IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 20, 2021, accepted September 30, 2021, date of publication October 6, 2021, date of current version October 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3118227

Automatic Detection of Epileptic Seizures in Neonatal Intensive Care Units Through EEG, ECG and Video Recordings: A Survey

BENEDETTA OLMI[1], LORENZO FRASSINETI[1,2], (Member, IEEE), ANTONIO LANATÀ[1], (Member, IEEE), AND CLAUDIA MANFREDI[1], (Member, IEEE)

[1]Department of Information Engineering, Università degli Studi di Firenze, 50139 Firenze, Italy
[2]Department of Medical Biotechnologies, Università di Siena, 53100 Siena, Italy

Corresponding authors: Benedetta Olmi (benedetta.olmi@stud.unifi.it) and Lorenzo Frassineti (lorenzo.frassineti@student.unisi.it)

This work was supported in part by the Italian Ministry of Foreign Affairs and International Cooperation under Project MX18MO14.

ABSTRACT In Neonatal Intensive Care Units (NICUs), the early detection of neonatal seizures is of utmost importance for a timely, effective and efficient clinical intervention. The continuous video electroencephalogram (v-EEG) is the gold standard for monitoring neonatal seizures, but it requires specialized equipment and expert staff available 24/24h. The purpose of this study is to present an overview of the main Neonatal Seizure Detection (NSD) systems developed during the last ten years that implement Artificial Intelligence techniques to detect and report the temporal occurrence of neonatal seizures. Expert systems based on the analysis of EEG, ECG and video recordings are investigated, and their usefulness as support tools for the medical staff in detecting and diagnosing neonatal seizures in NICUs is evaluated. EEG-based NSD systems show better performance than systems based on other signals. Recently ECG analysis, particularly the related HRV analysis, seems to be a promising marker of brain damage. Moreover, video analysis could be helpful to identify inconspicuous but pathological movements. This study highlights possible future developments of the NSD systems: a multimodal approach that exploits and combines the results of the EEG, ECG and video approaches and a system able to automatically characterize etiologies might provide additional support to clinicians in seizures diagnosis.

INDEX TERMS Deep learning, ECG, EEG, HRV, machine learning, neonatal seizures, neonatal seizure detection, NICUs, NSD, video analysis, seizure detection.

I. INTRODUCTION

"Neonatal seizures are defined as paroxysmal alterations of neurological functions, that occur within the 28th day of life in full-term newborns" [1]. The occurrence of seizures is quite common during the neonatal period, especially in preterm newborns: the estimated incidence is about 1-5/1000 live births, and 8.6/1000 in Neonatal Intensive Care Units (NICUs) [2]. The immature brain is characterized by high hyper-excitability due to poor inhibitory mechanisms and a surplus of excitatory neurotransmitters. Thus, weakly propagated fragmentary seizures can be generated [3], [4].

In NICUs, the early detection of neonatal seizures is of utmost importance for an effective and efficient clinical intervention. Seizures occurring in newborns are quite different from those of adults and children. Poor clinical manifestations characterize up to 70% of all neonatal seizures, thus they can be confused with normal neonatal behaviour [5], [6]. For this reason, electroencephalographic (EEG) monitoring is considered the most appropriate diagnostic technique to identify neonatal seizures. Specifically, the American Clinical Neurophysiology Society (ACNS) has recently defined continuous video EEG (vEEG) as the gold standard in the diagnosis of neonatal seizures [7], [8]. In fact, EEG records the spontaneous electrical cerebral activity, and video recordings allow monitoring possible clinical manifestation of seizures. Commonly, both video and EEG signals are evaluated and interpreted by visual inspection. However, this process is time-consuming and requires expert staff available 24/24h. Therefore, computer-based and machine

The associate editor coordinating the review of this manuscript and approving it for publication was Gang Wang.

learning techniques would be helpful to support seizure detection [9], [10].

Over the years, several studies have proposed Neonatal Seizure Detection (NSD) systems to automatically detect and characterize critical events, using specific Artificial Intelligence (AI) techniques. These systems are mainly based on algorithms applied to EEG, electrocardiogram (ECG) and video signal. Specifically, EEG is usually investigated to identify the presence of irregularities or characteristic trends due to seizures [9], [11]–[22]. ECG is analyzed to evaluate alterations of the heart rate variability due to changes in the control of the cardiovascular system [23]–[25]. Few studies in the literature attempted to improve the NSD systems' performances by investigating the combination of EEG and ECG signals [26]–[28]. Finally, video recordings are examined to detect the presence of possible ''unusual'' movements of the newborn induced by the seizure [29]–[36]. Some papers describe the main existing approaches for neonatal seizure detection [37], [38]. However, in the last years, the interest in developing NSD systems increased thanks to the progress in the artificial intelligence field, and several novel methods have been introduced.

The purpose of this study is to present a survey of the main NSD systems developed in the last ten years that implement AI techniques to detect and report the temporal occurrence of neonatal seizures. Expert systems based on the analysis of EEG, ECG and video recordings are investigated, and their usefulness as support tools for the clinical teams in diagnosing neonatal seizures in NICUs is evaluated. The search was performed in June 2021 based on the Scopus database using the following keywords: 'Neonatal seizure detection'. This search identified about 1196 articles. The search was then refined considering papers published in the last ten years and using the MeSH terms: 'Automated systems / EEG monitoring / HRV / motion detection' AND 'Neonatal seizure', 'Seizure detection' AND 'NICU', 'image/video' AND 'processing' AND 'Neonatal seizure / NICU'. Some more papers previously published were also considered to be milestones in the development of NSD systems. Among all the papers, those focusing on expert systems for the automatic analysis of multi-channel EEG, ECG and video signals in NICUs were selected. Papers based on the amplitude-EEG (aEEG) and single-channel EEG were excluded. Thus, 27 papers were retained for this survey and will be summarized here.

This paper is organized as follows: Section II introduces and explains the main metrics used to report and evaluate the NSD systems' performances.

In Section III, 13 (Table 1, 2, 3) papers dealing with NSD EEG-based systems are summarized. Several studies focus on EEG signals, as they allow investigating the electrical activation of neuronal patterns that represents the main parameter for a first assessment of brain function.

Section IV presents 3 (Table 4) NSD ECG-based systems. The ECG-based analysis is of interest in NSD being routinely performed without requiring specialized training. However, identifying seizures through ECG analysis is still challenging. Thus 3 studies (Table 5) attempt to improve the NSD systems' performances by investigating the combination of EEG and ECG signals. Eight NSD video-based systems (Table 6) are presented in Section V. Video analysis is an appealing contact-less approach for seizure detection based on neonatal gestures. Finally, Sections VI, VII are devoted to discussing the NSD systems as clinical decision support tools, highlighting possible future developments of the NSD systems.

II. MATERIAL AND METHODS

A. PERFORMANCE ASSESSMENT

A standardized performance assessment framework for the seizure detection task is currently missing, and the metrics used to report NSD systems results vary in the literature [39]. Therefore, a comparison of the proposed approaches is challenging [39], [40].

The main metrics used to describe the performance of seizure detection systems can be divided into epoch-based and event-based metrics [39], [40].

The epoch-based metrics are based on the segmentation of the signals into specific time windows, called ''epochs''. This technique is a typical pre-processing step in the NSD systems. The set of analyzed epochs is divided into two classes: the seizure epochs are conventionally named ''positive'', and the non-seizure epochs as ''negative''. Seizure detection is thus a binary problem. Generally, the classifiers developed for seizure detection provide the probability that a certain epoch belongs to the positive/negative class. The performance of the systems is obtained by evaluating the decisions made by the classifier against the manual labelling made by one or more experts in neonatal EEG for each epoch.

The decision made by the classifier can be represented by the so-called confusion matrix, made of four categories: true positives (TP), i.e. epochs correctly labelled as seizures; false positives (FP), i.e. epochs incorrectly labelled as seizure; true negatives (TN) refer to correctly labelled non-seizure epochs; false negatives (FN) that are epochs incorrectly labelled as non-seizure [39].

In the literature, three main metrics are widely used: Sensitivity (SEN), Specificity (SPE) and Accuracy (ACC). SEN (1) is defined as the ratio of the number of epochs correctly labelled as seizures and the total number of seizure epochs [40]; SPE (2) is defined as the number of epochs correctly labelled as non-seizures over the total number of non-seizure epochs [39]; ACC (3) is defined as the ratio of the number of epochs correctly labelled as seizures and non-seizure and the total number of epochs.

$$\mathrm{SEN} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{1}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})} \tag{2}$$

$$\tag{3}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{(\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP})} \tag{4}$$

Most papers also report the Receiver Operator Characteristic (ROC) curves, obtained by plotting SEN against SPE (or 1-SPE). The Area Under the ROC Curve (AUC) is another crucial parameter for comparing the performances of different systems [39]. Sometimes the Precision-Recall curves are used as an alternative to ROC curves: Precision is defined as the percentage of correctly labelled seizure epochs, and Recall is the same as Sensitivity [39].

Usually, the time interval between the start and the end time instant of a seizure labelled by the experts is called "event". The main event-based metrics are:

- Good Detection Rate (GDR), which is the overall percentage of the seizure events correctly identified by the system 40]. A seizure event is correctly identified if the system detects at least one epoch during the event.
- False Discovery Rate (FDR), which is the overall percentage of the seizure events incorrectly identified by the system [40].
- False Detection per Hour (FDH), which describes the number of seizures events identified by the system in 1 h that have no overlap with the events labelled by the expert [39].
- Mean False Detection Duration (MFDD), proposed by Temko *et al.* [39], "*is assessed by averaging the duration of all false detections produced by the system at a single operating point (with a chosen threshold)*".

The existing NSD systems can be divided into patient-independent and patient-specific ones.

The patient-independent approach aims at developing systems able to detect seizures across different subjects. Usually, these systems are validated by implementing the leave one-subject out (LOSO) cross-validation: "*this way, all but one patients' data is used for training and the remaining patient's data is used for testing. This procedure is repeated until each patient has been a test subject and the mean result is reported*" [16]. This operation evaluates the systems' ability to generalize the classification: once trained on all available data, it allows achieving performances similar to those obtained by the system with an unknown dataset [16].

The patient-specific approach aims at developing systems in which the classifiers' architecture is designed for each patient. In the patient-specific models, the k-fold cross-validation and the hold out validation are usually implemented [41]. The patient-specific approach shows higher performances than the patient-independent one, but it requires pre/peri-natal data that cannot be obtained [42].

Moreover, the comparison of the existing systems is challenging as open access neonatal datasets are rarely available. The Helsinki dataset [43] is the only public one containing neonatal EEG recordings with annotations of seizures to the best of our knowledge. It collects multi-channel EEG signals from 79 full-term newborns at the NICU of the Helsinki University Central Hospital. The recordings have a mean duration of 1 h and were obtained using 19 electrodes in the so-called double-banana layout. Only critical events with a duration > 10 s are considered. Three experts separately annotated the signals: 39 out of 79 newborns have seizure activity with the unanimous consensus of the three experts.

III. NSD EEG-BASED SYSTEMS

In this section, the main NSD systems based on the analysis of the EEG recordings are summarized.

These systems aim at distinguishing the seizure epochs from the non-seizure ones investigating EEG recordings. Generally, algorithms developed for the seizure detection task provide the probability that a certain epoch belongs to the seizure / non-seizure class. Threshold values to take decisions must be defined.

Several studies proposed computer-based systems based on three approaches: the heuristic, the data-driven and the deep-learning approaches.

A. THE HEURISTIC ALGORITHMS

The heuristic algorithms are based on the empirical definition of rules, threshold values, and specific parameters obtained testing the data. Specifically, these algorithms are usually based on the morphology of the EEG traces, mimicking the visual inspection made by the clinician that searches for a variation in the signal trend from the regular background activity, usually looking for repetitive waveforms characterized by the presence of spikes or regular oscillations [37].

