Urdal, J.; Engan, K.; Eftestøl, T.; Naranjo Ornedo, V.; Haud, IA.; Yeconia, A.; Kidanto, H.... (2020). Automatic identification of stimulation activities during newborn resuscitation using ECG and accelerometer signals. Computer Methods and Programs in Biomedicine. 193:1-9. https://doi.org/10.1016/j.cmpb.2020.105445

# Automatic identification of stimulation activities during newborn resuscitation using ECG and Accelerometer signals

Jarle Urdal*, Kjersti Engan*, Trygve Eftestøl*, Valery Naranjo†,
Ingunn Anda Haug‡, Anita Yeconia§, Hussein Kidanto¶, Hege Ersdal‖**
*Dep. of Electrical Engineering and Computer Science, University of Stavanger, Norway
Email: {jarle.urdal, kjersti.engan}@uis.no
† Instituto de Investigación e Innovación en Bioingeniería (I3B), Universitat Politécnica de Valéncia, Spain
‡Strategic Research, Laerdal Medical AS, Stavanger, Norway
§Haydom Lutheran Hospital, Haydom, Manyara, Tanzania
¶School of Medicine, Aga Khan University, Dar es Salaam, Tanzania
‖Dep. of Anesthesiology and Intensive Care, Stavanger University Hospital, Norway
**Dep. of Health Sciences, University of Stavanger, Norway

***Abstract*—*Background and Objective:* Early neonatal death is a worldwide challenge with 1 million newborn deaths every year. The primary cause of these deaths are complications during labour and birth asphyxia. The majority of these newborns could have been saved with adequate resuscitation at birth. Newborn resuscitation guidelines recommend immediate drying, stimulation, suctioning if indicated, and ventilation of non-breathing newborns. A system that will automatically detect and extract time periods where different resuscitation activities are performed, would be highly beneficial to evaluate what resuscitation activities that are improving the state of the newborn, and if current guidelines are good and if they are followed. The potential effects of especially stimulation are not very well documented as it has been difficult to investigate through observations. In this paper the main objective is to identify stimulation activities, regardless if the state of the newborn is changed or not, and produce timelines of the resuscitation episode with the identified stimulations. *Methods:* Data is collected by utilizing a new heart rate device, NeoBeat, with dry-electrode ECG and accelerometer sensors placed on the abdomen of the newborn. We propose a method, NBstim, based on time domain and frequency domain features from the accelerometer signals and ECG signals from NeoBeat, to detect time periods of stimulation. NBstim use causal features from a gliding window of the signals, thus it can potentially be used in future realtime systems. A high performing feature subset is found using feature selection. System performance is computed using a leave-one-out cross-validation and compared with manual annotations. *Results:* The system achieves an overall accuracy of 90.3% when identifying regions with stimulation activities. *Conclusion:* The performance indicates that the proposed NBstim, used with signals from the NeoBeat can be used to determine when stimulation is performed. The provided activity timelines, in combination with the status of the newborn, for example the heart rate, at different time points, can be studied further to investigate both the time spent and the effect of different newborn resuscitation parameters.**

***Index Terms*—Newborn resuscitation, Activity recognition, Automatic annotation, Machine learning**

## I. Introduction

Early neonatal death is a worldwide challenge with 1 million newborn deaths every year, and the vast majority of these are found in low and low-middle income contries [1]. The primary cause of these deaths are complications during labour and birth asphyxia [1], [2]. Guidelines on newborn resuscitaion are published by both the World Health Organization and others [3], [4]. The general guideline is to start resuscitation within the first minute after birth if the newborn is unable to start breathing [5]. A gap between the medical guidelines and what is actually performed in practice has also been observed [6]. While resuscitation immediately after birth is a crucial part of saving these lives, the full understanding of how to best apply therapeutic activities is not well documented. Therapeutic activities includes stimulation of the newborn, like firmly rubbing the back and drying, removal of mucus and obstructions in the airways by suction, and bag mask ventilation. The amount of activities performed during resuscitation, such as tactile stimulation and bag-mask ventilation, has been shown to be correlated with the 24-hour outcome of the newborn [7]. Further analyses should be conducted to study the importance of factors like duration and order of these therapeutic activities.

Safer Births[1] is a large and collaborative research project with the goal of establishing new knowledge and develop new innovative products to save lives at birth. One of the goals of this collaborative project is to construct a system that can automatically detect time periods where different resuscitation activities are performed. Such a system can be used as part of a debriefing system, and will make it possible to evaluate a large number of episodes to find out which resuscitation activities that are improving the state of the newborn, if current guidelines are good and if they are followed. The state of the newborn can effectively be evaluated by assessing the heart rate [4], and a change in the observed heart rate may be the result of prolonged resuscitation activities. There might as well be potential for real time decision support during resuscitation. A number of sensor data have been collected during

[1] http://www.saferbirths.com/

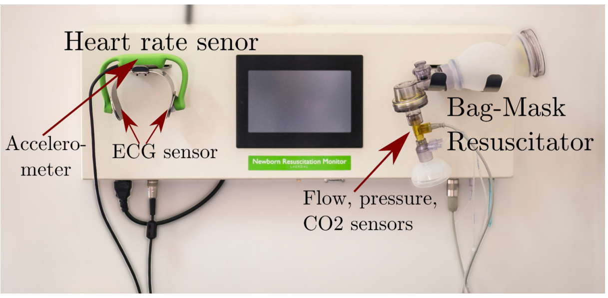


Fig. 1. Laerdal Newborn Resuscitation Monitor with the various sensors indicated. The measured heart rate is shown on the LCD to give feedback to the health care personnel. The green buckle with accelerometer and dry-electrode ECG is a prototype of the NeoBeat. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

newborn resuscitation at partner hospitals in Tanzania during the research project; pressure and flow from the bag-mask resuscitator (BMR), dry-electrode electrocardiogram (ECG) signals and signals from an accelerometer using a prototype of the NeoBeat[2], attached over the abdomen of the newborn.

The detection of bag-mask ventilation can relatively easy be performed using the flow and pressure signals from sensors mounted in the BMR [8], and detection and recognition of treatment activities during newborn resuscitation using deep neural networks on videos of resuscitation [9] has been described in earlier work from this research group. Videos of the resuscitation is often not available, or the view can be blocked by some of the activities. Thus it would be beneficial to be able to detect *stimulation* based on the NeoBeat signals. As tactile stimulation during newborn resuscitation involves some kind of repetitive movement, we hypothesize that these activities can be picked up using an accelerometer attached to the newborn through the use of NeoBeat.

