A Hybrid DCNN-SVM Model for Classifying Neonatal Sleep and Wake States Based on Facial Expressions in Video

Muhammad Awais, Xi Long, *Senior Member, IEEE*, Bin Yin, Saadullah Farooq Abbasi, Saeed Akbarzadeh, Chunmei Lu, Xinhua Wang, Laishuan Wang, Jiong Zhang, Jeroen Dudink, and Wei Chen, *Senior Member, IEEE*

***Abstract*—Sleep is a natural phenomenon controlled by the central nervous system. The sleep-wake pattern, which functions as an essential indicator of neurophysiological organization in the neonatal period, has profound meaning in the prediction of cognitive diseases and brain maturity. In recent years, unobtrusive sleep monitoring and automatic sleep staging have been intensively studied for adults, but much less for neonates. This work aims to investigate a novel video-based unobtrusive method for neonatal sleep-wake classification by analyzing the behavioral changes in the neonatal facial region. A hybrid model is proposed to monitor the sleep-wake patterns of human neonates. The model combines two algorithms: deep convolutional neural network (DCNN) and support vector machine (SVM), where DCNN works as a trainable feature extractor and SVM as a classifier. Data was collected from nineteen Chinese neonates at the Children's Hospital of Fudan University, Shanghai, China. The classification results are compared with the gold standard of video-electroencephalography scored by pediatric neurologists. Validations indicate that the proposed hybrid DCNN-SVM model achieved reliable performances in classifying neonatal sleep and wake states in RGB video frames (with the face region detected), with an accuracy of 93.8 $\pm$ 2.2% and an F1-score 0.93 $\pm$ 0.3.**

***Index Terms*—Neonatal sleep monitoring, video and image analysis, facial expression, deep convolutional neural network, support vector machine.**

Manuscript received July 31, 2020; revised February 25, 2021 and April 2, 2021; accepted April 6, 2021. Date of publication April 15, 2021; date of current version May 11, 2021. This work was supported in part by the Shanghai Municipal Science and Technology Major Project under Grant 2017SHZDZX01, and in part by the National Key R&D Program of China under Grant 2017YFE0112000. *(Corresponding authors: Wei Chen; Xi Long; and Chunmei Lu.)*

Muhammad Awais, Saadullah Farooq Abbasi, and Saeed Akbarzadeh are with the Center for Intelligent Medical Electronics, Department of Electronic Engineering, School of Information Science and Technology, Fudan University, Shanghai 200433, China (e-mail: 17110720061@fudan.edu.cn; 18110720168@fudan.edu.cn; sd.akbarzadeh@gmail.com).

Xi Long is with the Philips Research, 5656 AE Eindhoven, The Netherlands, and also with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: x.long@tue.nl).

Bin Yin is with the Connected Care and Personal Health Department, Philips Research, Shanghai 200032, China (e-mail: bin.yin@philips.com).

Xinhua Wang is with the Department of Neurology, Children's Hospital of Fudan University, Shanghai 200032, China (e-mail: xiner.1211@163.com).

Chunmei Lu and Laishuan Wang are with the Department of Neonatology, Children's Hospital of Fudan University, Shanghai 200032, China (e-mail: luchunmei1975@163.com; laishuanwang@yahoo.com).

Jiong Zhang is with the Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles 90033 USA (e-mail: jiong.zhang@loni.usc.edu).

Jeroen Dudink is with the Department of Neonatology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, 3584 EA Utrecht, The Netherlands (e-mail: j.dudink@umcutrecht.nl).

Wei Chen is with the Center for Intelligent Medical Electronics, Department of Electronic Engineering, School of Information Science and Technology, and Human Phenome Institute,Fudan University, Shanghai 200433, China (e-mail: w_chen@fudan.edu.cn).

Digital Object Identifier 10.1109/JBHI.2021.3073632

I. Introduction

SLEEP is a natural quiescence state of the mind and body, which is associated with reduced responsiveness to external stimuli [1], [2]. According to research on human development in early life, sleep is an essential factor for the development of the nervous system in infants [3], [4]. Newborn babies usually sleep between 16 and 18 hours per day in equispaced periods. As age increases, sleep changes from an ultradian rhythm to a circadian rhythm [5]. Consistent evidence indicates that sleep is vital for the brain development of neonates (in particular for preterm infants) and help them in recovering from illness [2], [6]. Further, the reliable measures for the tracking and assessment of wake-sleep patterns, over multiple nights could potentially provide an indication of neonatal development over time [7], [8], [9], [10].

Sleep in infants can be scored as in multiple stages such as active sleep, deep sleep, and wake, using polysomnographic (PSG) and EEG features [1]–[4]. Indeed, PSG is the gold standard for sleep monitoring and VEEG is the gold standard for seizure detection. As EEG signal is the most important modality in both PSG and VEEG to measure the brain activity during sleep or seizure, neurologists in the hospital used PSG or Video-EEG data to annotate sleep states (wake and sleep), and sleep staging (active sleep, quiet sleep, and wake) respectively[11], [12], which requires a number of sensors and electrodes attached to an infant's body to collect electrophysiological signals. The attachment of adhesive sensors/electrodes can have adverse effects on an infant's skin, which could lead to an elevated risk of infections [13], [14]. In previous decades, several unobtrusive or minimally

2168-2194 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

obtrusive techniques have been developed for the objective of infant monitoring such as dry electrode [15], capacitive sensing [16], ballistocardiogram [17], remote ultra-wideband radar [18], [19], video camera [20], and near-infrared spectroscopy [21]. Among these, video monitoring appears to be a promising technique since it is contact-free and convenient to use both at home or in hospitals and it can analyze body movements of an infant [14]. Video-based approaches capture body movements that are highly associated with infant sleep-wake patterns [22]. In that study, a three-dimensional (3D) spatiotemporal-based motion detection method was proposed to quantify infants' full-body movements from video (i.e., video-based actigraphy) of adult sleepers and achieved an error of 5.8% in estimating sleep efficiency. Subsequently, Long *et al.* [23] employed the video-based actigraphy approach to identify the sleep/wake state for 10 healthy term infants. Tested on a monochrome video data set of daytime naps using a linear discriminant classifier, 92.0% mean accuracy was reported. However, they analyzed general full-body movements solely without specifying movement types and did not eliminate disturbances in video frames caused by other subjects or objects like caregiving activities, which likely result in increased false positives in detecting awakenings. Recent studies have indicated that the facial motor and neuron activity as well as facial expression of neonates are associated with certain physiological and behavioral changes during sleep [24], [25]. Ariyaratnam and Rood [26] measured facial skin temperature and, found that the highest temperature of the face was in the forehead area (c. 34 °C) and the lowest (c. 32 °C) in the cheek area. The temperature pattern changes in the facial region may be used to investigate and assess lesions in the peripheral branches of cranial nerves. Further, studies on cutaneous temperature manipulation [27] reveal that skin temperature variation strongly improved the two most typical age-related sleep problems a decreased slow wave sleep and an increased risk of early morning awakening. Inducing an increase of 0.4 °C in skin temperature was sufficient to almost double the proportion of nocturnal slow wave sleep and to decrease the probability of early morning awakening from 0.58 to 0.04. The characterization of neonatal facial expressions could provide essential information in classifying their sleep and wake states with reduced artifact interference as compared with using full-body movements. However, the analysis of facial expressions for unobtrusive neonatal sleep monitoring has not been studied thus far.