Liu *et al.* [11] developed a system based on autocorrelation analysis to characterize periodic activity in neonatal EEGs and distinguish seizures from background behaviour. A dataset of 12-channels EEG signals from 14 newborns was considered. The signal from each EEG channel was separately pre-processed and segmented into 30 s epochs. From the EEGs of 9 out of 14 newborns, 2-11 epochs containing seizure activity were selected; control EEG epochs were extracted from recordings of 11 newborns. The system gave: SEN $= 84\%$ and SPE $= 98\%$.

Gotman *et al.* [12] presented three methods to detect rhythmic discharges, multiple spikes, and very slow rhythmic discharges. They compared some important features about rhythmicity, power and stability of the spectrum of a specific epoch with those of an earlier epoch in the background. A dataset of 55 newborns was considered for the training step, coming from 3 centers: Montreal Children's Hospital, Montreal, Canada; Sydney Children's Hospital, Sydney, Australia; Texas Children's Hospital, Houston, Texas. The testing dataset was composed of EEG signals from 9 newborns at the Montreal Children's Hospital, Montreal, Canada; 14 newborns at the Sydney Children's Hospital, Sydney, Australia; 18 newborns at the Texas Children's Hospital, Houston, Texas [44]. The EEG signals were segmented into 10 s epochs with 75% overlap. The system gave: SEN $= 71\%$.

Deburchgraeve *et al.* [13] identified two major seizures patterns and developed two separate detection algorithms running in parallel. The first algorithm aimed at detecting high-frequency activity and the so-called "spike train seizures"; the second aimed at detecting low-frequency activity, and the so-called "oscillatory seizures". The detection

TABLE 1. NSD EEG-based systems based on the heuristic approach. adopted method, size of the datasets, epochs duration and systems' performances are summarized.

Heuristic approach				
	Method	**Dataset (newborns)**	**Epochs**	**Performance**
Liu et al. [11]	Autocorrelation analysis	14	30 s epochs	SEN[a] = 84% SPE[a] = 98%
Gotman et al. [12]	Detection of three characteristic patterns	55	10 s epochs, 75% overlap	SEN[a] = 71%
Deburchgraeve et al. [13]	Detection of two major seizures patterns	26	5 s epochs, 4 s overlap	SEN[a] = 88% FDH[a] = 0.66 h^{-1}
Navakatikyan et al. [14]	Detection of increased regularity in EEG wave sequences	66	N.A.[a]	SEN[a] = 83 – 95%

[a]SEN = Sensitivity, SPE = Specificity, FDH = False Discovery per Hour, N.A. = Not Available.

occurred if one or both algorithms detected a seizure. A dataset of EEG recordings of 26 full-term newborns at the NICUs of the Sophia Children's Hospital, Netherlands, was considered, among which 21 newborns had seizures. The signal from each EEG channel was separately pre-processed and segmented into 5 s epochs, with 4 s overlap. The combined algorithms gave: SEN = 88%, FDH = 0.66 h^{-1}.

Navakatikyan *et al.* [14] developed a neonatal detection system in which the EEG traces were divided into parallel wave sequences to mimic the manual segmentation made by an expert clinician. The algorithm aimed at detecting increased regularity in EEG wave sequences to detect seizure discharges. A dataset of multi-channel EEG from 61 newborns at Royal Brisbane and Women's Hospital, and Royal Children's Hospital, Brisbane, Australia, was considered. The recordings from 6 newborns were selected for the training, the recordings of the remaining 55 newborns were considered for the testing. The algorithm's performance was evaluated using three different methods. In the first method, the sensitivity was defined as the percentage of detected seizures marked by the specialist [14]. In the second method, the sensitivity was defined considering the duration of both seizures and events instead of their number [14]. In the third one, only the intersecting time of the detected event with a marked seizure was considered a match, or a true-positive time interval [14]. The three methods gave sensitivity values ranging between 83% and 95%.

Table 1 summarizes methods, datasets, pre-processing and performances of the studies mentioned above based on the heuristic approach.

B. THE DATA-DRIVEN ALGORITHMS

The data-driven approaches use machine-learning techniques based on the extraction of specific features to characterize the data and thus to make decisions. The features, rules and thresholds for the decision-making process are learned from the data during the training step. Generally, the EEG traces are segmented into epochs in which the signal is almost stationary, and the features are extracted from these epochs [37]. The features are defined in the frequency, time and information theory domains.

Thomas *et al.* [15] presented a real-time NSD system based on Gaussian Mixture Models (GMM) classifiers. The dataset used in this study was recorded in the NICU at Cork University Maternity Hospital, Cork, Ireland. It comprises 8-channel EEG signals from 55 full-term newborns with Hypoxic-Ischemic Encephalopathy (HIE), of which 17 had seizures. The dataset contained 267 h of EEG recordings and a total of 705 seizure events with a duration average of 3.89 min. This set was used for training and testing using LOSO cross-validation. The signal from each EEG channel was separately pre-processed and segmented into 8 s epochs using a sliding window with 50% overlap between epochs. From each epoch, 55 features were extracted, defined in time, frequency and information theory domains. The Principal Component Analysis (PCA) at 99% and the Linear Discriminant Analysis (LDA) were implemented to reduce the feature space's dimensionality and improve the classification, obtaining a subspace of 30 features. The features from each epoch and each channel were fed into Gaussian Mixture Model (GMM) classifiers. Each classifier provided the probability that a certain epoch belongs to the seizure/non-seizure class. These decisions for single channels were combined into a multi-channel decision. Then, the collar operation, which consists of joining consecutive outputs of the classifier, was implemented [39]. The system, trained on 30 features, gave: GDR = 79%, FDH = 0.5 h^{-1}, MFDD = 2 min, SPE = 93%, SEN = 76%. The test was then applied to the signals of three more patients, confirming these performances. This result highlights that the LOSO operation appropriately describes the system's ability to generalize the classification. False detections were caused by background activity, artefacts and seizure-like patterns. Missed seizures were seizures of short duration (< 1 min).

Temko *et al.* [16] replicated the above-mentioned study: they replaced the GMM classifiers with Support Vector Machine (SVM) classifiers. The proposed system correctly detected 89% of seizure events (GDR) with 1 false detection

TABLE 2. Main NSD EEG-based systems based on the data-driven approach. adopted method, size of the datasets, epochs duration and the systems' performance are summarized.

	Data-driven approach				
	Method	**Dataset (newborns)**	**Validation**	**Epochs**	**Performance**
Thomas et al. [15]	GMM	55 with HIE	LOSO [a]	8 s epochs, 50% overlap	GDR[a] = 79%, FDH[a] = 0.5 h^{-1} MFDD[a] = 2 min SPE[a] = 93% SEN[a] = 76%
Temko et al. [16]	SVM	55 with HIE	LOSO [a]	8 s epochs, 50% overlap	GDR[a] = 89%, FDH = 1 h^{-1} GDR[a] = 96%, FDH = 2 h^{-1} GDR[a] = 100%, FDH = 4 h^{-1}
Pavel et al. [17]	ANSeR	258	N.A.[a]	N.A.[a]	Distinction pathological/ healthy newborns: non-algorithm group: SEN[a] = 89.5%, SPE[a] = 89.1%, FDR[a] = 22.7% algorithm group: SEN[a] = 81.3%, SPE[a] = 84.4%, FDR[a] = 36.6% recognition of seizures hours: non-algorithm group: SEN[a] = 45.3% algorithm group: SEN[a] = 66%
Tapani et al. [18]	SVM, Autocorrelation analysis	79	LOSO [a]	32 s epochs, 28 s overlap	AUC[a] = 92% SEN[a] = 76% SPE[a] = 99%.

[a]LOSO = Leave One Subject Out, SEN = Sensitivity, SPE = Specificity, FDH = False Discovery per Hour, GDR = Good Detection Rate, MFDD = Mean False Discovery Duration, AUC = Area Under the ROC Curve, N.A. = Not Available.

in 1 h (FDH), 96% with 2 false detections and the 100% with 4 false detections.

Pavel *et al.* [17] developed and evaluated a new NSD system called ''ANSeR'' (Algorithm for Neonatal Seizure Recognition). They performed a *''multicentre, randomized, two-arm, parallel, controlled study''* [17] in eight NICUs across Ireland, Netherlands, Sweden and the UK. A dataset of 258 newborns (gestational age between 36 and 44 weeks) was considered. The newborns were split into two groups: ''the algorithm group'' and the ''non-algorithm group''. The first one was made of 128 newborns (32 of which with seizures) monitored using both cEEG and ANSeR algorithm. The other 130 newborns (38 with seizures) were assigned to the ''non-algorithm group'' and controlled with routine cEEG monitoring alone. The cEEG recordings were annotated twice by independent expert neurophysiologists. A patient was considered as a ''neonate with seizures'' if there was at least one seizure with an overlap of 30 s between the two experts' annotations (''confirmed seizure''). A time-interval of EEG recordings lasting 1 hour was defined ''seizure hour'' if there was at least one confirmed electrographic seizure within that hour. The ANSeR system displayed the seizure probability trend in real-time, and when a predefined threshold was reached, an audible and visible alarm was activated. In this study, only seizures with a duration > 30 s were considered. Although the performance in distinguishing between pathological and healthy newborns was not significantly different between the two groups (in the non-algorithm group: SEN $=$ 89.5%, SPE $=$ 89.1%, FDR $=$ 22.7%; in the algorithm group: SEN $=$ 81.3%, SPE $=$ 84.4%, FDR $=$ 36.6%), the use of ANSeR enhanced the recognition of seizures' hours. Specifically, in the non-algorithm group: SEN $=$ 45.3% and in the algorithm group: SEN $=$ 66%.

Tapani *et al.* [18] presented a NSD system based on the autocorrelation analysis that aims at highlighting the time-varying periodicity characteristic of seizure's epochs. They built a new and public database [43]. It was made of 18-channel EEG signals from 79 full-term newborns admitted in the NICU of the Helsinki University Central Hospital. The recordings have a mean duration of 1 h and were made using 19 electrodes in a double-banana layout. In this study, only critical events with a duration > 10 s were considered. Three experts annotated the dataset separately; 39 out of 79 newborns have seizure activity with unanimous consensus by the three experts. The signal from each EEG channel was separately pre-processed and segmented into 32 s epochs using a sliding window with 28 s of overlap between epochs. From each epoch, 21 features were extracted, defined in time, frequency and information theory domains, and characteristics of autocorrelation analysis. These features were fed into SVM classifiers. The single channel binary decisions were combined into a multi-channel binary decision, and then the collar operation was implemented. The system was trained and evaluated implementing the LOSO

cross-validation, and it gave: AUC = 92%; SEN = 76%; SPE = 99%.

In Table 2 a summary shows methods, datasets, validations and performances of the above-mentioned studies based on the data-driven approach.

C. THE DEEP-LEARNING ALGORITHMS

The choice of the features in data-driven methods is a crucial operation as it determines the classifiers' performances. The need for feature extraction can be overcome by introducing deep-learning algorithms that do not require hand-designed features [44]. Among the different types of existing Deep Neural Networks (DNN), the Convolutional Neural Networks (CNN) are the most used in image analysis and signal processing, where time series can be processed as images through spectrograms. Recently, CNNs and in general deep neural networks were used to analyze EEG recordings too.