Detection and recognition of activities using data from an accelerometer have previously been explored on healthy adults: static and dynamic activities such as sitting positions, walking versus running were found using an accelerometer mounted on the subjects back [10] and the subjects waist [11]. With the rise of wearable technology in every day life such as sport watches and cell phones, accelerometers are now available for activity recognition using commercially available devices [12]. To the authors knowledge, there are no reported correlation between the ECG morphology [13] and external stimulation, and no reported works utilizing accelerometer and ECG signals of newborns to automatically classify therapeutic activities, except from this research group.

Automatic detection of some sort of activity performed by the health care personnel, VuDetector, has previously been proposed by members of our research group [14]. VuDetector achived a sensitivty of 90% and a specificity of 80%. During a resuscitation, the newborn will be moved, covered and uncovered etc. and such activities will also be visible on the accelerometer signals from the attached NeoBeat, but are not considered therapeutic activities. In VuClassifier [15], we also proposed a first attempt of classification of the detected activities based on ECG and accelerometer signals, reported with an accuracy of 79.8 % when distinguishing stimulation and chest compression from other activities. The VuClassifier was, however, based on signal features extracted from detected activity events of variable duration, it needed statistics from the entire resuscitation episode, and as such, only suitable for retrospective analyses, and it was trained and tested on relatively few episodes, and needed further verification.

In this work the main objective is to propose a system, NBstim, for detecting time periods of *stimulation activities* based on the signals recorded by the NeoBeat placed on the abdomen of the newborn, using causal signal features, and as such, suitable for real time analysis. In combination with a method of detected bag-mask ventilation sequences [8], this can be used to create useful timelines illustrating the amount, duration and order of ventilation and stimulation performed in real world newborn resuscitation episodes, see figure 2. In the rest of the paper we start by explaining the data material and the manual annotations in section II. In Section III the proposed NBstim is explain in context with the larger system, thereafter the signal featuers are defined. In the experimental section, VuClassifier and NBstim is tested using a larger dataset and compared with manually annotated activities.

## II. Data material

The data material used in this work was collected at Haydom Lutheran Hospital (HLH), a rural hospital in Tanzania, between October 2013 and September 2016 by the Safer Births project. The research project was approved by the Regional Committee for Medical and Health Research Ethics (REK) in Norway (2013/110/REK vest) and National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/1434). Parenteral verbal consent was obtained for all resuscitated newborns. Within this research project, subprojects have been subject to randomized trails. For this particular work, the data collection has been part of an observational study, not an intervention study.

The data were collected using the Laerdal NeoBeat prototype, Figure 1, which is part of the research device Laerdal Newborn Resuscitation Monitor [8], [14], [15], [16]. The NeoBeat prototype measures the heart rate using two dry-electrode ECG sensors attached to a buckle, which is placed over the abdomen of the newborn. This design allows the health care personnel to quickly attach the ECG sensor to the newborn and monitor the heart rate, and can therefore focus on giving the best treatment possible without struggling with gel and placement of the ECG sensors. An example of ECG and accelerometer signals measured using the NeoBeat prototype is shown in Figure 2. Due to the combination of dry-electrode ECG sensors and an environment with a lot of movement, the measured signal contains more noise than what is seen when using traditional ECG in settings with less movement. In HLH, a resuscitation monitor is installed in each of the labour rooms and the midwifes are primarily responsible for the health care both during labours and potential resuscitation immediately after birth. The health care workers involved in the data collection were trained to follow the existing Helping Babies

[2]https://laerdalglobalhealth.com/products/neobeat-newborn-heart-rate-meter/

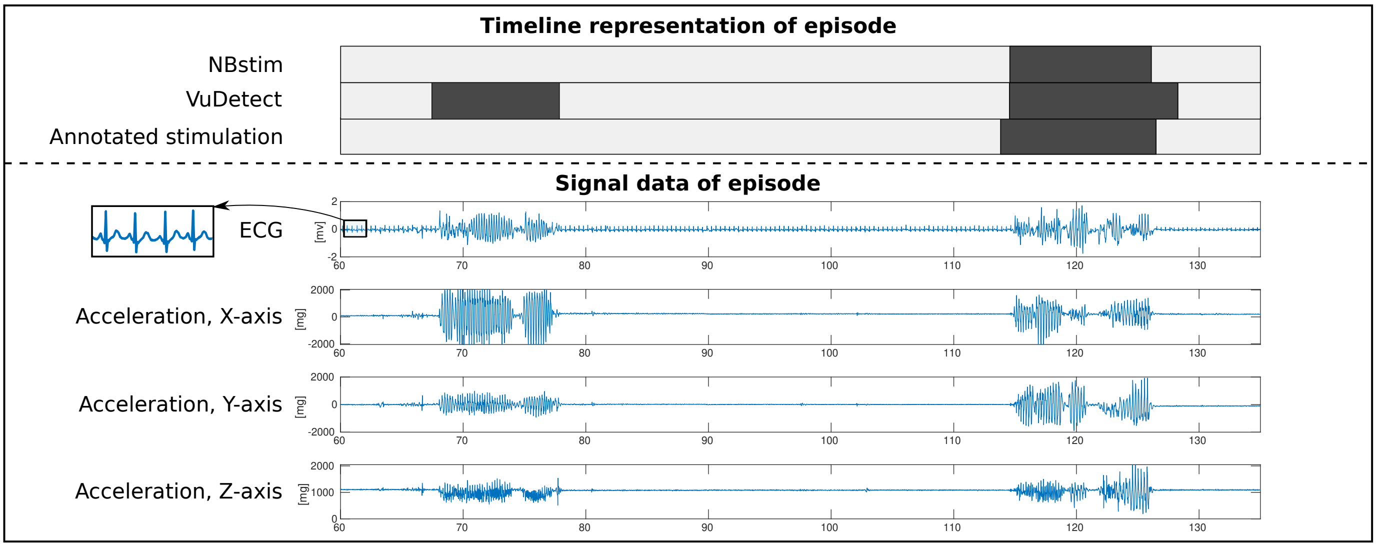


Fig. 2. Example segment of the measured ECG and accelerometer signals with the corresponding timeline representation. The signal with a lower amplitude in the beginning of the ECG signal corresponds to the expected QRS-complex. Two distinct regions with movement can be seen in the signals, and VuDetect identifies both activities. NBstim, proposed in this work, evaluates the regions found by VuDetect and classify them as stimulation or non-stimulation activities.

Breathe (HBB) guidelines [3] for newborn resuscitation. These guidelines state what should be checked, and what action to perform if the newborn is asphyxiated and need help to start breathing. The guidelines were posted on the wall above each resuscitation bed to remind the health care personnel to follow them. A limitation of the guidelines is, however, that they only defines what activity to perform, and not the amount, length, or how often the activity should be performed. Additional clinical data related to the labour and resuscitation was logged by designated research assistants present at the labour ward for the research project.