In recent years, deep learning (e.g., DCNN) algorithms have been widely used in different applications for biomedical image and physiological signal processing and analysis [28], [29]. Although their overall performances are promising, most of them were designed for a dedicated application rather than sleep monitoring in infants or neonates. For automated neonatal sleep state classification, for example, Koolen *et al.* [30] proposed a greedy algorithm using EEG recordings to classify active and quiet sleep for neonates; however, their approach may lead to a risk of overfitting by including too many redundant features. Further, Ansari *et al.* [31] designed an 18-layer CNN model to identify quiet sleep in preterm infants. A primary disadvantage is that the approach was based on a "trial-and-error" strategy, which might cause abundant room for further improving the network architecture. Palmu *et al.* [32] developed an EEG-based index for characterizing sleep states in early preterm infants. They found that infant's exhibit different sleep states, and a detection system that can handle artifacts (often occurring in EEG) is desired. To the best of our knowledge, this is the first study on video-based neonatal sleep monitoring using deep neural networks. In our previous research [33], we proposed to evaluate the use of existing pre-trained networks as a features extractor to perform neonatal sleep and wake states classification using different using video frames. From around 2-h Fluke video recording of seven neonates, we achieved a modest classification performance with an accuracy of 65.3%, with AlexNet using Fluke (RGB) video frames. This indicates that using a pre-trained model as a feature extractor could not fully suffice for highly reliable sleep and wake classification in neonates.

Various studies have demonstrated the effectiveness of combining DCNN for learning distinguishable features and support vector machine (SVM) for training a reliable classifier, thereby, outperforming DCNN with output activation function for classification [34], [36]. The success of the combinations inspire us to design a dedicated hybrid model for neonatal sleep-wake classification based on video data.

This work aims to investigate a non-contact approach (with a video camera) to classify sleep-wake states in neonates, with the advantage of no disruption for their sleep. More specifically, in this work, we will propose a hybrid DCNN-SVM model to perform the classification by exploiting facial expressions from (thermal and RGB) video frames. The paper is organized as follows. Section II describes the neonatal facial characteristics and the dataset information used in this study. The details of our proposed methodology are explained in Section III. Section IV presents and discusses the results. Finally, we conclude the work and provide future research directions in Section V.

II. Data Collection

Nineteen Chinese newborn infants were included in this study and the data collection was done with the help of a pediatrician at the Children's Hospital of Fudan University, Shanghai, China, between September 2017 and December 2018. The study protocol was approved by the internal ethics committee of the hospital. Since this study was observational in nature, no additional approval from an external ethics board was required. Table I provides detailed descriptions of patient demographics (including sex, gestational age, postmenstrual age, and weight), total sleep/wake time recorded, and the reason(s) for hospital admission per infant.

Video data were recorded using a Fluke infrared camera (TiX 580 Expert Series Thermal Imagers with a resolution of 640 $\times$ 480 pixels and a frame rate of 24 fps) [37]. It was placed with a "look-down" view to film the neonates in a good lighting condition, as depicted in Fig. 1. In addition, EEG data were collected simultaneously. One of the main advantages of using the Fluke TiX580 camera is that it can record multiple types of color palettes along with RGB videos. LaserSharp Auto Focus help us to calculate and displays the distance from the

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DEMOGRAPHICS, RECORDED SLEEP/WAKE TIME, AND REASON FOR HOSPITAL ADMISSION OF THE INFANTS INCLUDED IN THIS STUDY

#	Sex	GA (wk^{+d})	PMA (wk^{+d})	Weight (kg)	S/W (min)	Reason for admission
1	G	34^{+0}	35^{+5}	2.6	20/100	Preterm
2	B	39^{+0}	42^{+4}	4.5	20/100	Fever
3	G	36^{+4}	36^{+5}	2.6	100/10	Preterm/Anhelation
4	B	39^{+2}	40^{+6}	3.3	105/15	Cyanosis
5	B	37^{+1}	38^{+0}	2.9	50/70	Jaundice
6	G	35^{+3}	35^{+3}	2.2	55/65	Preterm
7	G	34^{+6}	35^{+3}	2.3	60/60	Preterm & Emesis
8	B	38^{+5}	39^{+0}	2.8	42/78	Hypoglycemia
9	B	37^{+5}	38^{+1}	3.7	55/65	Fever
10	B	38^{+6}	40^{+2}	3.0	25/95	Fever
11	G	40^{+4}	41^{+2}	2.7	90/30	Jaundice
12	G	34^{+1}	34^{+4}	2.0	60/60	Jaundice
13	B	35^{+3}	35^{+3}	2.2	65/55	Anhelation
14	G	35^{+0}	36^{+4}	3.2	95/25	Preterm
15	B	31^{+3}	31^{+3}	1.6	55/65	Preterm/Anhelation
16	G	40^{+5}	41^{+1}	3.2	55/65	Jaundice
17	B	36^{+1}	36^{+4}	3.1	95/25	Perterm/Jaundice
18	B	39^{+2}	39^{+3}	3.5	85/35	Jaundice
19	G	38^{+0}	38^{+4}	2.9	45/75	Jaundice

G: girl, B: boy, GA: gestational age, PMA: postmenstrual age during data collection, S: sleep, W: wake.

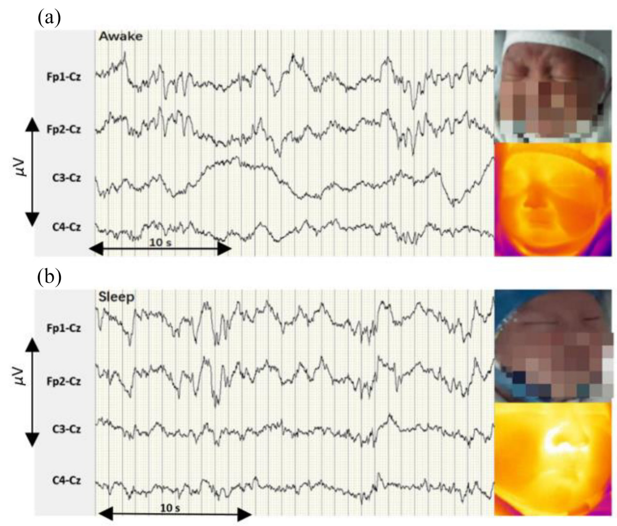

Fig. 2. Examples of 30-s EEG signal waveforms, and the corresponding RGB and thermal video frames of a neonate's face during (a) wake state and (b) sleep state. The amplitude scale (y-axis) represent in term of voltage (min:-100 μV, max:100 μV) vs time (x-axis).