Ansari *et al.* [19] compared a novel algorithm based on CNN, heuristic and data-driven algorithms. They implemented the heuristic model developed by Deburchgraeve *et al.* [13], based on the detection of ''spike-train seizures'' and ''oscillatory seizures'' and the data-driven one proposed by Thomas *et al.* [15], in which they replaced the SVM classifiers with Random Forest ones. A dataset of multi-channel EEG recordings from 48 full-term newborns was used. These patients, with assumed HIE, were admitted to the NICUS of Sophia Children's Hospital, Rotterdam, Netherlands. In the training step, 30.000 epochs of 90 s duration each were selected from 26 newborns from all available bipolar channels. For the test dataset, the recordings of the 22 remaining patients were segmented into 90 s epochs, with 60 s overlap. A CNN algorithm is proposed that does not need hand-designed features. After the training, a Random Forest classifier replaced the last five classifying layers to improve the performances. In this way, the remaining layers of the CNN work as an ''automatic feature extractor''. The overall performances highlights that the CNN-based method was more efficient than the data-driven ones (AUC = 83%; SEN = 77%; SPE = 78%; GDR = 77%; FDH = 0.90 h^{-1}), but less than the heuristic one (AUC = 88%; SEN = 77%; SPE = 90%; GDR = 77%; FDH = 0.63 h^{-1}). This could be due to the limited amount of data used for the study: indeed, CNN requires a large, varied and balanced database. Instead, the heuristic method, which aims at mimicking a human observer, is based on the clinicians' knowledge.

O'Shea *et al.* [20] presented a comparison between two feature-based and data-driven machine learning algorithms (Temko *et al.* [16], Tapani *et al.* [18]) and two novel systems based on Fully Convolutional Neural Network (FCNN): the 1D FCNN architecture, and the 2D FCNN one. In the training phase of the 1D FCNN, a single EEG channel was processed at a time, needed for the so-called ''strong labels'' that are seizure events annotated both in time and in space. The 2D FCNN architecture, which has multiple EEG channels as input, could be trained using only the ''weak labels'': in this case, the start and end time of the seizure were defined, but the spatial location was not specified. In the FCNN systems, all the EEG channels were segmented in 8 s epochs and sampled at 32 Hz. A number of 256 × 1 vectors was used as input of the 1D architecture, and 256 × N arrays as input of the 2D one, where N is the number of the EEG channels. They were trained and evaluated implementing the LOSO cross-validation on the Cork dataset [16] and tested on the Helsinki dataset [43]. It was shown that the FCNN architecture outperformed the other algorithms, confirming that the FCNN-based approaches can overcome the problem of finding appropriate features. Specifically, the 2D FCNN, based on weakly labelled data, showed the best performance (concatenated AUC = 95.6%). This architecture does not need the time consuming ''strong-labels'', therefore reducing the workload for the clinical annotators.

Tanveer *et al.* [21] developed a system based on the public dataset collected at the Helsinki University Hospital. Specifically, only the 19-channel EEG signals of the 39 newborns with seizure activity were considered. They presented three different 2D CNNs. Each model was trained and tested on one expert annotation. The EEG signals, sampled at 256 Hz, were segmented into 1 s epochs. To increase the number of seizure's samples, an overlap of 50% was used for seizure epochs. The models had as input all the EEG channels during a certain time window: 256 × 19 arrays were the input of the neural networks. To prevent overfitting, a categorical cross-entropy based loss function was introduced. It measures the distance between the output probabilities of the network and the truth values [45]. During the model training, the model weights are tuned to minimize the cross-entropy loss. The dataset was split into a training set (90%) and a validation set (10%), using the 10-cross fold validation technique. To define an overall prediction, a method based on all three model predictions was implemented that outperformed the single CNNs, giving: ACC = 96.3% and AUC = 99.3%.

Caliskan and Rencuzogullari [9] introduced a novel patient-specific NSD system based on the transfer learning technique, overcoming the CNNs' need for a large amount of data samples for training. Transfer learning is a machine learning tool that allows ''*to transfer the knowledge from the source domain to the target domain by relaxing the assumption that the training data and the test data must be independent and identically distributed*'' [46]. They presented some well-known pre-trained Deep CNNs (p-DCNNs) trained on the ImageNet database, such as AlexNet, GoogleNet, DenseNet and ResNet18. The last three layers of the networks were adapted to identify neonatal seizures. The public dataset of the Helsinki University Hospital [43] was considered. Specifically, in this study, a subset of multi-channel EEG signals from 39 epileptic newborns was considered (mean duration per patient 74 min). The signal from each EEG channel was separately pre-processed and segmented into 30 s epochs, with a 2 s sliding interval. The networks' inputs were created, converting the raw windowed EEG signals to 3 channel, 24 bits colour images. Specifically, a dataset of 106.796 images was built, 37.269 of which were labelled

TABLE 3. Main NSD EEG-based systems based on the deep-learning approach. adopted method, size of the datasets, epochs duration and the systems' performances are summarized.

	Deep-learning approach				
	Method	**Dataset (newborns)**	**Validation**	**Epochs**	**Performance**
Ansari et al. [19]	CNN + Random Forest	48 with presumed HIE	Hold out	90 s epochs, 60 s overlap	AUC[a] = 83%; SEN[a] = 77%; SPE[a] = 78%; GDR[a] = 77%; FDH[a] = 0.90 h^{-1}
O' Shea et al. [20]	2D FCNN	55 with HIE + 79	LOSO[a]	8 s epochs	concatenated AUC[a] = 95.6%
Tanveer et al. [21]	2D CNN	39	10 cross-fold	1 s epochs, 50% overlap	ACC[a] = 96.3% AUC[a] = 99.3%.
Caliskan et al. [9]	p-CNN	39	Hold out	30 s epochs, 2 s shift	Patient specific AUC[a] = 99%
O' Shea et al. [22]	fusion between the FCNN trained on the term newborns and the FCNN trained on the pre-term newborns	16 pre-term	N.A.[a]	SVM T-SDA[b]: 8 s epochs, 4 s overlap; DL T-SDA[b]: 8 s epochs, 7 s overlap;	AUC[a] = 95.4%.

[a] LOSO = Leave One Subject Out, SEN = Sensitivity, SPE = Specificity, FDH = False Discovery per Hour, GDR = Good Detection Rate, MFDD = Mean False Discovery Duration, AUC = Area Under the ROC Curve, N.A. = Not Available.

[b] SVM T-SDA = SVM based Term Seizure Detection Algorithm, DL T-SDA = Deep Learning based Term Seizure Detection Algorithm [22].

as a seizure. 50% of images were randomly selected for each patient to build the training and the test sets. To evaluate the classification performance, Caliskan *et al.* compared the results obtained using AlexNet, GoogleNet, Densenet and ResNet18 and built a conventional CNN with 6 layer depth. The accuracy and the AUC curve for all the 39 newborns showed that the DenseNet-based method has the best performance (mean AUC = 99%). A statistical analysis highlighted that all the p-DCNNs have better performance than the CNN, thus allowing the detection of neonatal seizures overcoming the limited dimension of the training data set.

Recently, O'Shea *et al.* [22] presented a novel system based on deep learning algorithms to detect neonatal seizures, focusing on preterm infants. In this study, they considered two of the above-mentioned algorithms, trained on datasets of full-term patients: the SVM-based one [16] and the FCNN-based one [20]. These approaches were tested on a dataset of 8-channel EEG of 16 preterm newborns (gestational age < 32 weeks) admitted to the NICUs of the Cork University Maternity Hospital, Ireland (total duration 575 h). Six out of 16 patients had seizure events, and the remaining 10 were control patients. The SVM-based algorithm (called "SVM T-SDA", where T stands for full-term newborns) gave AUC = 88.3%, and the FCNN-based one (called "DL T-SDA", where T stands for full-term newborns) gave AUC = 93.3%. Then, they retrained the algorithms on a dataset of 14-channel EEG recordings from 17 preterm newborns (gestational age < 32 weeks) admitted in the NICUs of Parma University Hospital, Italy (mean duration 1 h and 19 min). These algorithms were tested on the Cork preterm dataset with the following results: AUC = 89.7% with SVM P-SDA (where P stands for preterm newborns), and AUC = 93.5% with DL P-SDA (where P stands for preterm newborns). The gestational age (GA) strongly influences the morphology of the EEG signal. Therefore O' Shea *et al.* divided the training and test sets into 3 groups according to the GA of the newborns and developed SVM-based and FCNN-based specific algorithms for each GA group. Finally, they evaluated the fusion between the FCNN trained on the term newborns and the FCNN trained on the preterm newborns, divided into GA groups. The system obtained by the fusion of classifiers gave AUC = 95.4%.

Table 3 summarizes methods, datasets, validation and performance of the mentioned studies based on the deep-learning approach.

IV. NSD ECG-BASED SYSTEMS

This section describes the main NSD systems based on ECG analysis. Indeed, several studies suggest that neonatal seizures strongly influence cardiocirculatory activity. Goldberg *et al.* [47] considered a dataset of ECG signals from 9 paralyzed newborns, finding changes in ECG rhythmicity, heart rate, blood pressure and oxygenation. Therefore, they concluded that these fluctuations could be used as indicators of critical events. Similarly, Watanabe *et al.* [48] observed heart rate and respiratory rate changes during seizures in 215 newborns.

Although many pacemaker tissues exist that control heart contraction, heart rate and cardiac rhythmicity are largely regulated by the Autonomic Nervous System (ANS). Indeed, the sympathetic and parasympathetic nervous systems stimulate the heart by increasing and decreasing heart rate. Therefore, the evaluation of changes in inter-beat time intervals (Heart Rate Variability – HRV) can provide important information about the effects that seizures have on ANS's functions. For example, Bersani *et al.* [49] suggested

the HRV analysis as "*a possible marker of brain damage*" in the case of HIE. In order to perform the HRV analysis, the ECGs signals are usually pre-processed through a denoising procedure that preserves clinically relevant information. Afterwards, signals are segmented into individual beats [50], highlighting the QRS complexes.

Generally, the HRV spectrum is divided into spectral bands, and each of these bands is associated with different activities of the sympathetic and parasympathetic nervous systems [51]. In the literature, several ranges of frequency bands have been defined. In fact, these values strongly depend on the health of the ANS, age, and the patient's physiological conditions. The HRV signal can be obtained from the ECG, and its characteristics can be analyzed with different algorithms in the time or frequency domain. Several researchers have recently focused their studies on HRV analysis to detect seizures. ECG signal is routinely performed, and its recording is easier and less invasive than EEG [25].

Malarvili *et al.* [23] investigated the HRV signals by evaluating the Time-Frequency Distribution (TFD). The TFD is a bidimensional function that describes the instantaneous frequency of the signal in the combined time-frequency domain. This study considered a dataset of one-channel ECG signal of 5 newborns collected at the Royal Children's Hospital, Brisbane, Australia. This dataset consisted of 6 seizure events and 4 non-seizure events of 64 s each from 5 different newborns. All ECG traces were processed to extract the HRV signal. To analyse the HRV signal and recognize seizure events, the selected features were the first and the second conditional moments of the three spectral components: Low Frequency (0.03–0.07 Hz), Mid Frequency (0.07–0.15 Hz) and High Frequency (0.15–0.6 Hz). The LOSO cross-validation was implemented. By evaluating the overall performances, it was shown that the first conditional moment allows discriminating critical events from non-critical ones at low frequencies. This suggests that neonatal seizures mainly affect the HRV components in the low-frequency band, which are attributed to sympathetic activity by the authors (SEN = 83.33%, SPE = 100%).