The resuscitation monitor consists of a main processing unit with a display to show the measured heart rate, as well as the heart rate sensor and a bag-mask resuscitator (BMR). The NeoBeat prototype seen in green in Figure 1, contains dry-electrode ECG, sampled at 500 Hz, and a three-axes accelerometer to monitor movement of the newborn, sampled at 100 Hz. The ventilation bag includes pressure and flow sensors, sampled at 100 Hz, as well as a $CO_2$ sensor sampled at 20 Hz.

A total of 916 resuscitation episodes were recorded during the data collection period. A set of 76 randomly selected videos were annotated to obtain a timeline description of the resuscitation for further evaluation.

### A. Annotations

Videos of the resuscitation were annotated by two independent reviewers; one neonatologist and one human factors engineer. In cases with agreement score $< 80\%$, the two reviewers sat together and obtained consensus. The following categories were annotated: 1) *stimulation*, 2) *suction*, 3) *uncovered*, 4) *other*, 5) *obscured view*, and 6) *start/stop of resuscitation*. If the resuscitation lasted longer than seven minutes, only the first seven minutes were annotated. Stimulation and suction are considered two of the three primary treatment events performed during resuscitation in addition to ventilation of the newborn. *Uncovered* describes how much of the newborn covered by a blanket, this is considered an important information, as covering more of the newborn will result in a lower heat loss. The fourth category is all other activities that are considered as relevant for the treatment. This can for example be clamping of umbilical cord and injections.

The heart rate sensor is sometimes detached and later reattached during a resuscitation episode. As this will contribute to artifact's and missing data in the dataset, the author has manually annotated attachment of the heart rate sensor. Only time regions where the heart rate sensor is fully attached to the newborn will be used in the analysis.

### B. Dataset

Two episodes were excluded due to corrupted data. The dataset used in this work therefore consists of 74 episodes of newborn resuscitation, called $D1$ in this paper.

The example signals, seen in Figure 2, include two distinct areas where some activity clearly is happening. VuDetector identifies both areas in the timeline representation, shown as *detected activity*. The manual annotations, does however only recognize one of the regions as a stimulation activity.

Newborn resuscitation often involves multiple health care providers, resulting in multiple activities being performed at the same point in time. A data subset is created to study regions where only one activity is being performed on the newborn. This data subset, $D2$, consists of regions in the resuscitation where VuDetector has identified an activity, and where either only stimulation or no therapeutic activity are manually annotated. As the detected activity regions will not overlap perfectly with the annotated data, non-overlapping regions are removed. The subset, $D2$, consists of 15,958 time points of stimulation and 3653 time points of non-therapeutic

[3]https://shop.aap.org/helping-babies-breathe-2nd-ed-action-plan-wall-poster-paperback/

TABLE I
OVERVIEW OF THE DATA SUBSETS USED IN EXPERIMENTS. ALL SUBSETS IS BASED ON 74 EPISODES OF NEWBORN RESUSCITATION.

| Data set | Method | Inclusion criteria | Total duration of data set |
|---|---|---|---|
| $D1$ | Sliding window | Full episodes | 21,830 s |
| $D2$ | Sliding window | VuDetect and manually annotated stimulation or hold off | 1,961 s |
| $D3$ | Blocks of variable size | VuDetect and manually annotated stimulation or hold off | 1,961 s |

activities. Grouping these based on the manual annotations, we obtain $D3$ with 464 regions with stimulation, and 357 regions with non-therapeutic activities. An overview of the data subsets are shown in table I.

## III. ACTIVITY RECOGNITION SYSTEM

A block diagram of the proposed system for detecting and recognizing regions with activities during newborn resuscitation is shown in Figure 3. An example graphical user interface (GUI) of the system is shown in Figure 5. The system takes raw input from the NeoBeat prototype, do necessary processing and classifications on the signals, and present a timeline to the user describing what events occurs at various times during the resuscitation. The analysis is designed to be able to run in real time during resuscitation, or on request to obtain more details of a given resuscitation episode at a later time. To obtain a high resolution in the classified activity and allow for real time operation, stimulation is classified as time series signal of 10Hz. Where the features at index $i$ are causally extracted from index $i-k+1$ to $i$, where $k$ is the window size. As ventilation and stimulation activities can occur at the same time, the two classes are differentiated when presented to the user. The GUI, Figure 5, shows the manually annotated data in green and the classified annotation in cyan. When run on new unannotated data, only the cyan timelines will be visible to the user.

The following subsections will describe in more detail the various parts of the system, shown in Figure 3.

### A. Activity detection

Detection of time periods during resuscitation where activities are likely to be performed on the newborn has previously been proposed by our research group, VuDetector [14]. The method detects time regions based on the short time energy (STE) of the acceleration energy signal. The acceleration energy, $Acc(n)$ is found by

$$Acc(n) = \sqrt{Acc_x^2(n) + Acc_y^2(n) + Acc_z^2(n)} \tag{1}$$

Where $Acc_d$ is a low pass filtered version of the measured acceleration in axis $d \in \{x, y, z\}$, and $n$ is the index in the acceleration signal. The STE, $E(i)$, is then found by

$$E_{Acc}(i) = \sum_{n=i-N+1}^{i} \left(Acc(n) \cdot w\left(i-n\right)\right)^2 \tag{2}$$

Where the STE at index $i$ is computed using samples from the window of length $N$. The STE is thresholded to determine if an activity occurs at the current window. The method achieves a sensitivity of 90% and a specificity of 80%, both with a standard deviation of 6%. More details of VuDetector can be found in [14].

### B. Detection of ventilation

A method for detecting ventilation during newborn resuscitation based on the measured pressure signal in the BMR has previously been proposed by our research group with an accuracy of 95%, VuVentilation [8]. As ventilation and stimulation events can occur at the same time, ventilation is not taken into account when trying to recognize stimulation.

### C. Preprocessing of data

The recorded ECG signal is susceptible to noise from the power grid, and is therefore first filtered using a 50Hz notch filter prior to any analysis. In addition, the QRS wave amplitude is affected both by the condition of the heart as well as the sensor placement. Variations in the amplitude due to inferior sensor placement are not desired, and the ECG signal is therefore normalized based on the median R-height of the signal. R-waves in the ECG signal are found using a discrete wavelet transform with the *sym4* wavelet. The signal is then normalized using the formula:

$$ECG_{norm}(m) = \frac{ECG(m)}{\mathrm{median}(R_h)} \tag{3}$$

Where $R_h$ is a vector containing the height of the detected R-waves, and $m$ the index in the ECG signal.