Fig. 1. Examples indicating the setup for recording VEEG (video + EEG) data of a neonate in the hospital.

designated target and immediately adjusts the focus. For the neonatal facial database, we keep the distance between the subject and the camera to between 0.25 m and 0.36m. During the data collection process, a swaddle was used in order to keep the infant's position unchanged with his/her face visible (in a supine position) as much as possible. The experimental setup was designed according to the clinical study regulation of the hospital. For more details of the setup, please refer to our previous study [38]. For each neonate, a 120-min recording was collected to ensure the inclusion of one or two complete sleep-wake cycles [39].

Three neurologists, as per the hospital rostering system (one for each subject), performed manual sleep state scoring (on a 30-s basis) based on VEEG data [11]. NicoletOne system was used for EEG measurements and visual annotation. Electrodes were placed according to the standard 10-20 system for electrode placement [40]–[42]. Among the 19 infants, 15 include all the given electrodes except "T5 - 6," "F7 - 8" and "O1 - 2" (11 electrodes). For the remaining 4, "T5 - 6," "F7 - 8," "Cz" and "O1 - 2" were not recorded. These led to 10 channels included [41]. Note that, in neonates, recognizable EEG patterns during sleep are mostly visible at a postmenstrual age (PMA) of larger than 32 weeks. [43], [11]. Fig. 2 presents an example of 30-s EEG signal waveforms (from four channels) of a neonate during sleep and wakefulness with clear differences in term of both amplitude and frequency. In addition, the figure also indicates the corresponding RGB and thermal video frames from a neonate, exhibiting different facial expressions between sleep and wake states. During human annotation, the annotators observed the video when artifacts (for example due to body movements) appeared. The neurologists assigned either the label of "sleep" or "wake" for 30-s signal waveforms from VEEG; the parallel video data had a frame rate of 24 fps, corresponding to 720 frames per 30 s. Therefore, we labelled the 30-s video i.e., 720 frames as either "sleep" or "wake" depending on its corresponding VEEG 30-s sleep/wake annotation for classification.

III. METHODOLOGY

A. Face Region Detection

In order to enable the automatic identification of sleep or wake states for neonates by analyzing their facial expressions, it is necessary to have reliable neonatal face detection with minimal non-face related regions included. In our previous study [38], an automatic intensity-based method was used to detect an infant's facial region. For intensity-based face detection, initially, we transformed the RGB video frames into CIELAB color space after examining the color intensity values of each CIELAB channel, thereby the determining the threshold. The CIELAB single-channel frame was converted into binary frames and the connected area was separate from each other. Finally, the facial intensity region with the highest number of related/linked portions is imbricated on the original RGB video frames. Thereafter,

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

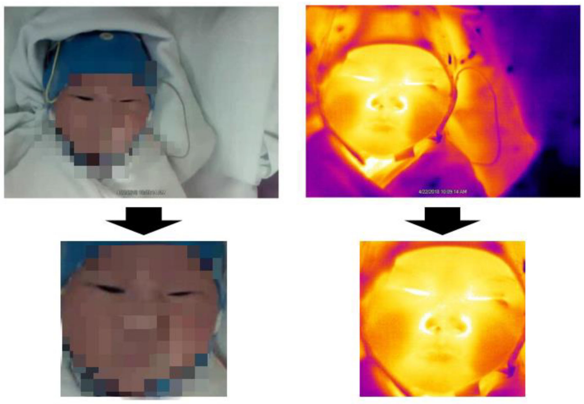

Fig. 3. Illustration of neonatal face region detection for both RGB and thermal videos using the intensity-based detection method [38].

the detected RGB facial region was mapped on the thermal video frames to extract the thermal facial region. Overall, we achieved a reliable performance of 95.8% accuracy. In the present study, this method was applied on raw RGB video frames to identify the facial region. Once the facial region was detected, it was then mapped onto the corresponding thermal video frames. Because DCNN prefers images with a fixed size, we thereby considered "cropping" the video frame resolution to 300 $\times$ 300 pixels with a centered infant face after face region detection (see Fig. 3), which sufficiently captured the entire face of all neonates with minimal inclusion of non-face related areas. Thereafter, we resized all the video frames to 224 $\times$ 224 pixels for the sake of allowing comparing with other well-known pre-trained neural networks (such as ResNet [44] and Inception [45], which used this resolution) in the future. More details of the intensity-based (face region) detection method can be found in our previous paper [38].

B. The Hybrid DCNN-SVM Model

1) DCNN: DCNN is a multi-layer neural network with a supervised learning architecture that can be viewed as the composition of a trainable classifier and an automatic feature extractor [46]. It has been proven that model generalization is attained with a large set of training data by using distortion techniques [47], [48]. Considering the usage of a relatively small dataset (small number of patients) in this study, training a complete DCNN classifier would not be the best choice for video-based neonatal sleep-wake classification. This is because infinitesimal optimization on loss function can be time-consuming and, more importantly, the trained model might not be able to achieve good generalization. Thus, rather than using the complete architecture of DCNN, we propose to extract valuable features with DCNN and then feed these features into a SVM model for classification.

Thus instead of using a highly diverse and complicated neural network, we considered a simplified DCNN architecture for this specific population. The architecture primarily includes five two-dimensional convolutional (Conv2D) layers with a rectified linear unit (*ReLU*) as the activation function for learning, and a flattened layer that concatenates the features followed by three fully connected layers (with *softmax* activation) for further classification of sleep and wake states. Here the reason for using *ReLU* instead of others like *tanh* is because it is time-efficient and can deal with the vanishing gradient problem [49]. Functioning as a feature extractor, convolutional layers are repossessed to discriminative features (or feature maps) through two operations: convolutional filtering and down-sampling (e.g., maxpooling). Commonly used filtering kernels with a size of 5 $\times$ 5 pixels (with stride $=$ 1 for convolving all pixels of input images) and down-sampling ratio of 2 are adopted [28]. Further the number of stacked feature maps (i.e., layer size) can have a significant influence on the generalization of a DCNN model [50]. Hence, we experimentally selected the layer sizes to be $\geq$32 by taking into account both accuracy and computation time. To avoid overfitting, we apply dropout layers (with a dropout rate of 0.1) at the end of each fully connected layer. In addition, for model training, the image batch size of 100 and epoch number of 100 are used. For DCNN-SVM, the output of each fully connected layer in DCNN is often considered to be multiple estimated normalized features of the input sample. In addition, the corresponding activation function can be used to calculate and optimize the features for each output unit.

2) SVM: SVM is a classification algorithm developed from the generalized portrait algorithm in pattern recognition and subsequently a linear SVM with hard margins was established [51]. It is considered a robust classifier, where the decision boundary can be determined by learning the maximum-margin separating hyperplane [52]. In this task for sleep-wake classification with video frames, a binary soft-margin linear SVM classifier [53] is applied. Let us consider an *n*-sample training dataset with feature vectors $\{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ and the corresponding labels $\{y_1, y_2, \ldots, y_i, \ldots, y_n\}$ where $y_i \in$ {-1, 1}; SVM attempts to identify the maximum-margin hyperplane by minimizing a loss function *U,* such that

$$\min_{w,b,s} \left(U = \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^{l} s_i \right), \tag{1}$$

subject to

$$y_i \left(\langle w, x_i \rangle + b \right) \geq 1 - s_i \text{ and } s_i \geq 0 \ \left(i = 12, \ldots, n \right), \tag{2}$$

where *w* is the weight vector, s_i are the slack variables, *b* is a scalar quantity, and the weight parameter *C* controls the trade-off between the minimum and maximum classification error. Here, we chose $C = 1$.

