring. For this, we used the LSTM-based (long short-term memory) procedure that we introduced in [5] as "Top-Level-Classification-Block". This algorithm works very similarly to the behavioral state classification, that we explain in detail later, and it uses similar features calculated on the entire images. The vectors, constructed from extracted features, were processed by an RNN, which had a sequential input and a classification output. The network contained two fully connected (FC) layers with a Rectified Linear Unit (ReLU) activation function and a stateful LSTM cell. The final layer is a SoftMax one, which provides the classification probability-type output with the following classes:

- Baby is present and a respiration-like signal is detected.

- Baby is present and is showing intensive motion, interpreted as random and frequent self-motion.
- Caring or other intervention happens.
- No motion or minimal motion can be found in the incubator, but the baby is detected.
- The baby is not detected in the scene; empty incubator.
- Multiple subcategories incorporating unacceptable camera image quality and possible errors: low light conditions, blurry view, camera image is saturated, or consecutive frames do not differ from each other.

The top-level classification achieved 97.9% sensitivity and 97.5% specificity on the data set introduced in [5].

3.2. Feature Extraction

The feature extraction module started with the detection of different regions of interest (ROI). We started from the assumption that there were regions in the image that carried special information, separate from the other areas that could be useful to us in classifying activity phases (Figure 3). The examples are as follows:

- The whole picture;
- The baby's whole body;
- The area around the abdomen and torso;
- A ring around the baby's abdomen including the limbs;
- The infant's face.

Figure 3. Regions that may be of key importance in determining activity phases. These are used as different ROIs in feature extraction.

Finding pixels for these regions was a segmentation task. Except for the baby's entire body, we segmented each area and collected movement, color and intensity information from them separately.

It was easy to conclude that these regions listed above carried specific information. The need to use the full picture was trivial. The abdominal region is the area where respiration can be observed, which is an important descriptor for determining activity phases [8–10]. The "ring around the baby's abdomen" was also important: this area was likely to include the majority of the limbs on the image, if the parameters were chosen correctly, and so we can extract information about the limb movements of the baby. The last region was the baby's face, giving us information about whether the baby's eyes are open or not, which is also an important descriptor.

The next step of this module was the feature extraction itself. To determine what features to use, we evaluated the ones we used in [5], which was in a similar domain. The selected features to extract were as follows:

- Image brightness;
- HSV (hue, saturation, value) image;

- Optical flow;
- Euclidean map;
- Respiration.

The listed features were calculated for all the ROIs that we defined as important regions above. The ROIs were stored as binary images (masks) that contained ones and zeros. The area designated by the ROI was obtained by multiplying the current image by the given ROI, using elementwise matrix multiplication. After the different features were calculated on the masked images, intra-frame statistics like the mean value and standard deviation were calculated by:

$$F_{mean} = \mu = \frac{1}{N}\sum_{i=1}^{N} I_t(i), \quad (1)$$

$$F_{std} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_t(i)-\mu)^2} \quad (2)$$

where N is the number of pixels, and $I_t(i)$ is the intensity value of the i-th pixel of the current frame. Inter-frame statistics like optical flow or the Euclidean map (F_{euc}) were also calculated:

$$f_x u + f_y v + f_t = 0;\ u = \frac{dx}{dt};\ v = \frac{dy}{dt}; \quad (3)$$

$$F_{euc} = \frac{dx}{dt}\frac{dx}{dt} + \frac{dy}{dt}\frac{dy}{dt}, \quad (4)$$

where f_x and f_y are the image gradients and f_t is the gradient along time, while u and v represents the displacements in the x and y directions during time dt.

Our basic assumption was that these intra- and inter-frame statistics, calculated for each of the regions, should be sufficient to classify the current activity state of infants. To analyze the data from a dynamic perspective, we had to transform them to another form which can be interpreted by RNNs. Since we did not need pixel-level interpretation, we converted them to 1D waveforms. This was usually carried out by calculating the average value for each feature.

3.3. Classification

As we classified using dynamic time-series data, it was evident to use recurrent neural networks, which have already been used successfully in many applications to solve time–domain classification problems like EEG classification [11–15]. LSTM and GRU (Gated Recurrent Unit) are such recurrent models. For our experiments, we implemented an LSTM stack and a GRU stack as well, extended with fully connected layers. Stacks constructed from these recurrent architectures were suitable for working on the multivariate input features. The expected outputs were the earlier mentioned activity phases, which were annotated by doctors on our video recordings.

Considering the parameters of the used RNN stack, we examined using a 2-layer LSTM, and a 2-layer GRU network with a cell state of 35 in both cases, and a hidden state of 35 lengths, respectively.

During the 100-epoch long training, we used the "CrossEntropy" loss function and "Adam-Optimizer". The database was split into three different sets: the training set, the test set and the validation set. We used Comet-ML (Comet.ml, NY, USA) for the visualization of loss, epoch loss, and for hyperparameter tuning.

4. Discussion

Our approach combines the activity classification with predictions of expected activity and with an observation of a broader view—of the incubators providing automatic statistics regarding procedure time, empty view, parents presence, etc. over an extended period of hospitalization.

Our system starts with the scene analysis, where we apply an LSTM-based method. This approach provided an impressive performance in the "scene-analyzing" step:

- Accuracy: 98.1%;
- Sensitivity: 97.5%;
- Specificity: 97.9%;
- Precision: 98.1%.

For the classification, two recurrent neural networks were tested. These were LSTM- and GRU-based. Our experiments showed that we could achieve slightly higher accuracy in this application by using GRU. These networks work on sequences of a certain fixed length (L) that heavily affects their performance and necessary hyperparameters. We tested a set of different input sequence lengths and found L = 200 to be the best.