### D. Feature extraction

The proposed system utilize a subset of the ECG and acceleration features used by VuClassifier [15]. An overview of the features used in VuClassifier can be found in column *fSet*1in table II. Features in the time domain, such as energy, RMS, entropy, and in the frequency domain, wavelet, were defined for both the acceleration and ECG signals by Vu [15]. The wavelet features were extracted using a 6-level decomposition using the Daubechies mother wavelet. These features are denoted $Ea$ for the energy corresponding to the approximation and $D1-D6$ for the energy corresponding to the detail at each level. More details of these features can be found in [15].

VuClassifier [15] considers the entire region detected by VuDetector as a single class. This approach will introduce unwanted misclassifications in all cases where the detected activity region does not perfectly match the true stimulation activity in both time and length. As two or more events performed by the health care workers are unlikely to have the same duration, this approach also extracts features from windows at various lengths throughout each episode which is also not desirable. To handle these challenges, a sliding window of fixed size is introduced. Initial experiments were conducted using various length of the sliding window, with only minor differences between window sizes. A sliding window of 1 s, with a 900 ms overlap were chosen to achieve a

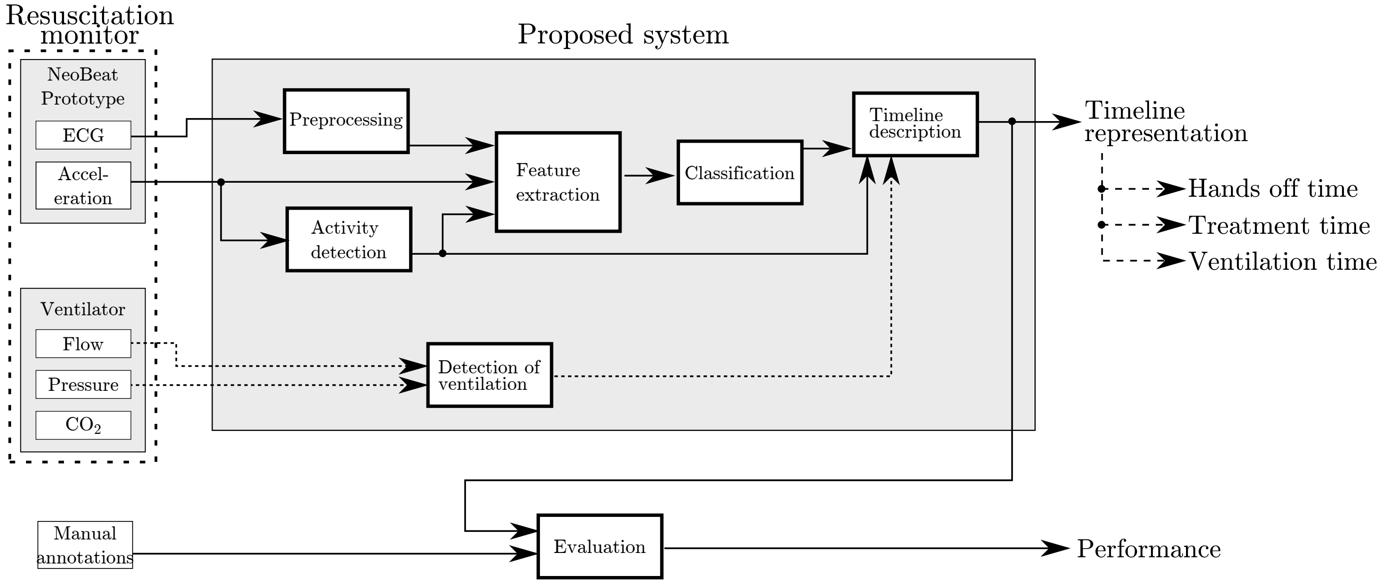


Fig. 3. Block diagram of the proposed systems. Using raw inputs directly from the Laerdal NeoBeat prototype, a timeline describing the resuscitation are computed and presented to the user for further analysis of the event. The dotted line to and from the block *detection of ventilation* illustrates an expansion of the system using a previously proposed method from our research group.

high resolution in the classification, so that all feature values are recalculated every 0.1 s, and as such can be seen as a function of a time index, i, at a sample rate of 10 Hz.

Based on visual observation of the resuscitation activities applied by the health care personnel, it is clear that stimulation often contains some repetitive movements, i.e. rubbing the back of the newborn. We therefore consider the accelerometer signals to be the most important signals for describing this repetitive movement. To represent these movements in the analysis, three new features are defined for each axes in the accelerometer signal. Resulting in a total of 9 new features. Let $P^i_{Acc,d}(f)$ denote the Short Time Fourier Transform (STFT) of window $i$ in the accelerometer signal in axis $d \in \{x, y, z\}$ as a function of the frequency, $f$. The first feature, $A_{Pmax,d}$, describes the maximum amplitude in the frequency domain,

$$A_{Pmax,d}\,(i) = \max\left(P^i_{Acc,d}(f)\right) \tag{4}$$

The second feature, $A_{f,d}(i)$, describes the frequency this maximum occurs at, according to

$$A_{f,d}(i) = \text{argmax}\left(P^i_{Acc,d}\,(f)\right) \tag{5}$$

The third feature describes the highest frequency with an amplitude above a set threshold, given by

$$A_{fT,d}(i) = \max\{f : P^i_{Acc,d}(f) > T\} \tag{6}$$

An overview of the STFT features are seen in column *fSet*4in table II.

### E. Classification

Initial tests were conducted using classifiers such as Naive Bayes, SVM [17], and RUSBoost [18]. Due to only minor differences in performance, the Naive Bayes classifier is used throughout the experiments due to its low computational complexity. The classifier is designed to distinguish between stimulation and non-stimulation activities. Discrimination between these two classes are conducted on all regions identified by VuDetector.

## IV. Experiments

Three experiments were designed. The first experiment was conducted to validate the previously published activity classifier. In the second experiment, all features were computed using a sliding window of fixed size. New features are added, and a nested cross-validation with feature selection are conducted to illustrate the performance which is possible. A reduced feature set is then found using feature selection. The final experiment studies performance of NBstim on full episodes, and how post processing can be utilized to increase the performance. The first experiment utilize the dataset *D*3, the second experiment uses the corresponding points extracted using a sliding window defined in *D*2. The third experiment utilize *D*1of 74 resuscitation episodes of up to 7 minutes each.

### A. Experiment 1: validation of previous work

In VuClassifier [15] the activities were divided into three classes, 1) *chest compression*, 2) *stimulation*, and 3) *other*. The first two classes were proposed to be combined to obtain a classification of treatment versus non-treatment. In the present work, we focused on distinguishing stimulation from non-stimulation activities, thus the two first classes were combined, and considered stimulation. The third class was interpreted as all non-stimulation activities. Validation of VuClassifier was conducted using the dataset *D*3.