3) The Hybrid Model: We consider features from the three fully connected layers of DCNN after all convolutional layers. This results in three hybrid models including DCNN-SVM1, DCNN-SVM2, and DCNN-SVM3, with feature numbers 2048, 256, and 64, respectively. Fig. 4 illustrates the schematic network architecture of the proposed hybrid DCNN-SVM models. First, the video frames after face region detection and resampling are directed to the input layer, and the original DCNN with the output layer is trained until the training process converged. Thereafter, we take the output from each fully connected layer as feature vectors for training and testing an SVM classifier for classifying sleep and wake states. It is expected that, for this specific dataset

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

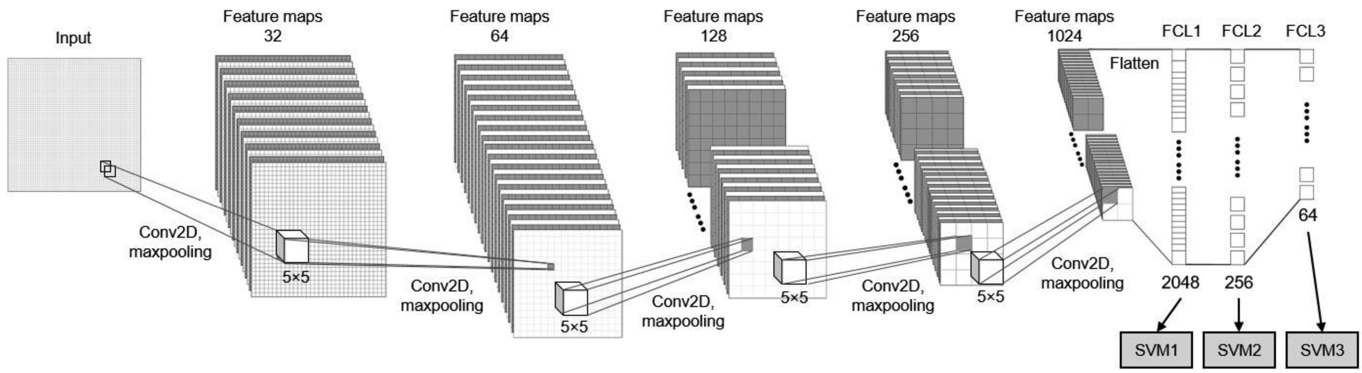

Fig. 4. The schematic network architecture of our proposed hybrid DCNN-SVM models combining DCNN layers and an SVM classifier, including DCNN-SVM1, DCNN-SVM2, and DCNN-SVM3. Conv2D: two-dimensional convolutions, FCL: fully connected layer.

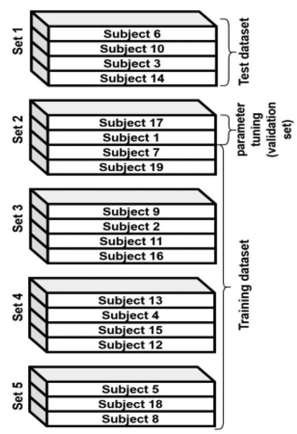

Fig. 5. The indicated partition of training, validation and test sets is an example during one cross-validation round.

and task, the hybrid models can synergize the strengths of both DCNN and SVM algorithms.

We used as Estar-S2600IP ((Intel Xeon(R) CPU E5-2650v4 @ 2.64GHz × 32) processor with 3 Nvidia 1080Ti Graphics cards with 64 GB RAM and another Dell precision Tower 7910(Intel Xeon(R) CPU E5-2687W v4 @ 3.00GHz × 24) with an Nvidia 1080Ti Graphics card to perform all the data processing. The overall computation time for a single fold is approximately 5-7 days, however once the model is trained it will cost much less time than training

C. Cross Validation and Evaluation Metrics

Given the relatively small data set, a five-fold cross validation was applied in this work in order to generate classification results for all data samples. The nineteen neonates were randomly divided into five folders (with three or four per fold). Fig. 5 depicts the overall description of our dataset and subject division for the five fold cross validation. During each round of cross validation, data from four folds were used for model training and parameter tuning, and data from the other fold were used for testing. After five rounds, sleep-wake classification results were obtained and the mean ± standard deviation (std) over infants are computed. With respect to wake detection, metrics including specificity, sensitivity, precision, accuracy, and F1-score were used to assess the classification performance. In this study, we compare the performance on the basis of raw RGB and thermal video frames as well as the corresponding cropped frames automatically detected facial region).

IV. Results and Discussion

With neonatal face region detection, approximately 0.65 out from among 0.68 million of video frames (in 224 × 224 pixels) were retained in the dataset for training and testing. To understand the separation of features between sleep and wake states, we visually compare the normalized feature values (from all features) in both classes on boxplots using facial region (Fig. 6) and raw (Fig. 7) video frames with different types of video (RGB and thermal) and classification models. Clearly, the feature values after the first fully connected layer (in correspondence to the DCNN-SVM1 model) using RGB video frames with the detected neonatal facial region appear to have the most separable distributions between sleep and wake times, indicating the promise in sleep-wake classification, whereas the other methods indicate a strong overlap between the two states.

The five-fold cross validation results for neonatal sleep-wake classification using different types (RGB and thermal) and sizes (raw and race region) of video frames as well as different DCNN-SVM models are summarized and compared in Tables II and III. Note that the choices of DCNN and SVM settings (or hyperparameters) were determined experimentally based on training accuracy. A reliable classification performance was achieved (see Table II), including an accuracy (mean ± std) of 93.8 ± 2.2% and an F1-score of 0.93 ± 0.03, at a specificity, a sensitivity, and a precision of 93.7 ± 1.6%, 93.8 ± 3.6%, and 92.9 ± 2.4%, respectively. As indicated in the tables, the hybrid model DCNN-SVM1 trained on cropped RGB video frames with face region detection significantly outperforms all other models. With regard to the detected face region, using the raw video frames may cause the inclusion of other surroundings (e.g., swaddle and cloths) that are irrelevant to infants' facial expression. These interferences can bring confounding factors

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

TABLE II
PERFORMANCE OF NEONATAL SLEEP-WAKE CLASSIFICATION (FIVE-FOLD CROSS VALIDATION) USING DIFFERENT DCNN-SVM MODELS BASED ON AUTOMATICALLY CROPPED FACE REGION VIDEO FRAMES (224X224 PIXELS)