The results with the better performing GRU stack and L = 200 are as follows:

- Accuracy: 82.60%;
- Sensitivity: 78.40%;
- Specificity: 87.67%;
- Precision: 88.46%.

The total number of trainable parameters of the GRU stack was 23,014.

This performance could be further increased by using additional features, for example we can specifically detect the eye and add its condition (opened or closed) as an additional feature. Besides adding eye detection, our ongoing work primarily focuses on increasing our database as sleep and wake scoring depends heavily on variables that are susceptible to the high heterogeneity of the subjects. We are also experimenting with measurements undertaken with different devices. Tuning the network hyperparameters and using different sensitivity thresholds is also still in progress.

5. Conclusions

We have shown that an RNN-based network may be suitable for classifying behavioral states using video recordings. We have shown a feature extraction method that is one possible way to create an input that can be used by these networks. We hypothesize that the presented automated behavioral state classification architecture may later be suitable to replace human observation to a certain level.

Author Contributions: Conceptualization, I.J. and Á.N.; methodology, I.J., Á.N., P.F. and M.S. (Máté Siket); software, I.J. and Á.N.; validation, J.V. and M.S. (Miklós Szabó); formal analysis, Á.N.; investigation, I.J. and Á.N.; resources, Á.Z. and M.S. (Miklós Szabó); data curation, I.J., Á.N., M.S. (Máté Siket) and J.V.; writing—original draft preparation, Á.N. and I.J.; writing—review and editing, I.J.; visualization, I.J. and Á.N.; supervision, P.F., Á.Z. and M.S. (Miklós Szabó); project administration, Á.Z. and M.S. (Miklós Szabó); funding acquisition, Á.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hungarian grant NKFIH-1019658 and the Ministry of Innovation and Technology NRDI Office, Hungary within the framework of the Artificial Intelligence National Laboratory Program.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Semmelweis University (SE IRB number 265/2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethics and personal rights.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Long, X.; Espina, J.; Otte, R.A.; Wang, W.; Aarts, R.M.; Andriessen, P. Video-based actigraphy is an effective contact-free method of assessing sleep in preterm infants. *Acta Paediatr.* **2020**, *110*, 1815–1816. [CrossRef] [PubMed]
2. Yavuz-Kodat, E.; Reynaud, E.; Geoffray, M.-M.; Limousin, N.; Franco, P.; Bourgin, P.; Schroder, C.M. Validity of actigraphy compared to polysomnography for sleep assessment in children with autism spectrum disorder. *Front. Psychiatry* **2019**, *10*, 551. [CrossRef] [PubMed]
3. Liao, W.-H.; Yang, C.-M. Video-based activity and movement pattern analysis in overnight sleep studies. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.
4. Unno, M.; Morisaki, T.; Kinoshita, M.; Saikusa, M.; Iwata, S.; Fukaya, S.; Yamashita, Y.; Nakayama, M.; Saitoh, S.; Iwata, O. Validation of actigraphy in hospitalised newborn infants using video polysomnography. *J. Sleep Res.* **2021**, *31*, e13437. [CrossRef] [PubMed]
5. Nagy, Á.; Földesy, P.; Jánoki, I.; Terbe, D.; Siket, M.; Szabó, M.; Varga, J.; Zarándy, Á. Continuous camera-based premature-infant monitoring algorithms for nicu. *Appl. Sci.* **2021**, *11*, 7215. [CrossRef]
6. Warren, I.; Mat-Ali, E.; Green, M.; Nyathi, D. Evaluation of the family and infant neurodevelopmental education (FINE) programme in the UK. *J. Neonatal Nurs.* **2019**, *25*, 93–98. [CrossRef]
7. Read, D.J.; Henderson-Smart, D.J. Regulation of breathing in the newborn during different behavioral states. *Annu. Rev. Physiol.* **1984**, *46*, 675–685. [CrossRef] [PubMed]
8. Maurya, L.; Kaur, P.; Chawla, D.; Mahapatra, P. Non-contact breathing rate monitoring in newborns: A review. *Comput. Biol. Med.* **2021**, *132*, 104321. [CrossRef] [PubMed]
9. Jorge, J.; Villarroel, M.; Chaichulee, S.; Guazzi, A.; Davis, S.; Green, G.; McCormick, K.; Tarassenko, L. Non-contact monitoring of respiration in the neonatal intensive care unit. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 286–293.
10. Rossol, S.L.; Yang, J.K.; Toney-Noland, C.; Bergin, J.; Basavaraju, C.; Kumar, P.; Lee, H.C. Non-contact video-based neonatal respiratory monitoring. *Children* **2020**, *7*, 171. [CrossRef] [PubMed]
11. Wang, P.; Jiang, A.; Liu, X.; Shang, J.; Zhang, L. LSTM-based EEG classification in motor imagery tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 2086–2095. [CrossRef] [PubMed]
12. Zeng, H.; Yang, C.; Dai, G.; Qin, F.; Zhang, J.; Kong, W. EEG classification of driver mental states by deep learning. *Cogn. Neurodynamics* **2018**, *12*, 597–606. [CrossRef] [PubMed]
13. Hu, X.; Yuan, S.; Xu, F.; Leng, Y.; Yuan, K. Scalp EEG classification using deep bi-LSTM network for seizure detection. *Comput. Biol. Med.* **2020**, *124*, 103919. [CrossRef] [PubMed]
14. Rana, R. Gated recurrent unit (gru) for emotion classification from noisy speech. *arXiv* **2016**, arXiv:1612.07778v1.
15. Chen, J.X.; Jiang, D.M.; Zhang, Y.N. A hierarchical bidirectional gru model with attention for eeg-based emotion classification. *IEEE Access* **2019**, *7*, 118530–118540. [CrossRef]