### B. Experiment 2: improvement of activity classifier

To improve the usability of an improved version of the classifier, and facilitate for real-time classification, only causal features were implemented. The feature set *fSet*1 include

TABLE II
OVERVIEW OF ALL FEATURES. THE DASHED LINE SEPARATE FEATURES FROM VU [15] AND NEW FEATURES PROPOSED IN THIS WORK FOR THE ACCELEROMETER SIGNALS.

| Feature number | | Feature name | Description | fSet 1, VuClassifier [15] | fSet 2, Vu window | fSet 3, Vu reduced | fSet 4, STFT features | fSet 5, Final set |
|---|---|---|---|---|---|---|---|---|
| 1 | Acceleration signal | $\mathrm{A}_{te}(i)$ | Total energy | 1 | 1 | 1 | - | 1 |
| 2 | | $A_{globalMax}(i)$ | Max acceleration value in episode | 1 | - | - | - | - |
| 3 | | $A_{Ea}(i)$ | Ea, energy approximation | 1 | 1 | 1 | - | 1 |
| 4 | | $A_{Ed1}(i)$ | Ed1, energy detail, level 1 | 1 | 1 | 1 | - | 1 |
| 5 | | $A_{Ed2}(i)$ | Ed2, energy detail, level 2 | 1 | 1 | - | - | - |
| 6 | | $A_{Ed3}(i)$ | Ed3, energy detail, level 3 | 1 | 1 | - | - | - |
| 7 | | $A_{Ed4}(i)$ | Ed4, energy detail, level 4 | 1 | 1 | - | - | - |
| 8 | | $A_{Ed5}(i)$ | Ed5, energy detail, level 5 | 1 | 1 | - | - | - |
| 9 | | $A_{Ed6}(i)$ | Ed6, energy detail, level 6 | 1 | 1 | - | - | - |
| 10 | | $A_{autocorr}(i)$ | Energy of the auto-correlation signal | 1 | 1 | 1 | - | 1 |
| 11 | | $A_{winMax}(i)$ | Max acceleration value in window | 1 | 1 | - | - | - |
| 22 | | $A_{valley}(i)$ | Valley energy | 1 | 1 | 1 | - | 1 |
| 13 | | $\overline{A}(i)$ | Mean of acceleration energy | 1 | 1 | - | - | - |
| 14 | | $\overline{A}_x(i)$ | Mean of acceleration$_x$ | 1 | 1 | - | - | - |
| 15 | | $\overline{A}_y(i)$ | Mean of acceleration$_y$ | 1 | 1 | 1 | - | 1 |
| 16 | | $\overline{A}_z(i)$ | Mean of acceleration$_z$ | 1 | 1 | - | - | - |
| 17 | | $A_\sigma(i)$ | Standard deviation of acc energy | 1 | 1 | 1 | - | - |
| 18 | | $A_{\sigma,x}(i)$ | Standard deviation of acc$_x$ | 1 | 1 | 1 | - | - |
| 19 | | $A_{\sigma,y}(i)$ | Standard deviation of acc$_y$ | 1 | 1 | 1 | - | - |
| 20 | | $A_{\sigma,z}(i)$ | Standard deviation of acc$_z$ | 1 | 1 | - | - | - |
| 21 | | $A_e(i)$ | Entropy of acc energy | 1 | 1 | - | - | - |
| 22 | | $A_{H,x}(i)$ | Entropy of acc$_x$ | 1 | 1 | 1 | - | 1 |
| 23 | | $A_{H,y}(i)$ | Entropy of acc$_y$ | 1 | 1 | - | - | - |
| 24 | | $A_{H,z}(i)$ | Entropy of acc$_z$ | 1 | 1 | 1 | - | - |
| 25 | | $A_{RMS,e}(i)$ | RMS of acc energy | 1 | 1 | - | - | - |
| 26 | | $A_{RMS,x}(i)$ | RMS of acc$_x$ | 1 | 1 | - | - | - |
| 27 | | $A_{RMS,y}(i)$ | RMS of acc$_y$ | 1 | 1 | 1 | - | - |
| 28 | | $A_{RMS,z}(i)$ | RMS of acc$_z$ | 1 | 1 | 1 | - | 1 |
| 29 | | $A_{corr,xy}(i)$ | Correlation, acc$_x$, acc$_y$ | 1 | 1 | 1 | - | 1 |
| 30 | | $A_{corr,xz}(i)$ | Correlation, acc$_y$, acc$_z$ | 1 | 1 | 1 | - | 1 |
| 31 | | $A_{corr,yz}(i)$ | Correlation, acc$_y$, acc$_z$ | 1 | 1 | 1 | - | 1 |
| 32 | | $A_{Fmax,x}(i)$ | Highest amplitude in STFT x-axis | - | - | - | 1 | 1 |
| 33 | | $A_{Fmax,y}(i)$ | Highest amplitude in STFT y-axis | - | - | - | 1 | 1 |
| 34 | | $A_{Fmax,z}(i)$ | Highest amplitude in STFT z-axis | - | - | - | 1 | 1 |
| 35 | | $A_{F,x}(i)$ | Frequency of peak amplitude, x-axis | - | - | - | 1 | 1 |
| 36 | | $A_{F,y}(i)$ | Frequency of peak amplitude, y-axis | - | - | - | 1 | - |
| 37 | | $A_{F,z}(i)$ | Frequency of peak amplitude, z-axis | - | - | - | 1 | 1 |
| 38 | | $A_{Fthresh,x}(i)$ | Highest frequency over a threshold, x-axis | - | - | - | 1 | - |
| 39 | | $A_{Fthresh,y}(i)$ | Highest frequency over a threshold, y-axis | - | - | - | 1 | 1 |
| 40 | | $A_{Fthresh,z}(i)$ | Highest frequency over a threshold, z-axis | - | - | - | 1 | 1 |
| 41 | ECG signal | $ECG_{te}(i)$ | Total energy | 1 | 1 | - | - | - |
| 42 | | $ECG_{globalMax}(i)$ | Max ecg value in episode | 1 | - | - | - | - |
| 43 | | $ECG_{Ea}(i)$ | Ea, energy approximation | 1 | 1 | 1 | - | 1 |
| 44 | | $ECG_{Ed1}(i)$ | Ed1, detail, level 1 | 1 | 1 | 1 | - | - |
| 45 | | $ECG_{Ed2}(i)$ | Ed2, energy detail, level 2 | 1 | 1 | 1 | - | - |
| 46 | | $ECG_{Ed3}(i)$ | Ed3, energy detail, level 3 | 1 | 1 | 1 | - | - |
| 47 | | $ECG_{Ed4}(i)$ | Ed4, energy detail, level 4 | 1 | 1 | 1 | - | 1 |
| 48 | | $ECG_{Ed5}(i)$ | Ed5, energy detail, level 5 | 1 | 1 | - | - | - |
| 49 | | $ECG_{Ed6}(i)$ | Ed6, energy detail, level 6 | 1 | 1 | - | - | - |
| 50 | | $ECG_{corr}(i)$ | Energy of the auto-correlation signal | 1 | 1 | - | - | - |
| 51 | | $\overline{ECG}_(i)$ | Mean value of ECG | 1 | 1 | 1 | - | 1 |
| 52 | | $ECG_\sigma(i)$ | Standard deviation of ecg | 1 | 1 | - | - | - |
| 53 | | $ECG_E(i)$ | Entropy of the ECG | 1 | 1 | - | - | - |
| 54 | | $ECG_{RMS}(i)$ | RMS of the ECG | 1 | 1 | 1 | - | 1 |
| 55 | | $ECG_{winMax}(i)$ | Max ECG value in window | 1 | 1 | 1 | - | 1 |

the two features, $ECG_{globalMax}$ and $A_{globalMax}$, which are computed using entire episodes, and were therefore omitted.