Model	Type of video	Specificity	Sensitivity	Precision	Accuracy	F1-score
DCNN-SVM1	RGB	**93.7 ± 1.6%**	**93.8 ± 3.6%**	**92.9 ± 2.7%**	**93.8 ± 2.2%**	**0.93 ± 0.03**
	Thermal	40.0 ± 5.9%	81.1 ± 3.0%	67.9 ± 7.5%	58.6 ± 2.9%	0.74 ± 0.05
DCNN-SVM2	RGB	23.5 ± 2.8%	78.0 ± 4.2%	65.4 ± 9.8%	52.7 ± 3.6%	0.68 ± 0.07
	Thermal	78.4 ± 7.8%	35.0 ± 3.3%	41.7 ± 4.4%	58.8 ± 4.1%	0.44 ± 0.09
DCNN-SVM3	RGB	57.5 ± 6.0%	76.3 ± 7.1%	70.7 ± 7.4%	67.2 ± 5.4%	0.69 ± 0.06
	Thermal	52.4 ± 2.5%	75.6 ± 8.3%	60.3 ± 5.5%	64.6 ± 6.6%	0.63 ± 0.07

TABLE III
PERFORMANCE OF NEONATAL SLEEP-WAKE CLASSIFICATION (FIVE-FOLD CROSS VALIDATION) USING DIFFERENT DCNN-SVM MODELS BASED ON RAW VIDEO FRAMES (640 × 480 PIXELS)

Model	Type of video	Specificity	Sensitivity	Precision	Accuracy	F1-score
DCNN-SVM1	RGB	13.7 ± 1.6%	94.7 ± 2.8%	59.4 ± 1.4%	54.9 ± 2.3%	0.73 ± 0.02
	Thermal	25.9 ± 5.0%	67.4 ± 4.2%	42.2 ± 3.3%	46.1 ± 3.1%	0.52 ± 0.04
DCNN-SVM2	RGB	25.5 ± 4.7%	75.1 ± 3.5%	49.8 ± 3.4%	51.0 ± 3.1%	0.60 ± 0.03
	Thermal	29.4 ± 4.0%	77.6 ± 6.6%	49.7 ± 6.3%	56.9 ± 5.4%	0.58 ± 0.09
DCNN-SVM3	RGB	46.5 ± 6.2%	57.4 ± 6.0%	49.3 ± 4.2%	50.8 ± 3.0%	0.53 ± 0.05
	Thermal	45.7 ± 5.6%	59.1 ± 5.8%	53.3 ± 6.5%	49.0 ± 5.7%	0.56 ± 0.06

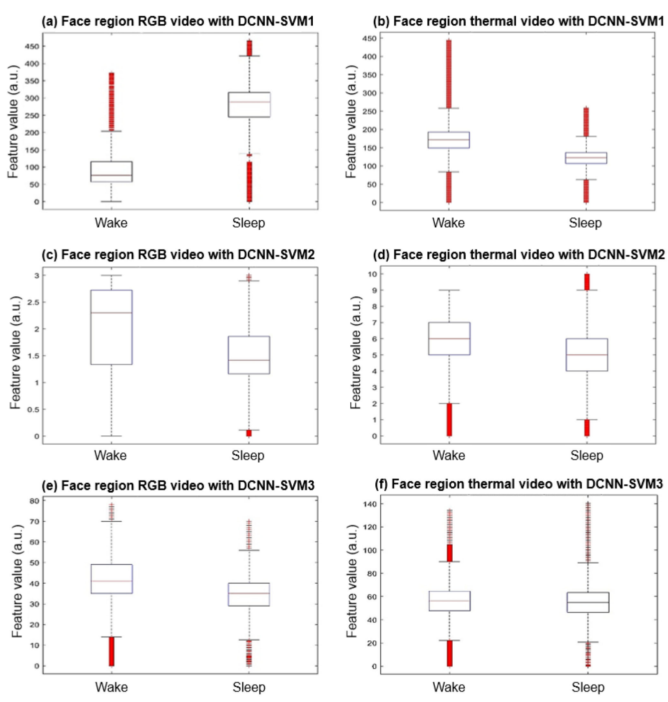

Fig. 6. Boxplots of all feature values in sleep and wake states using face region RGB (a, c, and e) and thermal (b, d, and f) video frames for different DCNN-SVM models.

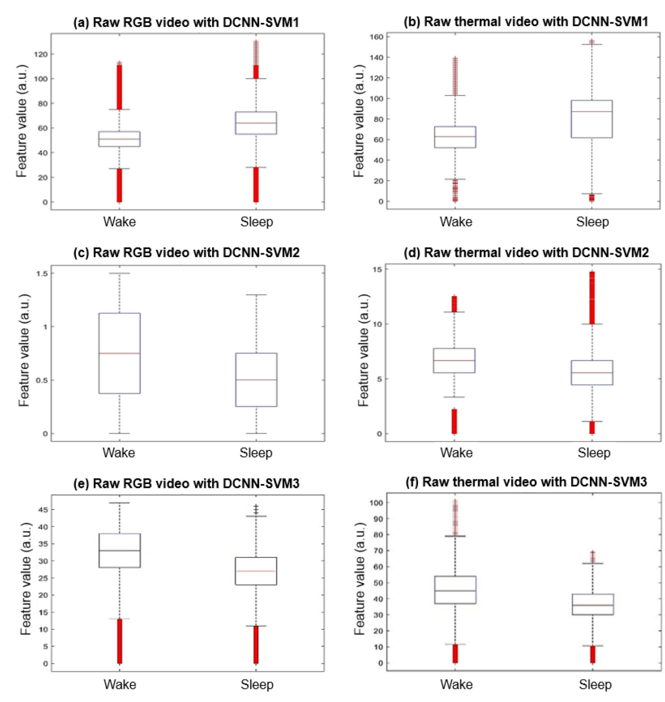

Fig. 7. Boxplots of all feature values in sleep and wake states using raw RGB (a, c, and e) and thermal (b, d, and f) video frames for different DCNN-SVM models.

to the DCNN feature learning and, thus, may fail in classifying sleep and wake states. In comparison with RGB, thermal video appears to be unable to reveal numerous details of an infant's facial expression, thereby producing a "blurred" face image. It is interesting to find that using more fully connected layers (two or three) with less generated features can decrease classification performance, thereby indicating ineffective neural network learning. Forcing a largely reduced feature set would likely miss valuable features that represent certain important facial expression information associated with neonatal sleep and wake states.

On the other hand, we also evaluated a "standalone" DCNN model without combining DCNN with SVM, where we found unpromising classification performance. The general learning

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
COMPARISON OF OUR PROPOSED VIDEO-BASED APPROACH WITH THE *STATE-OF-THE-ART* APPROACHES IN NEONATAL SLEEP-WAKE CLASSIFICATION

Author, year	Data/Sensor	Population	Age (PMA)	Algorithm	Accuracy
Harper, 1987 [57]	Respiration, HRV*	25 normal infants	First week, 1-6 months	DA[a]	77.0%
Sadeh, 1991 [58]	Actigraphy	11 children	12-48 months	SDA[b]	85.3%
Sazonova, 2004 [59]	Respiration, actigraphy	12 term and 14 preterm infants	34-46 weeks	LVQ-NN[c]	77-92%
So, 2005 [60]	Actigraphy	13 term and 9 preterm infants	<6 months	Thresholding	93.7%
Lewicke, 2008 [61]	HRV*	190 term/preterm infants	~ 45.5 ± 7.3 weeks	SVM	86.5%[#]
Tilmanne, 2009 [62]	Actigraphy	354 term/preterm infants	Mostly <6 months	ANN[d]	~80%
Fraiwan, 2014 [63]	EEG	14 term and 13 preterm infants	40 weeks	ANN[d]	81-86%
Long, 2019 [23]	Infrared video	10 healthy infants	3-9 months	LD[f]	92.0%
Werth, 2020 [64]	ECG	34 preterm infants	33 ± 2 weeks	RNN[e]	0.25/0.44[§]
This study	**RGB video**	**19 infants (various diseases)**	**31-43 weeks**	**Hybrid DCNN-SVM**	**93.8%**