Greene *et al.* [24] introduced a NSD ECG-based system using a Linear Discriminant (LD) classifier. A dataset of 8 ECG recordings from 7 full-term newborns admitted in NICU for HIE was considered. It was made of 520 seizure events (mean duration 3.86 min). Seven out of 8 ECG signals were recorded in the NICUs of the Unified Maternity Hospitals in Cork, Ireland, the remaining one was recorded in the NICU of Kings College Hospital, London. The ECG signals were segmented into 60 s epochs. An epoch was defined as a seizure epoch if 50% of its duration was interested in the critical activity. The R peaks were detected using an appropriate QRS detection algorithm, and features describing RR intervals' properties in time, frequency and information theory domains were extracted. These features were fed into a supervised LD classifier that looks for the best linear combination of features to distinguish seizure and non-seizure classes. Greene *et al.* developed a patient-specific and a patient-independent system. The patient-specific approach showed better results than the patient-independent one. It was evaluated by implementing a ten-fold cross-validation on each record. Specifically, each record was iteratively and randomly split into 10 folds, and 9 of these folds were used to train the classifier; the remaining one was used to test the classification. The obtained results were averaged, and the classifier's performance for each patient was evaluated. The patient-specific system gave: ACC = 66.04%; SEN = 75.52%; SPE = 57.70%. The patient-independent systems were validated implementing the LOSO operation, giving ACC = 61.80%; SEN = 78%; SPE = 51.75%. The patient-specific approach shows higher performances than the patient-independent one: however, it requires patient-specific data that cannot be obtained before the baby is born [42].

Doyle *et al.* [25] investigated the utility of HRV analysis to develop a NSD based on a SVM classifier. The Cork dataset was considered [16]. Specifically, only the recordings of 14 out of 17 newborns were considered with a total duration of 207.86 h. They are characterized by the presence of 697 seizure events (mean duration 3.83 min). Firstly, the HRV signal was extracted from the ECGs and segmented into 60 s epochs. Then, from each epoch, 62 features defined in time and frequency domains were extracted. These features were fed into two SVM classifiers: one characterized by a linear kernel and the other by a non-linear kernel. The two systems were evaluated by implementing the LOSO cross-validation. Both the systems gave: mean AUC = 60% and mean SEN = 60%. Later, the feature selection operation was implemented to select the most suitable features to discriminate between seizure and non-seizures epochs and prevent redundancy problems. Therefore, the non-linear SVM system was re-trained with a subset of 35 features, giving a lower value of AUC (55%): in fact, some features relevant for some patients were removed by the selection operation.

Table 4 summarizes methods, datasets, pre-processing and performances of the mentioned studies based on the ECG analysis.

A. NSD SYSTEMS BASED ON THE COMBINATION OF EEG AND ECG

Few studies in the literature attempted to improve the NSD systems' performances by investigating the combination of EEG and ECG signals.

Greene *et al.* [26] considered two methods for combining ECG and EEG signals: the early integration (EI) and the late integration (LI). The first one is based on a single feature vector, obtained concatenating the EEG and ECG features and fed them into a classifier. The late integration made use of one classifier for each signal: two output probabilities are combined to define an overall probability of seizure. A dataset of 12 recordings from 10 full-term newborns admitted to NICU for HIE was considered. Ten out of 12 recordings were made in the Unified Maternity Hospitals in Cork, Ireland (sampled at 256 Hz); the remaining one was made at Kings

TABLE 4. Main NSD ECG-based systems. method, size of the datasets, validation methods and the systems' performances are summarized.

	ECG-based systems				
	Method	**Dataset (newborns)**	**Validation**	**Epochs**	**Performances**
Malarvili et al. [23]	Heuristic TFD	5	LOSO [a]	64 s epochs	SEN[a] = 83.33% SPE[a] = 100%
Greene et al. [24]	LD	7	10-fold	60 s epochs	patient-specific: ACC[a] = 66.04% SEN[a] = 75.52% SPE[a] = 57.70%;
			LOSO [a]		patient-independent ACC[a] = 61.80% SEN[a] = 78% SPE[a] = 51.75%
Doyle et al. [25]	SVM	55 with HIE	LOSO [a]	60 s epochs	mean AUC[a] = 60% mean SEN[a] = 60%

[a]LOSO = Leave One Subject Out, SEN = Sensitivity, SPE = Specificity, FDH = False Discovery per Hour, GDR = Good Detection Rate, MFDD = Mean False Discovery Duration, AUC = Area Under the ROC Curve, ACC = Accuracy.

College Hospital, London (sampled at 200 Hz). Each recording was composed of multi-channel EEGs and one-channel ECGs. The EEG signals were annotated by expert clinicians that detected 633 seizure events (mean duration 4.60 min). The ECG signals were segmented into non-overlapping 60 s epochs as described in [24]. A total of 16.384 samples was obtained: 15.360 for a record sampled at 256 Hz and 12.000 for a record sampled at 200 Hz. The R peaks were detected from each epoch, and 6 features describing RR intervals' properties in time, frequency and information theory domains were extracted. The signal from each EEG channel was separately pre-processed and segmented into non-overlapping 8 s epochs (2048 samples at 256 Hz). From each epoch, 6 features defined in time, frequency and information theory domains were extracted. Each feature vector from different channels was concatenated to create a single ''super feature vector''. A sorting function was implemented to remove information about the spatial location of the seizure by distinguishing feature values of ''channel involved in a seizure'' and ''channels not-involved''. The ECG signal was segmented into non-overlapping 60 s epochs and the EEG signal into non-overlapping 8 s epochs: to combine the information of the two signals, the ECG frame rate was matched to the EEG frame rate by interpolation. The EI and LI frameworks were developed using LD classifiers, and both approaches were evaluated in patient-specific and patient-independent configurations. The 10-fold cross-validation was implemented for the patient-specific classifier, while for the patient-independent one, the LOSO cross-validation was implemented. The performances were evaluated by averaging the results across recordings. In the patient-specific framework the EI approach gave: GDR = 95.82%; FDR = 11.23%; ACC = 86.32%; SEN = 76.37%; SPE = 88.77%, while the LI approach gave: GDR = 97.52%; FDR = 13.18%; ACC = 84.66%; SEN = 74.08%; SPE = 86.82%. In the patient-independent framework the EI approach gave: GDR = 81.44%; FDR = 28.57%; ACC = 71.51%; SEN = 71.73%; SPE = 71.43% while the LI approach gave: GDR = 81.27%; FDR = 33.05%; ACC = 68.89%; SEN = 74.39%; SPE = 66.95%. The patient-specific approach shows higher performances than the patient-independent one. However, as mentioned above, the patient-specific approach is not suitable for neonatal application [42]. The patient-independent performances result appealing, but their clinical utility is limited by the high FDR.

Based on the above-mentioned study [26], Mesbah *et al.* [27] investigated the early integration (''feature fusion'') and the late integration (''classifier fusion''), introducing some changes and novelties. They considered a different dataset and segmented the EEG and ECG signal into epochs of different duration. Moreover, they selected different sets of features to analyse the signals and considered different types of classifiers. The dataset was made of EEG-ECG recordings from 8 full-term newborns admitted to the Royal Brisbane Hospital, Brisbane, Australia. A paediatric neurologist annotated the EEG, hence 13 seizure events were identified (mean duration 2.54 min). Then, the ECG signals were segmented into 64 s epochs. Twenty-one seizure epochs and 13 non-seizure epochs were randomly selected and considered. The EEG signals were segmented into 64 s epochs too, and each epoch was further divided into non-overlapping 12.8 s epochs. Moreover, Mesbah *et al.* introduced some non-stationary features: in ECG analysis, they considered features defined in time and in time-frequency domains; in EEG analysis, they considered features defined in time, frequency, time–frequency, and time-scale domains. For both signals, the feature selection operation was implemented to prevent redundancy problems, and the LOSO cross-validation was implemented. The feature fusion framework, based on 1-NN classifier trained on the selected features, gave: SEN = 95.20%; SPE = 88.60%. The classifier fusion framework, based on linear classifier (ECG)

TABLE 5. Main NSD systems based on the combination between eeg and ecg. method, size of the datasets, validation methods and the systems' performances are summarized.

Systems based on combination between EEG and ECG						
	Method	**Dataset (newborns)**	**Validation**	**Epochs**	**Performances**	
Greene et al. [26]	Early and Late fusions with LD classifiers	10	10-fold LOSO[a]	60 s epochs	EI[a] patient-specific GDR[a] = 95.82% FDR[a] = 11.23% ACC[a] = 86.32% SEN[a] = 76.37% SPE[a] = 88.77%; LI[a] patient-specific GDR[a] = 97.52% FDR[a] = 13.18% ACC[a] = 84.66% SEN[a] = 74.08% SPE[a] = 86.82%;	EI[a] patient-independent GDR[a] = 81.44% FDR[a] = 28.57% ACC[a] = 71.51% SEN[a] = 71.73% SPE[a] = 71.43%; LI[a] patient-independent GDR[a] = 81.27% FDR[a] = 33.05% ACC[a] = 68.89% SEN[a] = 74.39% SPE[a] = 66.95%
Mesbah et al. [27]	Early and Late fusion with 1-NN and linear classifiers	8	LOSO[a]	64 s epochs	EI[a] with 1-NN SEN[a] = 95.20 SPE[a] = 88.60; LI[a] with linear classifier and 1-NN classifier SEN[a] = 95.20 SPE[a] = 94.30	
Temko et al. [28]	SVM	38	LOSO[a]	60 s epochs	AUC[a] = 86.8%	

[a]LOSO = Leave One Subject Out, SEN = Sensitivity, SPE = Specificity, FDH = False Discovery per Hour, GDR = Good Detection Rate, MFDD = Mean False Discovery Duration, AUC = Area Under the ROC Curve, ACC = Accuracy, EI = early integration, LI = late integration.

and 1-NN classifier (EEG) trained on the selected features, gave: SEN = 95.20%; SPE = 94.30%.

Temko *et al.* [28] investigated automated multimodal prediction of outcome in newborns with HIE, based on features extracted from clinical analysis, EEG and ECG signals. A dataset of video-multichannel-EEGs and ECGs from 38 full-term newborns admitted in the NICUs of Cork University Maternity Hospital was considered. From these recordings, 1-h segments per patient free from visual artefacts were selected. The 1-h segments from each EEG channel were further segmented into 60 s non-overlapping epochs. A set of 57 features defined in time, frequency, and information theory domains describing the brain symmetry were extracted from each epoch. To combine the information across channels, the mean value of the features was considered. The R peaks were detected using an appropriate QRS detection algorithm, defining the HRV signal. As for the EEG signal, the 1-h recording was segmented into 60 s non-overlapping epochs. A set of 60 features defined in time, frequency, and information theory domains was extracted from each epoch. Regarding the clinical features, the Apgar score, the initial pH and the Base deficit were analyzed. The EEG and HR features were synchronized, and the clinical features were characterized by one value per patient that was replicated for each epoch. The SVM classifier and the LOSO cross-validation were implemented. The feature selection operation, named Recursive Feature Elimination (RFE), was applied in each iteration to the EEG and HRV sets. The classifier trained on 12 features from EEG, HR and Apgar showed the best performances, giving AUC = 86.8%.

Table 5 summarizes methods, datasets, pre-processing and performances of the mentioned studies based on the combined ECG-EEG analysis.

V. NSD VIDEO-BASED SYSTEMS

This section aims at outlining the most significant papers about NSD video-based systems. The newborns' movements can provide crucial information about their physio-pathological state. The analysis of movement characteristics and properties can be useful for a timely diagnosis of neurological and neurodevelopmental disorders.

Over the years, many approaches were proposed to evaluate the newborns' movements involving their body and head through video analysis. Indeed "*limbs and head are the infant body parts mostly affected by seizure-caused motion*" [29]. At present, the detection and classification of neonatal seizures based on video recordings cannot replace EEG analysis but allows creating a contact-less seizure detection system as a support to the clinical decision [52]. Indeed, Malone *et al.* [53] showed that "*health care professionals have difficulty in discriminating between neonatal seizure and non-seizure movements*" analyzing video recordings of the movements only. They considered a dataset made of video clips of 11 newborns with EEG-confirmed seizures (clonic and subtle), and 9 newborns with random movements. These videos were recorded at King's College Hospital London, United Kingdom, and at Cork University Maternity Hospital,

Ireland. The recordings were examined by 137 health care professionals: the seizure events were correctly identified only by 20% to 50% of the professionals (41% on average). Therefore, developing additional systems based on video analysis capable of identifying even inconspicuous movements would be useful.