Selecting a smaller feature subset, with a high performance, from a larger set can be achieved using multiple approaches. Exhaustive search is rarely used in datasets with many features due to the heavy computational cost related to do validation of every possible feature combination in a dataset. A common approach is to use a greedy method, such as a forward selection or backward elimination, as they are fast and robust against overfitting [19]. A wrapper based nested cross-validation [20] with a modified feed forward approach, where the 5 best features in an iteration was used in this work to determine the best feature combination in the next iteration.

The nested cross-validation with feature extraction scheme is used to determine the performance which can be obtained using a feature subset. A new feature selection is then conducted to identify the optimal subset, *fSet*3, from *fSet*2. The new feature set, *fSet*4, are then be included. A second round of nested cross-validation and feature selection will be performed to identify the potential performance and the feature set, *fSet*5, from *fSet*3 ∪ *fSet*4.

### C. Experiment 3: full episodes

As the proposed system is designed to annotate full episodes, it is important to present the performance which the end user will see. As a result of this, the performance on full episodes are computed using a leave-one-out validation on full episodes using the feature set *fSet*5.

The complete system classifies stimulation with a resolution of $10Hz$. It is however a reasonable assumption that activities performed by health care workers do not change at such a speed, nor do they last as short time. By taking these two factors into consideration, a post processing scheme were introduced, with the potential of eliminating short segments where the activity was misclassified.

One of the most basic post processing schemes consists of doing a majority voting within a detected activity region, this approach has the same challenges as VuClassifier [15], and will therefore not be considered for further analysis. As one of the challenges of this approach is misclassifications at the borders, a second post processing scheme consists of classifying the edge regions alone, while leaving the bigger middle section to be classified as a single activity. This scheme can solve the edge problem, but determining the ideal size of these edges can pose a challenge. An alternative processing scheme is based on the idea that the detected activity region could include one or more areas with actual stimulation. The change between stimulation and non-stimulation activities should, however, not be able to change as fast as original classification. This post processing scheme can easily be implemented using a median filter on the classified timeline.

## V. Results

### A. Experiment 1

The VuClassifier was trained on the original dataset described in [15], and then used to classify the subset $D3$. The performance can be seen in table III, and the used features can be seen in column *fSet*1in table II.

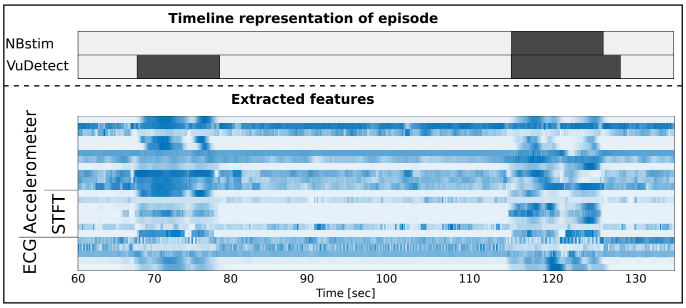


Fig. 4. Overview of the computed feature values for the same time period illustrated in Figure 2 using the 23 final features, *fSet*5. Each row in the heat map illustrate the values of a given feature over time. A darker color indicates a higher value. The feature order is the same as shown in table IV. All features are normalized in the region $[0, 1]$ for visualization purposes. NBstim evaluates the regions found by VuDetect, and classify them as stimulation or non-stimulation activities.

TABLE III
PERFORMANCE USING THE FEATURES AND CLASSIFIER PROPOSED BY VU ET AL.ON BLOCKS WITH VARIABLE LENGTH FOUND AS TIME SEGMENTS WHERE AN ACTIVITY IS FOUND BY VUDETECTOR AND EITHER STIMULATION OR NO ACTIVITY IS MANUALLY ANNOTATED, $D3$

| Method | #features | Sensitivity | Specificity | Accuracy |
|---|---|---|---|---|
| VuClassifier [15] | 46 | 88.4% | 1.7% | 50.7% |

TABLE IV
PERFORMANCE OF VARIOUS FEATURE SETS COMPUTED USING SLIDING WINDOW ON TIME POINTS WHERE AN ACTIVITY IS FOUND BY VUDETECTOR AND EITHER STIMULATION OF NO ACTIVITY IS MANUALLY ANNOTATED, $D2$. THE PERFORMANCE IS COMPUTED USING A 3-FOLD NESTED CROSS-VALIDATION

| Feature set | #features | Sensitivity | Specificity | Accuracy |
|---|---|---|---|---|
| *fSet*2 | 44 | 61.7% | 65.5% | 61.7% |
| *fSet*3 | 24 | 63.3% | 65.5% | 63.7% |
| *fSet*4 | 9 | 75.6% | 43.7% | 69.7% |
| *fSet*3∪ *fSet*4 | 33 | 57.4% | 69.8% | 59.7% |
| *fSet*5 | 23 | 67.3% | 62.1% | 66.4% |

### B. Experiment 2

The performance using all 44 causal window based features in *fSet*2 on the dataset $D2$, can be seen in table IV. The 44 features are then reduced to 24 features using feature selection. The chosen features can be seen in column *fSet*3in table II, and the performance of the reduced feature set in table IV. A feature extraction is conducted the feature subset consisting of *fSet*3∪ *fSet*4, resulting in *fSet*5as seen in table II. Performance for each subset is computed using a nested cross-validation with feature extraction. The performance of these three feature sets are seen in table IV.

A visualization of the computed values for the final feature set, *fSet*5, is shown in Figure 4. Each row corresponds to a given feature, and a darker color indicate a higher value in the computed feature value. For visualization purposes, all features are normalized to $[0, 1]$.