*HRV: heart rate variability derived from ECG. [a]Discriminant analysis, [b]stepwise discriminant analysis, [c]learning vector quantization neural network, [d]artificial neural network, [e]recurrent neural network, [f]linear discriminant. [#]Results with 30% data have been rejected, [§]Results were measured by Cohen's kappa coefficient instead of accuracy (0.25 in classifying active sleep and wake, 0.44 in classifying quiet sleep and wake).

procedure of DCNN is similar to but an extension of the multi-layer perceptron (MLP). MLP attempts to reduce error by parameter optimization in the training dataset. When the first separating hyperplane is established via backpropagation, the training process ends at either the local or the global minima. For some of the data, the contribution of using MLP/DCNN as a classifier could be less than that of using SVM; this is on account of, the benefit of the SVM loss function in minimizing the generalization error with its superior regularization effects on unseen data with a certain dispersal of training data [54].

It is important to note that in order to include more data for DCNN learning, we considered all video frames as samples where their labels were derived from the corresponding 30-s window based sleep scoring. In contrast, the visual scoring of sleep states by pediatric sleep scorers as a gold standard relies on the "general picture" of VEEG data over the entire window. The transitions between sleep and wake often occur much less than for 30 s accompanied with changes in the amplitudes of EEG theta and delta activity or with the presence of arousals [11], [55]. This implies that within a 30-s window, it is possible that a few video frames can be more related to sleep while others are related to wake, or vice versa. In this case, the frame-based labelling is likely imprecise. However, this is not a big issue since the number of transitions between sleep and wake states are very small in our dataset (only a few per infant). Nevertheless, in order to be consistent with the human scoring "resolution," 30-s based decisions (sleep or wake) can be further made by analyzing all frame-based classification results within each 30s, for example, using a majority voting method.

Moreover, the proposed method in this study considers only "static" video frames with maximum visibility of their facial expressions and does not analyze temporal information (e.g., body or head movements over time). In fact, the level of infants' movements has been demonstrated to be highly associated with their sleep and wake states [23], [58]. With regard to this situation, combining body movement and changes in facial expression (or motor activity) is anticipated to further improve the classification performance. For example, incorporating a recurrent neural network model (such as long short-term memory [56]) after DCNN layers to learn time-variant context between video frames would help exploit body movements and facial motor activities over time, which merits further investigation.

Further, we compared the results of our method with others reported in the literature, as presented in Table IV. Since existing studies on video-based neonatal or infant sleep monitoring are limited, we also included those with the deployment of actigraphy or electrophysiologic signals such as electrocardiography (ECG), heart rate (variability), respiration, and EEG signals. Unlike this study, most previous studies used an obtrusive or invasive device with a sensor or electrode(s) attached to an infant's skin. In addition, a large number of those studies validated their approaches only on a cohort of healthy term infants or older children. Nevertheless, our proposed approach indicates better superiority in classifying infant sleep and wake states compared with the *state-of-the-art* approaches. Although video-based approaches have shown advantages for neonatal sleep monitoring, there are limitations. First, neonates routinely undertake caretaking from pediatricians or nurses in hospital settings [65]. The caregiving behaviors would lead to "occlusion" in video, which implies that the camera cannot always "see" the infant, thereby making face detection and then sleep-wake classification challenging. Second, the lighting or illumination in hospital rooms or neonatal intensive care units can change dramatically and, occasionally, can even dim to a few lux [66], which is difficult for the detection of infant's face region with RGB video. Thus, the validity of our approch in a low "ambient light" illumination is unclear and must be explored in the future.

In this study, to understand the overall separation of the discriminative facial features between sleep and awake states, we visually compare the normalized feature values (from all features) in both classes using boxplots as shown in Fig. 5 and Fig. 6. The feature values after the first fully connected layer (in correspondence to the DCNN-SVM1 model (Fig. 5a)) using RGB video frames with detected neonatal face region seem to have the most separable distributions between sleep and wake, indicating the promise in sleep-wake classification, whereas the other methods show a strong overlap between the two states. However, as our proposed approach is based on DCNN and SVM model that involves a large number of hyper-parameters features are challenging to be fully understandable. However,

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

in the explainable AI area, visualizing network activations (or activation maps) for CNN layers could provide a hint that which facial parts (e.g., edge of the face, eyes, mouth, etc.) are key features.

As mentioned earlier, this preliminary study for proof-of-concept was conducted in a well-controlled environment, aiming at examining the feasibility of neonatal sleep-wake classification through the learning of facial expressions. During the data collection process, we attempted to ensure that the infants were lying in bed with their face up in order to increase the visibility of their facial expressions. However, in real life, infants can move and change their sleeping position often with their face not always visible for example, while in a side or a prone position. Therefore, a larger set of data including more infants without the requirement of a specific sleeping position (visibility of face) must be acquired in future work to verify and improve the algorithm. For example, when the infant's face is not visible, we may use body movement to identify sleep/wake [67]. Combining such video-based actigraphy and face expressions will potentiall solve the problem. Furthermore, the used method to detect infants' face was based on a straightforward intensity–based algorithm. In the future, designing a robust automatic face segmentation/tracking algorithm could provide more accurate facial region for further sleep classification.

To enhance overall classification performance, it would be interesting to classify neonatal sleep and wake states by combining both RGB and thermal video frames and evaluate the performance of our proposed hybrid DCNN-SVM model. A potential challenge is that it might require using multiple cameras or at least multiple channels (RGB and thermal) in real applications, resulting in largely increased computational complexity.

Literature study reveals that, the EEG patterns of neonates with and without certain clinical events (e.g., seizure) are different [68], [69]. Seizure or pathological crying might be biasing the sleep state classification results. In future work, analysis of these abnormalities using camera-based AI models will be clinically relevant. In addition, identification of some other neurological abnormalities (such as sleep disorder, birth asphyxia, and encephalopathy) could be the future direction.

The promising results found in this paper compel us to go further in term of classifying a neonate's sleep states (active, deep, and awake), followed by creating a home-based monitoring system to follow the infant's sleep pattern and quality when she/he is discharged from the hospital. Moreover, in the hospital it will be helpful for pediatricians to set up a simple device to collect neonatal sleep data without any wires and this system could function as an aid tool for neurologists by automatically annotating the sleep states of neonates. Furthermore, video data analysis could possibly be used to detect vital signs such as heart rate [70], [71], respiration rate [20], SpO2 [72]. In realistic scenarios, having all those signals reliably measured with a video camera is challenging as the results can be influenced by many factors including motion artifacts, lighting variations, etc. Further research is required to improve the capability and robustness of measuring these vital signs using video-based approaches for the ultimate goal of classifying neonatal sleep states.