Three main approaches in newborns motion detection can be found:

- frame differencing that aims at highlighting the patient's movement by evaluating the difference between consecutive video-frames.
- optical flow, based on the relative movement between the observer and the scene. It allows computing the speed vector associated with each pixel of the frame.
- tracking methods, based on the selection of regions of interest and their tracking in a sequence of frames [38], [33].

Ntonfo *et al.* [29] developed a system that aimed at distinguishing the clonic (''periodic seizures over short time intervals'') and myoclonic (''seizures that are brief, rapid, single or arrhythmic repetitive jerks'') seizures. This system applies the optical flow technique to define the maximum optical flow vector amplitude (MIMP) to detect the part of the body interested by a strong and pathological movement. Around the MIMP, the Region of Interest (ROI) was selected on which the subsequent analysis was focused. Moreover, the ROI was tracked in the image sequences using the template matching technique that is based on the Mean Absolute Difference (MAD) similarity measure. To characterize the motion, each RGB frame of the sequences was converted into the greyscale, and then the frame difference between two consecutive frames was computed. In this way, a sequence of frames in which the movement was highlighted was obtained. These frames were converted into a binary scale by selecting a suitable threshold [1]. Taking into account that the bright binary pixels were related to the moving body parts, the average luminosity motion signal was defined by evaluating the average number of white pixels in each frame. To distinguish the clonic and myoclonic seizures, the periodicity of the average luminosity motion signal was evaluated by defining the Broadening Factor, ''*an indicator of how impulsive the entire movement is*'' [29], and the maximum distance between consecutive pairs of zeros of the average luminosity motion signal.

Later, Ntonfo *et al.* [30] presented another system that distinguished between clonic and myoclonic seizures by analyzing gesture trajectories. They defined the MIMPs as in [29] and tracked them in the frame sequences using the template matching technique. The final movement trajectories were formed ''*by joining all the points given by consecutive MIMP coordinates in a sequential manner*'' [30]. Some trajectory features were defined and analyzed by clustering to create groups of movements with similar characteristics. A cluster with high cardinality highlights a repetitive movement. A dataset of 2 recordings from 2 newborns at the Department of Gynecology, Obstetric, and Neonatal Sciences of the University of Parma, Italy, was considered. One patient was affected by clonic seizures, the other one by random movements.

Pisani *et al.* [31] developed and validated a system that aims at identifying clonic seizures from other movements and noise. A dataset of 23 video recordings was analyzed from 12 full-term newborns admitted to the NICUs of Parma University Hospital. These videos, containing 78 seizures of clonic type, were analyzed by visual inspection. 502 noise events, with a total duration of 04:44:08 h (mean duration 00:00:34) and 668 motor events with a total duration of 04:15:22 h (mean duration 00:00:23) were identified. Each frame of the video recordings was converted into a grey scale, and the frame difference between two consecutive frames was computed. The average luminosity motion signal was obtained as described in [29]. The periodicity of the signal was evaluated by defining the Normalized AutoCorrelation Function (NACF) and the Cumulative Mean Normalized Difference Function (CMNDF) [1]. The periodicity was analyzed considering: ''*disjoint consecutive frame windows, where each window lasts 10 s*; *two interlaced windows, with 50% overlapping; three consecutive interlaced windows, with 50% overlap between consecutive pairs*'' [31]. The described procedure was also applied to 6 video recordings of 5 healthy newborns (total duration 04:34:29 h, mean duration 00:45:45 h). In these videos, 426 motor events (total duration 01:19:02 h, mean duration 00:00:11 h) and 99 noise events (total duration 00:14:00 h, mean duration 00:00:08 h) were detected. The system developed using two interlaced windows gave the best performances in detecting clonic seizures: AUC = 79.6%; SEN = 71%; SPE = 69%. In detecting motor and environmental phenomena, the system developed using three interlaced windows gave: SPE = 97%.

Cattani *et al.* [32] developed a system based on the average luminosity motion signal analysis, obtained as described in [30]. To study the periodicity of the signal and detect the clonic seizures, the Maximum Likelihood criterion was adopted. The motion signals were acquired through multiple cameras, and depth sensors were considered. Specifically, three video cameras were set up in the NICUs of the University Hospital of Parma: two cameras recorded the newborn from the front and the side and the third camera was attached to the cot to focus on the face. A dataset of 4 recordings of a newborn with the three cameras was collected. The first 2 videos were characterized by the presence of pathological movements related to clonic seizures, while physiological movements characterized the remaining 2 videos. The analysis was performed considering two 10 s interlaced windows with 50% overlapping. The system gave: SEN $\approx$ 90%, SPE $\approx$ 90%, outperforming the systems based on one or two cameras only.

To ease the distinction between seizure and non-seizure events, Karayiannis *et al.* [33] presented a system performing a post-seizure analysis based on newborns' motor activity. They defined the temporal motion strength through the spatiotemporal decomposition of an image sequence. In this

way, a specific subband of the decomposed image sequence was identified, detecting motion between consecutive frames. The subband was processed by applying median filters and segmented, implementing an adaptive version of the k-means algorithm (k = 3). The white areas in the segmented frames display the moving parts of the body. The temporal motion strength was defined by evaluating the average of white areas in consecutive frames. Moreover, the temporal motor activity signal was defined by tracking anatomical sites, such as right leg, left hand, and right hand, through a modified version of the Kanade-Lucas-Thomasi (KLT) algorithm. These anatomical sites were projected to both the horizontal and the vertical axes across the frames. Four video recordings from the Clinical Research Centers for Neonatal Seizures (CRCNS), Houston, TX, were considered: 2 out of 4 were characterized by the presence of myoclonic seizures, the other 2 by the presence of focal clonic seizures.

Later, Karayiannis and Tao [34], [35] proposed an improved method to extract the temporal motion strength signal. As above, frame differencing was implemented to highlight the moving parts of the body, but the resulting frames were segmented with the vector form of the k-means algorithm (k = 4 clusters of vectors). In this way, the number of vectors made of background pixels erroneously classified as moving body parts was reduced.

Moreover, Karayiannis *et al.* [36] developed a Feed-Forward Neural Network (FFNN) to classify and recognize myoclonic, focal clonic seizures and physiological motion. Two different approaches were compared regarding the temporal motion strength signal: the one described in [34] and another based on the optical flow [54]. In the latter, the velocity vectors associated with each pixel of the frame were defined, and the area containing all the pixels with a speed greater than a defined threshold was computed. Similarly, to define the motor activity signal, the predictive block matching technique [55], [56] was compared to methods involving other models of blocks (''robust motion trackers'') [57]. In these studies, a block of pixels of predefined dimension (''reference block'') was defined in the first frame of the considered sequence around the anatomical sites of interest. The location of the block was then predicted, looking for the most similar block in subsequent frames using Kalman filtering. A dataset of 240 videos from 43 newborns was considered. These recordings were made at the CRCNS. Specifically, 80 out of 240 records were characterized by the presence of myoclonic seizures, 80 by focal clonic seizures and the remaining 80 by physiological movements. A preliminary comparison between different techniques highlighted that the most reliable approach to estimate the motion strength signal was the one based on the optical flow, while the one based on the robust motion trackers was found the best to estimate the motor activity signal. The dataset was split into training (50%) and test (50%) sets. Firstly, features such as the variance of time intervals, energy ratio, maximum spike duration and the number of spikes were extracted from the motion strength signal. These features were fed into the FFNN, that gave SEN > 95%, SPE > 95% on the training set, and SEN > 90%, SPE > 95% on the test set. Then, features such as energy ratio, maximum spike duration, the variance of the time intervals between the extrema and number of extrema were extracted by the motor activity signal. These features were fed into the FFNN, that gave: SEN > 90%, SPE > 90% on the training set, and SEN < 90%, SPE > 90% on the test set. Finally, the features extracted from both motion strength and motor activity signals were fed into the FFNN, which gave SEN > 90%, SPE > 95% on the training set, and SEN < 90%, SPE > 90% on the test set.

Table 6 summarizes methods, datasets, pre-processing and performances of the mentioned studies based on the video analysis.

VI. DISCUSSION

This paper presents a survey of the expert systems developed in the last ten years for Neonatal Seizures Detection in NICUs.

Over the years, many approaches were proposed to automatically detect seizure activity in adults and children, investigating EEG and other physiological signals [58]. In fact, epilepsy, which is a neurological disease, can affect spontaneous electrical cerebral activity. Other signals that are under cerebral control can also provide information about the state of the brain [58]. For example, HR alterations commonly occur in adults with seizures [58], making heart rate analysis crucial for seizure detection. HR can be achieved through ECG or photoplethysmography (PPG). Also respiratory activity is relevant and can help in seizure detection [59]. In fact, seizure activity can frequently alter the normal and physiological respiratory rate. Irregular ventilation during seizures can be investigated by monitoring blood oxygenation: several studies showed increased cerebral oxygen saturation before seizures that can be efficiently measured using near-infrared spectroscopy (NIRS) techniques. Furthermore, seizure-related changes in sympathetic activity can be evaluated by investigating skin conductance (SC) modulation, or generally electrodermal activity (EDA). Motor manifestations of seizures can be analysed by examining the electromyographic (EMG) signal and using accelerometer-based (ACM) devices. While the applications of these methodologies to the newborn are very scarce, there are many studies and results regarding the adult and the child. The reasons concern not only the peculiar physiological characteristics of the newborn, as already pointed out above [60], but also the difficulty of applying and using adequate sensors in NICUs. Even though a survey concerning adult and child monitoring would be interesting, it is out of the scope of this work. We suggest survey papers [58], [59], and [61] to the interested reader. Thus, most of the expert systems developed for Neonatal Seizures Detection in NICUs summarized in this review are based on EEG, ECG and video analysis, as these signals are usually recorded and monitored in NICUs. Several studies investigated how the seizures occurrence

TABLE 6. Main NSD video-based systems method, size of the datasets, validation methods and the systems' performances are summarized.

Video-based systems						
	Method	**Dataset (newborns)**	**Validation**	**Epochs**	**Performances**	**Notes**
Ntonfo et al. [29]	• Optical flow, • Template matching	N.A. [a]	N.A. [a]	N.A. [a]	N.A. [a]	Distinction between clonic and myoclonic seizures
Ntonfo et al. [30]	• Optical flow, • Template matching, • Clustering gesture trajectories	2	Qualitative evaluation	N.A. [a]	N.A. [a]	Distinction between clonic and myoclonic seizures
Pisani et al. [31]	• Frame differencing	12	Binary statistical analysis	10 s epochs, 50% overlap	detecting clonic seizures: AUC[a] = 79.6% SEN[a] = 71% SPE[a] = 69%; detecting motor and environmental phenomena SPE[a] = 97%	Distinction between clonic seizures and random movements
Cattani et al. [32]	• Frame differencing (3 cameras)	1	N.A. [a]	10 s epochs, 50% overlap	SEN[a] ≈ 90% SPE[a] ≈ 90%	Detection of clonic seizures
Karayiannis et al. [33]	• Frame differencing, • KLT	4	Qualitative evaluation	N.A. [a]	N.A. [a]	Distinction between myoclonic and focal clonic seizures
Karayiannis et al. [34], [35]	• Frame differencing,	N.A.	N.A. [a]	N.A. [a]	N.A. [a]	Distinction between myoclonic and focal clonic seizures
Karayiannis et al. [36]	• FFNN, • Frame differencing, • Optical flow, • block matching, • robust motion trackers	43	Hold out	N.A. [a]	training set SEN[a] > 90% SPE[a] > 95% test set SEN[a] < 90% SPE[a] > 90%	Distinction between myoclonic and focal clonic seizures

[a] N.A. = Not Available, SEN = Sensitivity, SPE = Specificity, ACC = Accuracy.

affects the electrophysiological signals. Specifically, EEG is usually investigated to identify the presence of irregularities or characteristic trends due to seizures [11]–[22]; ECG is analyzed to evaluate the heart rate variability due to changes in the cardiovascular system during or close to ictal events [23]–[25], while video recordings are examined to detect the presence of possible ''unusual'' movements of the newborn induced by the seizure [29]–[36]. Only one recent study investigating the NIRS technique applied to newborns exists [62]. This paper summarizes and improves previous studies, highlighting the clinical relevance of the combined analysis of aEEG and NIRS signals. Indeed, seizures are characterized by a drop in cerebral oxygen saturation due to an increase in cerebral metabolic demand.