### C. Experiment 3

The performance when distinguishing between stimulation and non-stimulation activities in entire resuscitation episodes, $D1$, with and without a post processing scheme are shown in table V. A leave-one-out cross-validation is used, and the

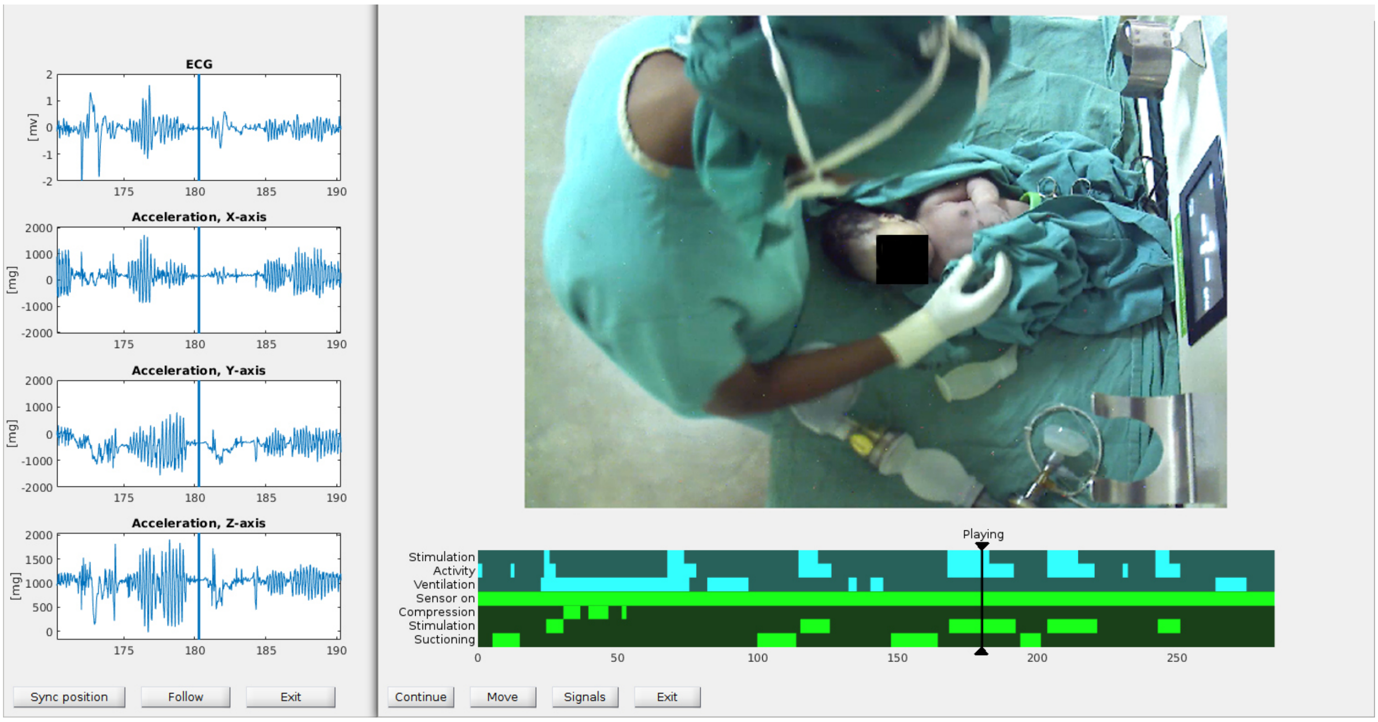


Fig. 5. Example graphical user interface for the proposed system. In this example, the timeline includes manually annotated data, shown in green, and automatically classified data, shown in cyan. ECG and acceleration signals are included to visualize the current measurements to the user.

TABLE V
PERFORMANCE OF NBSTIM WHEN DISTINGUISHING BETWEEN STIMULATION AND NON-STIMULATION IN FULL RESUSCITATION EPISODES WITH A MAXIMUM LENGTH OF 7 MINUTES, $D1$

| Method | Sensitivity | Specificity | Accuracy |
|---|---|---|---|
| No postprocessing | 68.3% | 93.1% | 88.6% |
| Median filtering | 69.2% | 94.8% | 90.3% |

classified timeline is compared to when stimulation or no-stimulation is manually annotated.

## VI. Discussion

Validation of VuClassifier achieves a high sensitivity and a low specificity in distinguishing between stimulation and non-stimulation events on the data set annotated by a neonatologist, more details in section II-A. This performance does not correspond to the accuracy of 79.8%, sentivity of 84% and specificity of 72.6% reported in Vu et al. [15]. The degradation in performance could be a result of how the new dataset is defined. The increased size should not affect performance, but how manual annotations are found may result in a change. In the original publication, Vu et al [15], the data was annotated by the author using several additional categories, and may therefore differ from how a trained clinician would annotate it.

By combining features in *fSet*1with a sliding window, *fSet*2, and a simple classifier, an accuracy of 61.7% is achieved, with a sensitivity and a specificity both above 60%. By adding more features to a system, the performance does not necessarily increase. The final feature set, *fSet*5, found using a feature selection approach, outperforms the original feature set,*fSet*2.

The performance when using the STFT features, *fSet*4, to distinguish between stimulation and non-stimulation activities achieve a high sensitivity and accuracy. The specificity is, however, reduced compared to the reduced feature set from Vu et al [15], *fSet*3. The final subset, *fSet*5, consisting of 23 features, achieves a sensitivity and accuracy which is lower, but the specificity has a large increase compared to using only the new features. While correct identification of the stimulation events are important, we want to keep the false positive rate as low as possible to better indicate how much stimulation is actually given during the resuscitation.

Using the obtained feature subset, *fSet*5, on complete episodes, $D1$, a large increase in performance is seen with the accuracy increasing from 66.4% in table IV to 88.6% in table V. This increase is a result of complete resuscitation episodes often include large time periods where no stimulation occurs. As many of these periods are not identified as interesting by VuDetector, the specificity will increase. By applying a median filter post processing scheme, the accuracy is further increased to 90.3%.

While the addition of new STFT features may be considered a small expansion of previous work, we consider the proposed causal system with a reduced feature set to be an important step towards utilizing this work for automatic annotation of newborn stimulation. The new features were proposed to describe repetitive movement observed when health care personnel performed stimulation, i.e. rubbing the back of the newborn. In the visualization of the final feature set, *fSet*5, we can see some differences between the feature values in the stimulation and non-stimulation regions. It is, however, challenging to determine how each feature is physiological linked to the performed activity.