V. Conclusion

In this study, we proposed a hybrid DCNN-SVM model for neonatal sleep-wake classification. Thermal and RGB video frames with infant facial expressions were exploited in a novel non-contact manner to train and test the proposed approach. Validations were performed on a dedicated data set with approximately 0.65 million video frames and high classification performance was achieved by our proposed hybrid model. Our results indicate that when the face is clearly visible, the proposed approach is reliable and effective to be used for classifying sleep and wake states in neonates.

References

[1] V. Bertelle *et al.*, "Sleep in the neonatal intensive care unit," *J. Perinat. Neonatal. Nurs.*, vol. 21, no. 2, pp. 140–148, Apr./Jun. 2007.

[2] S. Graven, "Sleep and brain development," *Clin. Perinatol.*, vol. 33, no. 3, pp. 693–706, Sep. 2006.

[3] H. Roffwarg *et al.*, "Ontogenetic development of human sleep-dream cycle," *Science*, vol. 152, no. 3722, pp. 604–619, Apr. 1966.

[4] L. Hoe *et al.*, "Neurophysiologic assessment of neonatal sleep organization: Preliminary results of a randomized, controlled trial of skin contact with preterm infants," *Pediatrics*, vol. 117, no. 5, pp. 909–923, May 2006.

[5] M. Mirmiran *et al.*, "Development of fetal and neonatal sleep and circadian rhythms," *Sleep Med. Rev.*, vol. 7, no. 4, pp. 321–334, Aug. 2003.

[6] J. Bayer *et al.*, "Sleep problems in young infants and maternal mental and physical health," *J. Paediatric Child Health*, vol. 43, no. 1-2, pp. 66–73, Jan. 2007.

[7] S. Gertner *et al.*, "Sleep–wake patterns in preterm infants and 6 month's home environment: Implications for early cognitive development," *Early Hum. Dev.*, vol. 68, no. 2, pp. 93–102, Jul. 2002.

[8] P. Peirano *et al.*, "Sleep-wake states and their regulatory mechanisms throughout early human development," *J. Pediatr.*, vol. 143, no. 4, pp. 70–79, Oct. 2003.

[9] A. Sadeh *et al.*, "Actigraphic home-monitoring sleep-disturbed and control infants and young children: A new method for pediatric assessment of sleep-wake patterns," *Pediatrics*, vol. 87, no. 4, pp. 494–499, Apr. 1991.

[10] C. Acebo *et al.*, "Estimating sleep patterns with activity monitoring in children and adolescents: How many nights are necessary for reliable measures?," *Sleep*, vol. 22, no. 1, pp. 95–103, Feb. 1999.

[11] M. Grigg, "The visual scoring of sleep in infants 0 to 2 months of age," *J. Clin. Sleep Med.*, vol. 12, no. 3, pp. 429–445, Mar. 2016.

[12] G. Damberger *et al.*, "The visual scoring of sleep and arousal in infants and children," *J. Clin. Sleep Med.*, vol. 3, no. 2, pp. 201–240, Mar. 2007.

[13] R. Lloyd *et al.*, "Overcoming the practical challenges of electroencephalography for very preterm infants in the neonatal intensive care unit," *Acta Paediatr,* vol. 104, no. 2, pp. 152–157, Feb. 2015.

[14] J. Werth *et al.*, "Unobtrusive sleep state measurements in preterm infants—A review," *Sleep Med. Rev.*, vol. 32, pp. 109–122, Apr. 2017.

[15] J. Linde *et al.*, "Normal newborn heart rate in th first five minutes of life assessed by dry-electrode electrocardiography," *Neonatology*, vol. 110, no. 3, pp. 231–237, Sep. 2016.

[16] J. Werth *et al.*, "Automated preterm infant sleep staging using capacitive electrocardiography," *Physiol. Meas.*, vol. 40, no. 5, Jun. 2019, Art. no. 055003.

[17] W. Lee *et al.*, "Physiological signal monitoring bed for infants based on load-cell sensors," *Sensors*, vol. 16, no. 3, pp. 409, Mar. 2016.

[18] J. Kim *et al.*, "Non-contact respiration monitoring using impulse radio ultrawideband radar in neonates," *R. Soc. Open Sci.*, vol. 6, no. 6, Jun. 2019, Art. no. 190149.

[19] R. Goederen *et al.*, "Radar-based sleep stage classification in children undergoing polysomnography: A pilot-study," *Sleep Med.*, Mar. 2021. [Online]. Available: https://doi.org/10.1016/j.sleep.2021.03.022.

[20] M. Villarroel *et al.*, "Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit," *NPJ Digit. Med.*, vol. 2, p. 128, Dec. 2019.

[21] M. Wolf *et al.*, "A review of near infrared spectroscopy for term and preterem newborns," *J. Near Infrared Spectrosc.*, vol. 20, no. 1, pp. 43–55, Jan. 2012.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.

[22] A. Heinrich *et al.*, "Body movement analysis during sleep based on video motion estimation," in *Proc. IEEE Healthcom*, Lisbon, Portugal, 2014, pp. 539–543.

[23] X. Long *et al.*, "Video-based actigraphy for monitoring wake and sleep in healthy infants: A laboratory study," *Sensors*, vol. 19, no. 5, Mar. 2019.

[24] G. Zamzmi, R. Kasturi, D. Goldgof, R. Zhi, T. Ashmeade, and Y. Sun, "A review of automated pain assessment in infants: Features, classification tasks, and databases," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 77–96, Nov. 2018.

[25] D. Holditch *et al.*, "Sleeping and waking state development in preterm infants," *Early Hum. Dev.*, vol. 80, no. 1, pp. 43–64, Oct. 2004.

[26] S. Ariyaratnam, and J. P. Rood, "Measurement of facial skin temperature," *J. Dent.*, vol. 18, no. 5, pp. 250–253, Oct. 1990.

[27] R. Raymann *et al.*, "Skin deep: Enhanced sleep depth by cutaneous temperature manipulation," *Brain*, vol. 131, no. 2, pp. 500–513, Feb. 2008.

[28] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 31–40, Jan. 2017.

[29] B. Rim *et al.*, "Deep learning in physiological signal data, a survey," *Sensors*, vol. 20, no. 4, pp. 969, Feb. 2020.

[30] N. Koolen *et al.*, "Automated classification of neonatal sleep states using EEG," *Clin. Neurophysiol.*, vol. 128, no. 6, pp. 1100–1108, Jun. 2017.

[31] A. Ansari *et al.*, "Quiet sleep detection in preterm infants using deep convolutional neural networks," *J. Neural Eng.*, vol. 15, Dec. 2018, Art. no. 066006.

[32] K. Palmu *et al.*, "Sleep-wake cycling in early preterm infants : Comparison of polysomnographic recordings with a novel EEG-based index," *Clin. Neurophysiol.*, vol. 124, no. 9, pp. 1807–1814, Sep. 2013.

[33] M. Awais *et al.*, "Can pre-trained convolutional neural networks be directly used as a feature extractor for video-based neonatal sleep and wake classification?," *BMC Res Notes*, vol. 13, pp. 507, Oct. 2020.