The lack of complete public datasets of neonatal seizures makes the implementation of an automated seizure detector in newborns more difficult. The availability of public electrophysiological and video signals datasets is indeed crucial for the development and evaluation of computer-based systems for the targeted task. To the best of our knowledge, the Helsinki dataset [43] is the only public one containing neonatal EEG recordings with multi-expert annotations of seizures. The majority of the NSD systems proposed in the literature are evaluated on private datasets only, making the comparison between the existing approaches unachievable.

Furthermore, this comparison is still challenging because the metrics used to report the results of the NSD systems vary in the literature [39]. Therefore, a standard set of metrics would be advisable to evaluate the usefulness and the efficiency of the developed techniques for the neonatal seizure detection task.

Another crucial issue concerns the validation methods applied to evaluate the generalization ability or precision of the proposed methods in the seizure detection task. As mentioned in this paper, the existing NSD systems can be divided into patient-independent, and patient-specific ones. The patient-independent approach, which aims at developing systems able to detect seizures across different subjects, is usually validated through the LOSO operation.

Instead, in the patient-specific approach, aiming at developing systems in which the classifiers' architecture is designed for each patient, the k-fold cross-validation and the hold out validation [41] are usually implemented. Although the patient-specific method is appealing, it requires specific electrophysiological data that cannot be obtained before birth [42]. Thus, the patient-independent approach is more advisable in the NSD task. In particular, the LOSO is preferable as it allows a good evaluation of the systems' ability to generalize the classification in small datasets [16]. Instead, if datasets are not quite large [63] other validation approaches, such as k-fold o hold out, tend to overestimate the performances of the systems [41].

The NSD EEG-based systems aim at detecting the neonatal seizure events analysing and characterizing the EEG recordings. These systems show better performance than those based on other electrophysiological signals: indeed, the EEG is the most appropriate diagnostic technique to detect neonatal epileptic seizures as it allows investigating the electrical activation of neuronal patterns. As shown in Table 3 the deep-learning—based approaches lead to a remarkable improvement in NSD performances, however these systems require larger datasets than heuristic and machine-learning—based methods. The interest in studying and analysing ECG recordings for the NSD task is growing more and more because the ECG signal is routinely performed, and its recording is easier and less invasive than the EEG one [25]. However, the performances of ECG-based patient-independent systems are not so appealing and at present they cannot replace the EEG-based systems.

Despite that, ECG analysis, and in particular the related HRV analysis, seems to be a promising marker of brain damage. Bersani *et al.* [49] presented a systematic review that highlights a possible relationship between HIE and abnormal HRV values, suggesting that HRV analysis may represent a valid alternative to EEG to detect the most common etiology of neonatal seizures. Statello *et al.* [64] also analyzed the behaviour of the sympathetic and the parasympathetic systems during neonatal seizures by investigating HRV indices. They found that the vagal-mediated HRV signal in newborns with seizures is lower than in healthy newborns and that a short-term increase in vagal-mediated HRV characterizes seizures.

To increase the performances of the developed NSD systems, some studies in the literature investigated the combination of EEG and ECG signals [26]–[28]. As a result, this combination has led to a more robust system for neonatal seizure detection than a system based on the ECG signal only. To the best of our knowledge, for ECG/HRV analysis, no deep-learning method was proposed in the literature for neonatal seizure detection. Considering the improvement obtained by DL techniques on EEG, these methods should also be evaluated on NSD experiments with ECG signals. Finally, to improve performances of ECG-based NSD, more efforts should be made in the study of brain-heart interactions [65], [66] during ictal events in newborns. Indeed, the link between the cardio-regulatory system and neonatal seizures is not yet fully understood. Some findings [27], [64] suggested that seizures can directly or indirectly alter the cardio-regulatory system. However, evidence about mechanisms occurring during these events and the corresponding etiology are still missing or incomplete. Identifying and measuring them might allow the use of more specific and useful features for the neonatal seizure detection task through ECG signals.

Computer-based systems based on video recordings analysis can be useful to characterize the newborns' movements and thus their physio-pathological state. The systems described in this work apply different and interesting approaches to the analysis of video recordings. Most of these papers aim to detect and distinguish clonic and myoclonic seizures characterized by intense clinical manifestations. However, up to 70% of all neonatal seizures are characterized by poor clinical manifestations [3], [6]. These seizures are called subtle and are characterized by eye deviations, repetitive opening and closing of the eyelids, sucking, oral-buccal-lingual movements, ''swimming'' or ''pedalling'' movements [6], [67]. As very few clinical correlates exist, the subtle seizures can be confused with normal neonatal behaviour [3], [6]. Malone *et al.* [53] highlighted the health care professionals' difficulty in identifying subtle seizures: while clonic seizures were correctly identified by about two out of three professionals on average (with identification of single cases ranging from 36% to 95%), subtle seizures were recognized only by an average of one out of three professionals (with individual detection between 20% and 50% at best). Therefore, developing additional systems based on video analysis, capable of identifying even inconspicuous movements and automating the semiology of facial expressions, would be useful. However, it is difficult to automatically recognize and track the newborns' faces in the NICUs, as electrodes or cannulas often cover part of the face, cameras may be inappropriately placed, and lighting may be poor. Furthermore, the majority of the systems mentioned in this study are based on video recordings with a single video-camera. Using more cameras could improve the systems' performances allowing a view of the newborn from

different perspectives and ensuring an adequate coverage of the observed scene, as already evaluated by. Karayiannis *et al.* [33]. They found that the use of multiple cameras improved performances in detecting clonic seizures [33].

The papers summarized here do not provide information about etiologies of the detected seizures. However, some of them [68], [69] pointed out that the cause of seizure events could be identified by analyzing seizure events themselves. Therefore, NSD systems able to automatically characterize the etiologies investigating the available electrophysiological and clinical signals could be an additional support tool to clinicians. Indeed, identifying etiologies is crucial to determine specific pharmacological treatments and subsequent prognoses.

Another crucial aspect concerns the methods of displaying the information obtained with the NSD systems. Few papers focus on developing an appropriate user interface and evaluate how it could affect seizure detection in a clinical environment [17], [70]. Temko *et al.* [70] investigated different ways to provide the output of a NSD system to clinicians, showing that a viable system interface is fundamental to assess the real usefulness of the NSD systems as support tools for the medical staff in the diagnosis of neonatal seizures in NICUs. Moreover, evaluating the seizure detection delay in NSD systems is crucial for understanding their clinical usefulness. The seizure detection delay is defined as ''the time delay between the seizure detected by the algorithm and the seizure onset marked by an expert'' [71]. This delay is heavily influenced by the time duration of the epochs in which the signal is segmented and the processing time required to run the algorithms [20], [50]. Specifically, the processing time is given by the algorithms' complexity and the computational performances. According to the seizure detection delay, two different types of expert systems applications are defined: online and offline. The first one ensures a timely, effective and efficient clinical intervention during the acquisition of electrophysiological and clinical signals. [10]. In fact, neonatal seizures may lead to acute neurological impairment and neonatal death. Thus they should be treated as soon as possible [72]. The offline analysis is useful as it marks out the seizure epochs. Thus the neurologists can examine the detected epochs necessary for a correct diagnosis [10]. Only few papers summarized in this survey explicitly define their algorithms as suitable for online analysis ([12]–[15], [17], [19], [20], [29], [31], [35]). Other papers do not give any information about the time delay in seizure detection or the kind of application.

To conclude, the purpose of this survey was to highlight the results of the analysis of EEG, ECG and video recordings for the identification of epileptic seizures in the newborn. This paper also aims at highlighting that the combined use of the three signals can lead to significant improvements providing complementary information. Cabon *et al.* [73] proposed a semi-automatic system for the estimation of the sleep stages of premature newborns in NICUs through video and audio recordings analysis; Chen *et al.* [74] presented a wearable sensor system for simultaneous recording of ECG and respiration in newborns to monitor the neonatal health status. However, to the best of our knowledge, none of the existing systems combine the three signals for the neonatal seizure detection task in NICUs. Implementing a multimodal approach that exploits the results of several domains could be useful for developing an efficient and reliable automatic system to support clinicians.

VII. CONCLUSION

This paper summarises the main attempts to develop NSD systems proposed in the last ten years. Several studies focused on the EEG analysis to define a system that automatically recognizes critical events. Indeed, investigating the EEG signal allows obtaining higher performances than other electrophysiological and clinical signals. ECG- and video-based systems have also been investigated: the former is based on evaluating the seizures influence on the heart rate, the latter on the recognition and characterization of ''unusual'' movements. It has been shown that the technological progress and the development of signal processing techniques allowed defining possible support tools for the medical staff, which could improve neonatal seizure detection in clinical scenarios. Moreover, it has been shown that the EEG, ECG and video signals provide complementary information. Therefore, a multimodal approach that exploits and combines the results of the three approaches could be investigated in future. NSD systems able to automatically characterize the etiologies investigating the available electrophysiological and clinical signals could be a valuable support for clinicians.

ACKNOWLEDGMENT

(Antonio Lanatà and Claudia Manfredi contributed equally to this work.)

REFERENCES

[1] G. Ferrari, G. M. Kouamou, C. Copioli, R. Raheli, and F. Pisani, ''Low-complexity image processing for real-time detection of neonatal clonic seizures,'' in *Proc. 3rd Int. Symp. Appl. Sci. Biomed. Commun. Technol. (ISABEL)*, Nov. 2010, pp. 1–5, doi: 10.1109/ISABEL.2010.5702898.

[2] F. Pisani, C. Spagnoli, R. Falsaperla, L. Nagarajan, and G. Ramantani, ''Seizures in the neonate: A review of etiologies and outcomes,'' *Seizure*, vol. 85, pp. 48–56, Feb. 2021, doi: 10.1016/j.seizure.2020.12.023.

[3] R. M. Pressler, ''Neonatal seizures,'' Int. League Against Epilepsy, Univ. Oxford Math. Inst., 2015, ch. 6. [Online]. Available: https://epilepsysociety.org.uk/sites/default/files/2020-08/Chapter06 Pressler2015_0.pdf

[4] R. A. Shellhaas, ''Seizure classification, etiology, and management,'' in *Handbook of Clinical Neurology*, vol. 162. Amsterdam, The Netherlands: Elsevier, 2019, pp. 347–361, doi: 10.1016/B978-0-444-64029-1.00017-5.

[5] R. M. Pressler and L. Lagae, ''Why we urgently need improved seizure and epilepsy therapies for children and neonates,'' *Neuropharmacology*, vol. 170, Jun. 2020, Art. no. 107854, doi: 10.1016/j.neuropharm.2019.107854.

[6] C. P. Panayiotopoulos, *The Epilepsies: Seizures, Syndromes and Management*. Oxfordshire, U.K.: Bladon Medical Publishing, 2005.

[7] T. Chang and A. du Plessis, ''Neurodiagnostic techniques in neonatal critical care,'' *Current Neurol. Neurosci. Rep.*, vol. 12, no. 2, pp. 145–152, Apr. 2012, doi: 10.1007/s11910-012-0254-y.