The signal composition, consisting of accelerometer and dry electrode ECG, as available from the NeoBeat has never been available before. A method utilizing these signals to automatically annotate when stimulation is being performed during resuscitation can greatly impact future studies of how stimulation activities affect the resuscitation process and newborn outcome. With such a system, like we propose with NBstim, studies exploring how stimulation activities affects the resuscitation process will no longer be limited to using a low number of manually annotated data, but information from larger data sets can be extracted. Statistics on how guidelines are followed can be extracted. It will also allow to extract information of the duration of stimulation activities during a resuscitation. Was it a continuous stimulation or multiple? When the newborn was not ventilated, was that due to hands-off time or stimulation? Such information may be vital when exploring how resuscitation outcomes are correlated with the stimulation activities during resuscitation. In a future scenario we might have NeoBeat available in some hospitals, video signals in other hospitals, ventilation data from some hospitals, or several of these components. We are planning to fuse the output of the NBstim algorithm with the output of the automated video analysis [9], and potentially output from ventilator signals [8] to produce more reliable timelines of activities, also including ventilation and suction when possible.

For further quality assurance and truth marking for validation, an interface similar to what is shown in Figure 5 can be used as an interactive annotation tool with the automatic detections as a first step, and with the option of refining these manually if needed.

### A. Limitations

Due to the small data set of only 74 resuscitations episodes, a total of 21,830s, the proposed system may can be seen as a feasibility study of the possibility of annotating stimulation based on the measured accelerometer and ECG signals. When identifying the feature set, a smaller subset of only 1,961s is utilized. This reduction is performed by only including time periods where some movement occur, and VuDetector identifies the movement as an activity, and either only stimulation or no therapeutic activity is performed. The advantage of using this smaller subset for the feature selection is that the method will identify features which are crucial in distinguishing stimulation and non-stimulation activities instead of focusing on patterns from other activities. Because of this limited data set with ground truth, further validation is required before applying the method in clinical practice.

## VII. CONCLUSION

In this work, we present a complete system for automatic identification of stimulation during newborn resuscitation. The system consists of an activity detector, and the proposed NBstim classifier with 23 features, 18 from the 100 Hz accelerometer signals in $X$,$Y$, and $Z$-directions and 5 from the 500 Hz dry-electrode ECG signal. Features are computed using a sliding window of 1 s with 900 ms overlap. NBstim achieves a high performance, with an accuracy of 90.3% in identifying stimulation, and could therefore be used as a replacement of time consuming manual annotation, or as an initial step in an interactive tool. The ultimate objective is to save lives at birth, and more specifically by studying what activities are performed by health care providers during resuscitation of asphyxiated newborns, if guidelines are followed, and if current guidelines are effective in saving lives.

The system can be used with the newly released Laerdal NeoBeat Newborn Heart Rate Meter, but a validation using a larger data set is required before implementing the method in clinical practice. In the Safer Births project, we are currently working on expand our data collection of newborn resuscitation, and we want to increase the number of manually annotated data.

In future work, we want to utilize NBstim for creating timelines for thousands of newborn resuscitation episodes. In combination with the immediate and 24-h outcome, available in the Safer Births project, we can extract vital statistics and potentially get a greater understanding of how stimulation activities affect resuscitation procedures and newborn outcomes.

## VIII. CONFLICT OF INTEREST

This work is part of the Safer Births project which has received funding from Laerdal Foundation, Laerdal Global Health, Skattefunn, Norwegian Ministry of Education and USAID. The work was partly supported by the Research Council of Norway through the Global Health and Vaccination Programme (GLOBVAC) project number 228203.

Ingunn Anda Haug is an employee of Laerdal Medical AS.

## REFERENCES

[1] S. Wright *et al.*, "Save the children: Ending newborn deaths: Ensuring every baby survives."

[2] H. Wang *et al.*, "Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the global burden of disease study 2013," *The Lancet*, vol. 384, no. 9947, pp. 957–979, 2014.

[3] WHO, "Guidelines on basic newborn resuscitation," World Health Organization, 2012.

[4] J. M. Perlman *et al.*, "Part 7: neonatal resuscitation: 2015 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations," *Circulation*, vol. 132, no. 16_suppl_1, pp. S204–S241, 2015.

[5] H. L. Ersdal, E. Mduma, E. Svensen, and J. M. Perlman, "Early initiation of basic resuscitation interventions including face mask ventilation may reduce birth asphyxia related mortality in low-income countries: a prospective descriptive observational study," *Resuscitation*, vol. 83, no. 7, pp. 869–873, 2012.

[6] M. Iriondo, M. Thió, E. Burón, E. Salguero, J. Aguayo, M. Vento, and N. R. G. N. of the Spanish Neonatal Society (SEN), "A survey of neonatal resuscitation in spain: gaps between guidelines and practice," *Acta paediatrica*, vol. 98, no. 5, pp. 786–791, 2009.

[7] J. Urdal *et al.*, "Signal processing and classification for identification of clinically important parameters during neonatal resuscitation," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Sept 2017, pp. 547–552.

[8] H. Vu *et al.*, "Automatic detection and parameterization of manual bag-mask ventilation on newborns," *IEEE journal of biomedical and health informatics*, 2016.

[9] Ø. Meinich-Bache, K. Engan, I. Austvoll, T. Eftestøl, H. Myklebust, L. Yarrot, H. Kidanto, and H. Ersdal, "Object detection during newborn resuscitation activities," *IEEE journal of biomedical and health informatics*, 2019.

[10] S. H. Lee *et al.*, "A study on the activity classification using a triaxial accelerometer," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, vol. 3, Sept 2003, pp. 2941–2943 Vol.3.

[11] P. Gupta and T. Dallas, “Feature selection and activity recognition system using a single triaxial accelerometer,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1780–1786, June 2014.

[12] F. Yang and L. Zhang, “Real-time human activity classification by accelerometer embedded wearable devices,” in *2017 4th International Conference on Systems and Informatics (ICSAI)*, Nov 2017, pp. 469–473.

[13] A. B. De Luna, V. N. Batchvarov, and M. Malik, “The morphology of the electrocardiogram,” *The ESC Textbook of Cardiovascular Medicine Blackwell Publishing*, pp. 1–35, 2006.

[14] H. Vu *et al.*, “Detection of activities during newborn resuscitation based on short-time energy of acceleration signal,” in *International Conference on Image and Signal Processing*. Springer, 2016, pp. 262–270.

[15] ——, “Automatic classification of resuscitation activities on birth-asphyxiated newborns using acceleration and ECG signals,” *Biomedical Signal Processing and Control*, vol. 36, pp. 20 – 26, 2017.

[16] ——, “Exploring the relationship between characteristics of ventilation performance and response of newborns during resuscitation,” in *International Joint Conference on Biomedical Engineering Systems and Technologies*. Springer, 2015, pp. 275–290.

[17] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[18] C. Seiffert *et al.*, “Rusboost: A hybrid approach to alleviating class imbalance,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 185–197, 2010.

[19] I. Guyon and A. Elisseeff, “An introduction to feature extraction,” in *Feature extraction*. Springer, 2006, pp. 1–25.

[20] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artificial intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.