[34] X. Niu, and C. Suen, "A novel hybrid CNN-SVM classifier for recognizaing handwritten digits," *Pattern Recogn.*, vol. 45, no. 4, pp. 1318–1325, Apr. 2012.

[35] D. Xue *et al.*, "CNN-SVM for microvascular morphological type recognition with data augmentation," *J. Med. Biol. Eng.*, vol. 36, pp. 755–764, Dec. 2016.

[36] A. Fred, and M. Agarap, "An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification," Feb. 2019, *arXiv:1712.03541*.

[37] *Fluke Corporation*, "*TiX500, tix520, tix560, tix580 expert series thermal imagers: users manual*," Revol. 1, Feb. 2015. Accessed: Apr. 11, 2020, [Online]. Available: https://www.instrumart.com/assets/Tix500-series-manual.pdf

[38] M. Awais *et al.*, "Novel framework: Face feature selection algorithm for neonatal facial and related attributes recognition," *IEEE Access*, vol. 8, pp. 59100–59113, 2020.

[39] M. Damberger, "Ontogeny of sleep and its functions in infancy, childhood, and adolescence," in *Sleep Disorders in Children*, Cham, Switzerland: Springer Nature, 2016, pp. 3–29.

[40] S. Abbasi *et al.*, "Automatic denoising and artifact removal from neonatal EEG," in *Proc. 3rd Int. Conf. Biol. Inf. Biomed. Eng.*, 2019, pp. 1–5.

[41] S. Abbasi *et al.*, "EEG-Based neonatal sleep-wake classification using multilayer perceptron neural network," *IEEE Access*, vol. 8, pp. 183025–183034, Nov. 2020.

[42] R. Homan *et al.*, "Cerebral location of international 10–20 system electrode placement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, Apr. 1987.

[43] R. Berry *et al.*, "The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications," *Amer. Acad. Sleep Med.*, 2014, Accessed: Apr. 10, 2021, [Online]. Available: https://aasm.org/clinical-resources/scoring-manual/

[44] K. He *et al.*, "Deep residual learning for image recognition," Dec. 2015, *arXiv:1512.03385*.

[45] C. Szegedy *et al.*, "Rethinking the inception architecture for computer vision," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, 2016, pp. 2818–2826.

[46] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural. Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.

[47] F. Lauer *et al.*, "A trainable feature extractor for handwritten digit recognition," *Pattern Recogn.*, vol. 40, no. 6, pp. 1816–1826, Jun. 2007.

[48] S. Dodge, and L. Karan, "Quality robust mixtures of deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5553–5562, Jul. 2018.

[49] X. Glorot *et al.*, "Deep sparse rectifier neural networks," in *Proc. Fourteenth Int. Conf. Artif. Intell. Statist., PMLR*, vol. 15, 2011, pp. 315–323.

[50] P. Zhou and J. Feng, "Understanding generalizaiton and optimization performance of deep CNNs," *PMLR*, vol. 80, pp. 5960–5969, 2018.

[51] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[52] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," in *Proc Mach. Learn. Appl.*, Cham, Switzerland: Springer Nature, 2001, pp. 249–257.

[53] I. Carmichael and J. S. Marron, "Geometric insights into support vector machine behavior using the KKT conditions," Oct. 2018, *arXiv:1704.00767*.

[54] Y. Tang, "Deep learning using linear support vector machines," Feb. 2015, *arXiv:1306.0239*.

[55] D. Weerd *et al.*, "Neonatal EEG. The international federation of clinical neurophysiology," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 52, pp. 149–157, 1999.

[56] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.,* vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[57] R. Harper *et al.*, "Machine classification of infant sleep state using cardiorespiratory measures," *Electroencephalogr. Clin. Neurophysiol.*, vol. 67, no. 4, pp. 379–387, Oct. 1987.

[58] A. Sadeh *et al.*, "Actigraphic home-monitoring sleep-disturbed and control infants and young children: A new method for pediatric assessment of sleep-wake patterns," *Pediatrics*, vol. 87, no. 4, pp. 494–499, Apr. 1991.

[59] E. Sazonov *et al.*, "Activity-based sleep-wake identification in infants," *Physiol. Meas.*, vol. 25, no. 5, pp. 1291–1304, Oct. 2004.

[60] K. So *et al.*, "Actigraphy correctly predicts sleep behavior in infants who are younger than six months, when compared with polysomnography," *Pediatr. Res.*, vol. 58, no. 4, pp. 761–765, Oct. 2005.

[61] A. Lewicke *et al.*, "Sleep versus wake classification from heart rate variability using computational intelligence: Consideration of rejection in classification models," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 108–118, Jan. 2008.

[62] J. Tilmanne *et al.*, "Algorithms for sleep-wake identification using actigraphy: A comparative study and new results," *J. Sleep Res.*, vol. 18, no. 1, pp. 85–98, Mar. 2009.

[63] L. Fraiwan and K. Lweesy, "Newborn sleep stage identification using multiscale entropy," in *Proc. IEEE MECBE*, Doha, Qatar, 2014, pp. 361–364.

[64] J. Werth *et al.*, "Deep learning approach for ECG-based automatic sleep state classification in preterm infants," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101663.

[65] E. Ingersoll and E. Thoman, "Sleep/wake states of preterm infants: Stability, developmental change, diurnal variation, and relation with caregiving activity," *Child Dev,* vol. 70, no. 1, pp. 1–10, Feb. 1999.

[66] M. Rea and M. Figueiro, "The NICU lighted environment," *Newborn Infant Nurs. Rev.*, vol. 16, no. 4, pp. 195–202, Dec. 2016.

[67] X. Long *et al.*, "Video-based actigraphy is an effective contact-free method of assessing sleep in preterm infants," *Acta Paediatr*, PMID: 33382134, Dec. 2021, doi: 10.1111/apa.15740.

[68] C. Panayiotopoulos, "*The Epilepsies: Seizures, Syndromes and Management*. Oxfordshire, U.K.: Bladon Medical Publishing, 2005. PMID: 20821848.

[69] B. Boashash *et al.*, "A methodology for time-frequency image processing applied to the classification of non-stationary multichannel signals using instantaneous frequency descriptors with application to newborn EEG signals," *EURASIP J. Adv. Signal Process*, vol. 117, May 2012.

[70] P. Addison *et al.*, "Video-Based physiologic monitoring during an acute hypoxic challenge: Heart rate, respiratory rate, and oxygen saturation," *Anesth. Analg.*, vol. 125, no. 3, pp. 860–873, Sep. 2017.

[71] R. Eline *et al.*, "The value of cardiorespiratory parameters for sleep state classification in preterm infants: A systematic review," *Sleep Med. Rev.*, vol. 58, 2021, Art. no. 101462.

[72] G. Jayatilaka *et al.*, "Non-contact infant sleep apnea detection," in *Proc. 14th Conf. Ind. Inf. Syst. (ICIIS)*, Kandy, Sri Lanka, 2019, pp. 260–265.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on June 17,2021 at 00:04:23 UTC from IEEE Xplore. Restrictions apply.