[8] R. A. Shellhaas, T. Chang, T. Tsuchida, M. S. Scher, J. J. Riviello, N. S. Abend, S. Nguyen, C. J. Wusthoff, and R. R. Clancy, "The American clinical neurophysiology society's guideline on continuous electroencephalography monitoring in neonates," *J. Clin. Neurophysiol.*, vol. 28, no. 6, pp. 611–617, Dec. 2011, doi: 10.1097/WNP.0b013e31823e96d7.

[9] A. Caliskan and S. Rencuzogullari, "Transfer learning to detect neonatal seizure from electroencephalography signals," *Neural Comput. Appl.*, vol. 33, no. 18, pp. 12087–12101, Sep. 2021, doi: 10.1007/s00521-021-05878-y.

[10] L. Frassineti, C. Barba, F. Melani, F. Piras, R. Guerrini, and C. Manfredi, "Automatic detection and sonification of nonmotor generalized onset epileptic seizures: Preliminary results," *Brain Res.*, vol. 1721, Oct. 2019, Art. no. 146341, doi: 10.1016/j.brainres.2019.146341.

[11] A. Liu, J. S. Hahn, G. P. Heldt, and R. W. Coen, "Detection of neonatal seizures through computerized EEG analysis," *Electroencephalogr. Clin. Neurophysiol.*, vol. 82, no. 1, pp. 30–37, 1992, doi: 10.1016/0013-4694(92)90179-L.

[12] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: Methods and initial evaluation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 356–362, Sep. 1997, doi: 10.1016/S0013-4694(97)00003-9.

[13] W. Deburchgraeve, P. J. Cherian, M. De Vos, R. M. Swarte, J. H. Blok, G. H. Visser, P. Govaert, and S. Van Huffel, "Automated neonatal seizure detection mimicking a human observer reading EEG," *Clin. Neurophysiol., Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 119, no. 11, pp. 2447–2454, Nov. 2008, doi: 10.1016/j.clinph.2008.07.281.

[14] M. A. Navakatikyan, P. B. Colditz, C. J. Burke, T. E. Inder, J. Richmond, and C. E. Williams, "Seizure detection algorithm for neonates based on wave-sequence analysis," *Clin. Neurophysiol., Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 117, no. 6, pp. 1190–1203, Jun. 2006, doi: 10.1016/j.clinph.2006.02.016.

[15] E. M. Thomas, A. Temko, G. Lightbody, W. P. Marnane, and G. B. Boylan, "Gaussian mixture models for classification of neonatal seizures using EEG," *Physiol. Meas.*, vol. 31, no. 7, pp. 1047–1064, 2010, doi: 10.1088/0967-3334/31/7/013.

[16] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with support vector machines," *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 464–473, Mar. 2011, doi: 10.1016/j.clinph.2010.06.034.

[17] A. M. Pavel, J. M. Rennie, L. S. de Vries, M. Blennow, A. Foran, D. K. Shah, R. M. Pressler, O. Kapellou, E. M. Dempsey, S. R. Mathieson, E. Pavlidis, A. C. van Huffelen, V. Livingstone, M. C. Toet, L. C. Weeke, M. Finder, S. Mitra, D. M. Murray, W. P. Marnane, and G. B. Boylan, "A machine-learning algorithm for neonatal seizure recognition: A multicentre, randomised, controlled trial," *Lancet Child Adolescent Health*, vol. 4, no. 10, pp. 740–749, 2020, doi: 10.1016/S2352-4642(20)30239-X.

[18] K. T. Tapani, S. Vanhatalo, and N. J. Stevenson, "Time-varying EEG correlations improve automated neonatal seizure detection," *Int. J. Neural Syst.*, vol. 29, no. 4, May 2019, Art. no. 1850030, doi: 10.1142/S0129065718500302.

[19] A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos, and S. Van Huffel, "Neonatal seizure detection using deep convolutional neural networks," *Int. J. Neural Syst.*, vol. 29, no. 4, May 2019, Art. no. 1850011. [Online]. Available: https://core.ac.uk/download/pdf/158601452.pdf, doi: 10.1142/S0129065718500119.

[20] A. O'Shea, G. Lightbody, G. Boylan, and A. Temko, "Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture," *Neural Netw.*, vol. 123, pp. 12–25, Mar. 2020, doi: 10.1016/j.neunet.2019.11.023.

[21] M. A. Tanveer, M. J. Khan, H. Sajid, and N. Naseer, "Convolutional neural networks ensemble model for neonatal seizure detection," *J. Neurosci. Methods*, vol. 358, Jul. 2021, Art. no. 109197, doi: 10.1016/j.jneumeth.2021.109197.

[22] A. O'Shea, R. Ahmed, G. Lightbody, E. Pavlidis, R. Lloyd, F. Pisani, W. Marnane, S. Mathieson, G. Boylan, and A. Temko, "Deep learning for EEG seizure detection in preterm infants," *Int. J. Neural Syst.*, vol. 31, no. 8, 2021, Art. no. 2150008. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/2106/2106.00611.pdf, doi: 10.1142/S0129065721500088.

[23] M. B. Malarvili, M. Mesbah, and B. Boashash, "Time-frequency analysis of heart rate variability for neonatal seizure detection," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, pp. 1–10, Dec. 2007. [Online]. Available: https://asp-eurasipjournals.springeropen.com/track/pdf/10.1155/2007/50396.pdf

[24] B. R. Greene, P. D. Chazal, G. B. Boylan, S. Connolly, and R. B. Reilly, "Electrocardiogram based neonatal seizure detection," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 673–682, Apr. 2007, doi: 10.1109/TBME.2006.890137.

[25] O. M. Doyle, A. Temko, W. Marnane, G. Lightbody, and G. B. Boylan, "Heart rate based automatic seizure detection in the newborn," *Med. Eng. Phys.*, vol. 32, no. 8, pp. 829–839, Oct. 2010, doi: 10.1016/j.medengphy.2010.05.010.

[26] B. R. Greene, G. Boylan, R. Reilly, P. Chazal, and S. Connolly, "Combination of EEG and ECG for improved automatic neonatal seizure detection," *Clin. Neurophysiol., Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 118, pp. 59–1348, Jun. 2007, doi: 10.1016/j.clinph.2007.02.015.

[27] M. Mesbah, M. Balakrishnan, P. B. Colditz, and B. Boashash, "Automatic seizure detection based on the combination of newborn multi-channel EEG and HRV information," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–14, Dec. 2012, doi: 10.1186/1687-6180-2012-215.

[28] A. Temko, O. Doyle, D. Murray, G. Lightbody, G. Boylan, and W. Marnane, "Multimodal predictor of neurodevelopmental outcome in newborns with hypoxic-ischaemic encephalopathy," *Comput. Biol. Med.*, vol. 63, pp. 169–177, Aug. 2015, doi: 10.1016/j.compbiomed.2015.05.017.

[29] G. Kouamou, G. Ferrari, F. Lofino, R. Raheli, and F. Pisani, "Extraction of video features for real-time detection of neonatal seizures," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2011, pp. 1–6, doi: 10.1109/WoWMoM.2011.5986193.

[30] G. M. K. Ntonfo, F. Lofino, G. Ferrari, R. Raheli, and F. Pisani, "Video processing-based detection of neonatal seizures by trajectory features clustering," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 3456–3460, doi: 10.1109/ICC.2012.6364396.

[31] F. Pisani, C. Spagnoli, E. Pavlidis, C. Facini, G. M. K. Ntonfo, G. Ferrari, and R. Raheli, "Real-time automated detection of clonic seizures in newborns," *Clin. Neurophysiol., Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 125, no. 8, pp. 1533–1540, Aug. 2014, doi: 10.1016/j.clinph.2013.12.119.

[32] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, and F. Pisani, "Monitoring infants by automatic video processing: A unified approach to motion analysis," *Comput. Biol. Med.*, vol. 80, pp. 158–165, Jan. 2017, doi: 10.1016/j.compbiomed.2016.11.010.

[33] N. B. Karayiannis, S. Srinivasan, R. Bhattacharya, M. Wise, J. D. Frost, Jr., and E. M. Mizrahi, "Extraction of motion strength and motor activity signals from video recordings of neonatal seizures," *IEEE Trans. Med. Imag.*, vol. 20, no. 9, pp. 965–980, Sep. 2001. [Online]. Available: https://uh.edu/~elee8zt/ieeepaper.pdf, doi: 10.1109/42.952733.

[34] N. B. Karayiannis and G. Tao, "Improving the extraction of temporal motion strength signals from video recordings of neonatal seizures," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill. (AVSS)*, Jul. 2003, pp. 87–92. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.568.657&rep=rep1&type=pdf

[35] N. B. Karayiannis and G. Tao, "An improved procedure for the extraction of temporal motion strength signals from video recordings of neonatal seizures," *Image Vis. Comput.*, vol. 24, no. 1, pp. 27–40, Jan. 2006. [Online] Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.3857&rep=rep1&type=pdf, doi: 10.1016/j.imavis.2005.09.009.

[36] N. B. Karayiannis, G. Tao, Y. Xiong, A. Sami, B. Varughese, J. D. Frost, M. S. Wise, and E. M. Mizrahi, "Computerized motion analysis of videotaped neonatal seizures of epileptic origin," *Epilepsia*, vol. 46, no. 6, pp. 901–917, Jun. 2005, doi: 10.1111/j.1528-1167.2005.56504.x.

[37] A. Temko and G. Lightbody, "Detecting neonatal seizures with computer algorithms," *J. Clin. Neurophysiol.*, vol. 33, no. 5, pp. 394–402, Oct. 2016, doi: 10.1097/WNP.0000000000000295.

[38] S. Cabon, F. Porée, A. Simon, O. Rosec, P. Pladys, and G. Carrault, "Video and audio processing in paediatrics: A review," *Physiol. Meas.*, vol. 40, no. 2, 2019, Art. no. 02TR02, doi: 10.1088/1361-6579/ab0096.

[39] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. B. Boylan, "Performance assessment for EEG-based neonatal seizure detectors," *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 474–482, Mar. 2011, doi: 10.1016/j.clinph.2010.06.035.

[40] A. H. Ansari, P. J. Cherian, A. Caicedo Dorado, K. Jansen, A. Dereymaeker, L. De Wispelaere, C. Dielman, J. Vervisch, P. Govaert, M. De Vos, G. Naulaers, and S. V. Huffel, "Weighted performance metrics for automatic neonatal seizure detection using multiscored EEG data," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1114–1123, Jul. 2018, doi: 10.1109/JBHI.2017.2750769.

[41] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of K-fold cross-validation," *J. Mach. Learn. Res.*, vol. 5, pp. 1089–1105, Dec. 2004. [Online]. Available: https://www.jmlr.org/papers/volume5/grandvalet04a/grandvalet04a.pdf

[42] A. Temko, A. K. Sarkar, G. B. Boylan, S. Mathieson, W. P. Marnane, and G. Lightbody, "Toward a personalized real-time diagnosis in neonatal seizure detection," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–14, 2017, doi: 10.1109/JTEHM.2017.2737992.

[43] N. J. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, "A dataset of neonatal EEG recordings with seizure annotations," *Sci. Data*, vol. 6, no. 1, Mar. 2019, Art. no. 190039, doi: 10.1038/sdata.2019.39.

[44] L. Frassineti, D. Ermini, R. Fabbri, and C. Manfredi, "Neonatal seizures detection using stationary wavelet transform and deep neural networks: Preliminary results," in *Proc. IEEE 20th Medit. Electrotech. Conf. (MELECON)*, Jun. 2020, pp. 344–349, doi: 10.1109/MELECON48756.2020.9140713.

[45] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020. [Online] Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8943952, doi: 10.1109/ACCESS.2019.2962617.

[46] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. ICANN*, 2018, pp. 270–279. [Online]. Available: https://arxiv.org/pdf/1808.01974.pdf, doi: 10.1007/978-3-030-01424-7_27.

[47] R. N. Goldberg, S. L. Goldman, R. E. Ramsay, and R. Feller, "Detection of seizure activity in the paralyzed neonate using continuous monitoring," *Pediatrics*, vol. 69, no. 5, pp. 583–586, 1982.

[48] K. Watanabe, K. Hara, S. Miyazaki, M. Kuroyanagi, S. Asano, K. Kondo, K. Kuno, H. Jose, and K. Iwase, "Electroclinical studies of seizures in the newborn," *Folia Psychiatrica Neurologica Japonica*, vol. 31, no. 3, pp. 383–392, 1977, doi: 10.1111/j.1440-1819.1977.tb02625.x.

[49] I. Bersani, F. Piersigilli, D. Gazzolo, F. Campi, I. Savarese, A. Dotta, P. P. Tamborrino, C. Auriti, and C. Di Mambro Bersani, "Heart rate variability as possible marker of brain damage in neonates with hypoxic ischemic encephalopathy: A systematic review," *Eur. J. Pediatrics*, vol. 180, pp. 1–11, Nov. 2020, doi: 10.1007/s00431-020-03882-3.

[50] C. Bock, P. Kovacs, P. Laguna, J. Meier, and M. Huemer, "ECG beat representation and delineation by means of variable projection," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 10, pp. 2997–3008, Oct. 2021, doi: 10.1109/TBME.2021.3058781.

[51] M. B. Malarvili and M. Mesbah, "Newborn seizure detection based on heart rate variability," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2594–2603, Nov. 2009, doi: 10.1109/TBME.2009.2026908.

[52] G. M. K. Ntonfo, "Monitoring and diagnosing neonatal seizures by video signal processing," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 13, no. 2, pp. 58–60, 2014. [Online]. Available: https://www.raco.cat/index.php/ELCVIA/article/download/281638/369507

[53] A. Malone, C. A. Ryan, A. Fitzgerald, L. Burgoyne, S. Connolly, and G. B. Boylan, "Interobserver agreement in neonatal seizure identification," *Epilepsia*, vol. 50, no. 9, pp. 2097–2101, Sep. 2009, doi: 10.1111/j.1528-1167.2009.02132.x.

[54] N. B. Karayiannis, B. Varughese, G. Tao, J. D. Frost, M. S. Wise, and E. M. Mizrahi, "Quantifying motion in video recordings of neonatal seizures by regularized optical flow methods," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 890–903, Jul. 2005, doi: 10.1109/TIP.2005.849320.

[55] N. B. Karayiannis, A. Sami, J. D. Frost, M. S. Wise, and E. M. Mizrahi, "Automated extraction of temporal motor activity signals from video recordings of neonatal seizures based on adaptive block matching," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 676–686, Apr. 2005, doi: 10.1109/TBME.2005.845154.

[56] N. B. Karayiannis, A. Sami, J. D. Frost, M. S. Wise, and E. M. Mizrahi, "Quantifying motion in video recordings of neonatal seizures by feature trackers based on predictive block matching," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2004, pp. 1447–1450, doi: 10.1109/IEMBS.2004.1403447.

[57] N. B. Karayiannis, Y. Xiong, J. D. Frost, M. S. Wise, and E. M. Mizrahi, "Quantifying motion in video recordings of neonatal seizures by robust motion trackers based on block motion models," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 6, pp. 1065–1077, Jun. 2005. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.80.1179&rep=rep1&type=pdf, doi: 10.1109/TBME.2005.846715.

[58] A. Van de Vel, K. Cuppens, B. Bonroy, M. Milosevic, K. Jansen, S. Van Huffel, B. Vanrumste, P. Cras, L. Lagae, and B. Ceulemans, "Non-EEG seizure detection systems and potential SUDEP prevention: State of the art: Review and update," *Seizure*, vol. 41, pp. 141–153, Oct. 2016, doi: 10.1016/j.seizure.2016.07.012.

[59] A. Ulate-Campos, F. Coughlin, M. Gaínza-Lein, I. S. Fernández, P. L. Pearl, and T. Loddenkemper, "Automated seizure detection systems and their effectiveness for each type of seizure," *Seizure*, vol. 40, pp. 88–101, Aug. 2016, doi: 10.1016/j.seizure.2016.06.008.

[60] L. Frassineti, A. Parente, and C. Manfredi, "Multiparametric EEG analysis of brain network dynamics during neonatal seizures," *J. Neurosci. Methods*, vol. 348, Jan. 2021, Art. no. 109003, doi: 10.1016/j.jneumeth.2020.109003.

[61] C. Baumgartner, J. P. Koren, and M. Rothmayer, "Automatic computer-based detection of epileptic seizures," *Frontiers Neurol.*, vol. 9, p. 639, Aug. 2018, doi: 10.3389/fneur.2018.00639.

[62] G. F. T. Variane, V. Y. Chock, A. Netto, R. F. R. Pietrobom, and K. P. Van Meurs, "Simultaneous near-infrared spectroscopy (NIRS) and amplitude-integrated electroencephalography (aEEG): Dual use of brain monitoring techniques improves our understanding of physiology," *Frontiers Pediatrics*, vol. 7, p. 560, Jan. 2020, doi: 10.3389/fped.2019.00560.

[63] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1. Amsterdam, The Netherlands: Elsevier, 2018, pp. 542–545. [Online]. Available: https://www.researchgate.net/profile/Daniel-Berrar/publication/324701535_Cross-Validation/links/5cb4209c92851c8d22ec4349/Cross-Validation.pdf

[64] R. Statello, L. Carnevali, D. Alinovi, F. Pisani, and A. Sgoifo, "Heart rate variability in neonatal patients with seizures," *Clin. Neurophysiol., Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 129, no. 12, pp. 2534–2540, Dec. 2018, doi: 10.1016/j.clinph.2018.10.001.

[65] D. Hendrikx, L. Thewissen, A. Smits, G. Naulaers, K. Allegaert, S. Van Huffel, and A. Caicedo, "Using graph theory to assess the interaction between cerebral function, brain hemodynamics, and systemic variables in premature infants," *Complexity*, vol. 2018, pp. 1–15, Aug. 2018, doi: 10.1155/2018/6504039.

[66] G. Valenza, A. Greco, C. Gentili, A. Lanata, L. Sebastiani, D. Menicucci, A. Gemignani, and E. P. Scilingo, "Combining electroencephalographic activity and instantaneous heart rate for assessing brain–heart dynamics during visual emotional elicitation in healthy subjects," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2067, May 2016, Art. no. 20150176, doi: 10.1098/rsta.2015.0176.

[67] J. M. Rennie, "Neonatal seizures," *Eur. J. Pediatrics*, vol. 156, no. 2, pp. 83–87, Jan. 1997. [Online]. Available: http://web4.uwindsor.ca/users/b/boulos/main.nsf/9d019077a3c4f6768525698a00593654/703453ccb701463685256ea50004a158/FILE/SEISURES.pdf, doi: 10.1007/s004310050559.

[68] M. L. Nunes, E. G. Yozawitz, S. Zuberi, E. M. Mizrahi, M. R. Cilio, S. L. Moshé, P. Plouin, S. Vanhatalo, and R. M. Pressler, "Neonatal seizures: Is there a relationship between ictal electroclinical features and etiology? A critical appraisal based on a systematic literature review," *Epilepsia Open*, vol. 4, no. 1, pp. 10–29, Mar. 2019, doi: 10.1002/epi4.12298.

[69] T. T. Sands and T. L. McDonough, "Recent advances in neonatal seizures," *Current Neurol. Neurosci. Rep.*, vol. 16, no. 10, pp. 1–10, Oct. 2016, doi: 10.1007/s11910-016-0694-x.

[70] A. Temko, W. Marnane, G. Boylan, and G. Lightbody, "Clinical implementation of a neonatal seizure detection algorithm," *Decis. Support Syst.*, vol. 70, pp. 86–96, Feb. 2015, doi: 10.1016/j.dss.2014.12.006.

[71] L. S. Vidyaratne and K. M. Iftekharuddin, "Real-time epileptic seizure detection using EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 2146–2156, Nov. 2017, doi: 10.1109/TNSRE.2017.2697920.

[72] W. M. J. Apers, L. S. de Vries, F. Groenendaal, M. C. Toet, and L. C. Weeke, "Delay in treatment of neonatal seizures: A retrospective cohort study," *Neonatology*, vol. 117, no. 5, pp. 599–605, 2020, doi: 10.1159/000509282.

[73] 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," *Biomed. Signal Process. Control*, vol. 52, pp. 362–370, Jul. 2019, doi: 10.1016/j.bspc.2019.04.011.

[74] H. Chen, W. Chen, S. Bao, C. Lu, L. Wang, J. Ma, P. Wang, H. Lu, F. Shu, and S. B. Oetomo, "Design of an integrated wearable multi-sensor platform based on flexible materials for neonatal monitoring," *IEEE Access*, vol. 8, pp. 23732–23747, 2020. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8976137

BENEDETTA OLMI was born in Firenze, Italy, in 1995. She received the B.S. degree in electronics and telecommunications and the M.S. degree in biomedical engineering from the Università degli Studi di Firenze, Firenze, in 2018 and 2021, respectively.

From January 2021 to June 2021, she got a scholarship at the Department of Information Engineering, Università degli Studi di Firenze, where she has been holding a research grant, since July 2021. Her research interests include biomedical signal analysis and machine learning algorithms for biomedical applications.

LORENZO FRASSINETI (Member, IEEE) was born in Firenze, Italy, in 1992. He received the B.S. degree in electronics and telecommunications and the M.S. degree in biomedical engineering from the Università degli Studi di Firenze, Firenze, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the School of Genetics, Oncology and Clinical Medicine, University of Siena, in collaboration with the Department of Information Engineering, Università degli Studi di Firenze. His research interests include the study of machine learning algorithms for biomedical applications, automatic detection of neurological disorders, biosignal analysis, and sonification techniques.

ANTONIO LANATÀ (Member, IEEE) received the Ph.D. degree. He is currently an Associate Professor of bioengineering at the Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy. His research interests include design and implementation of wearable system for physiological monitoring and statistical and nonlinear biomedical signal processing. He is the author of more than 120 international scientific contributions in these fields published in peer-reviewed international journals, conference proceedings, books, and book chapters. He is an Associate Editor of *Bionics and Biomimetics* (Frontiers), *Bioengineering* (MDPI), *Bioelectronics* (MDPI), *Biosensors* (MDPI), *Algorithms* (MDPI), *Electronics* (MDPI), *Animals* (MDPI), and *The Open Cybernetics and Systemics Journal*. Recently, he has also been an Associate Guest Editor of several special issues for *Bioengineering* (MDPI), *Electronics—Bioelectronics* (MDPI), *Frontiers in Bioengineering and Biotechnology*, and *Frontiers in Neurorobotics*.

CLAUDIA MANFREDI (Member, IEEE) is currently an Associate Professor of biomedical engineering at the Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy. Her research interests include biomedical signal processing, with application to the analysis of EEG, ECG, video recordings, and the human voice. On the latter subject, she has been organizing the Biannual International Workshop MAVEBA, since 1999. She is the author of more than 120 peer-reviewed papers indexed in Scopus. She is a member of IEEE BME Society, Italian Bioengineering Group (GNB), International Speech and Communication Association (ISCA), Collegium Medicorum Theatri (CoMeT), Pan European VOice Conference (PEVOC), and World Voice Day. She is a member of the Editorial Board of *Biomedical Signal Processing and Control* (Elsevier Ltd.), an Editor of the MAVEBA proceedings series (FUP, Italy), and a Guest Editor of special issues of the journals, including *European Acoustics Association* (Acustica-Acta Acustica), *Medical Engineering and Physics* (Elsevier Ltd.), and *Biomedical Signal Processing and Control* (Elsevier Ltd.).