Sleep Versus Wake Classification From Heart Rate Variability Using Computational Intelligence: Consideration of Rejection in Classification Models

Aaron Lewicke*, *Member, IEEE*, Edward Sazonov, *Member, IEEE*, Michael J. Corwin, Michael Neuman, Stephanie Schuckers, *Senior Member, IEEE*, and CHIME Study Group

***Abstract*—Reliability of classification performance is important for many biomedical applications. A classification model which considers reliability in the development of the model such that unreliable segments are rejected would be useful, particularly, in large biomedical data sets. This approach is demonstrated in the development of a technique to reliably determine sleep and wake using only the electrocardiogram (ECG) of infants. Typically, sleep state scoring is a time consuming task in which sleep states are manually derived from many physiological signals. The method was tested with simultaneous 8-h ECG and polysomnogram (PSG) determined sleep scores from 190 infants enrolled in the collaborative home infant monitoring evaluation (CHIME) study. Learning vector quantization (LVQ) neural network, multilayer perceptron (MLP) neural network, and support vector machines (SVMs) are tested as the classifiers. After systematic rejection of difficult to classify segments, the models can achieve 85%–87% correct classification while rejecting only 30% of the data. This corresponds to a Kappa statistic of 0.65–0.68. With rejection, accuracy improves by about 8% over a model without rejection. Additionally, the impact of the PSG scored indeterminate state epochs is analyzed. The advantages of a reliable sleep/wake classifier based only on ECG include high accuracy, simplicity of use, and low intrusiveness. Reliability of the classification can be built directly in the model, such that unreliable segments are rejected.**

***Index Terms*—Heart rate variability, infants, reliability, sleep/wake classification.**

I. Introduction

AUTOMATIC computer classification is an important tool in biomedical research, particularly, in the analysis of large data sets where manual classification is not feasible. Applications which have a large data set available for decision making may not require that each sample be classified. In addition, in some real-time applications, a decision may not be life-threatening such that a delay in making a decision can be tolerated. In this paper, we propose to incorporate reliability in the classification model in its design and training in order to improve classification performance. This approach does not require that each sample is classified. Instead, the model is trained to isolate and reject unreliable samples. This paper utilizes rejection in a classification model developed for classifying sleep/wake states in infants utilizing only the electrocardiogram.

II. Background

The classic technique of determining sleep state includes recording three different electrical signals: brain activity (EEG), eye movements (EOG), and muscle activity from the chin (EMG). The recorded activity is then classified into three major states: awake (AWK), nonrapid eye movement (non-REM) or quiet sleep (QS), and rapid eye movement (REM) or active sleep (AS). The QS state is characterized by a slowing down in brain activity, breathing, heart rates, body temperature, metabolic activity, muscle movement, and a decline in blood pressure. Unlike QS, AS can be characterized as an active brain in a paralyzed body. During AS, breathing and heart rate become irregular, blood flow to the brain increases, and body temperature increases. Aside from occasional twitches, body muscles are paralyzed on the level of the brainstem to prevent "acting out" of dreams that occur during AS.

Sleep scoring via an automatic approach has been studied extensively. Although many promising results have been obtained, the EEG signal has been used [1], [3], [10], [13], [15], [17], [19], [22], [30], [32], [34], [33]. This study considers development of an automated algorithm using only RR interval derived from ECG. Very little work has been performed using only HRV, but calculation of quiet/active sleep using cardiorespiratory (ECG and respiration) parameters has been explored to determine sleep states [12], [14], [25], [37]. Haddad *et al.* [12] used cardiorespiratory variables to distinguish REM sleep from quiet sleep, requiring the knowledge that the infant was asleep. The study was based on the measurements of nine (training) and five (testing) infants at one and four months of age. Gold standard decisions of sleep state were based on visual interpretations of EEG, EOG, and EMG. The Kolmogorov–Smirnov distances (KSD) were calculated to analyze the separation between quiet sleep and REM sleep. The results of Haddad *et al.* show that coefficient of variation of respiratory cycle time separated the data the best. There were 85 5-min epochs staged as quiet sleep

Manuscript received June 18, 2006; revised April 3, 2007. *Asterisk indicates corresponding author.*

*A. T. Lewicke is with the Department of Electrical and Computer Engineering, Clarkson University, 8 Clarkson Ave., Potsdam, NY 13699 USA (e-mail: lewickat@clarkson.edu).

E. S. Sazonov and S. A. C. Schuckers are with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699 USA (e-mail: esazonov@clarkson.edu; sschucke@clarkson.edu).

M. J. Corwin is with the Department of Pediatrics, Boston Medical Center, Boston University School of Medicine, Boston, MA 02215 USA and also with the BattelleCRO, Inc., Newton, MA 02459 USA(e-mail: mjcorwin1@carestat.com).

M. R. Neuman is with the Department of Biomedical Engineering, Michigan Technological University, Houghton, MN 49931-1295 USA (e-mail: mneuman@mtu.edu).

Digital Object Identifier 10.1109/TBME.2007.900558

and 85 staged as REM sleep; 79 of the quiet sleep epochs were correctly classified as quiet sleep (93%); 84 of the REM epochs were correctly classified as REM (∼99%). The same boundary was used for each age group with similar results. Harper *et al.* [14] used cardiorespiratory measures to determine sleep state by discriminant analysis. The measures calculated were median heart and respiratory rate, heart and respiratory interquartile range, and peak-to-trough amplitude for RR interval variation in two low-frequency bands. This study was performed on infants at one week and at one, two, three, four, and six months of life. A different model was created for each age (one, two, three, four, and six months) and had an overall correct classification of 84.8% [14]. Using only cardiac information decreased the classification to 82.0%. Using only respiratory measures decreased the classification to 80.0%.

Few researchers have studied sleep classification from electrocardiogram alone. Welch *et al.* used beat-to-beat heart rate data from eight healthy, young male adults [37] divided into 128 beat epochs (3200 total epochs) to classify sleep stages. Discriminant analysis and Bayes classifier models were made for each of the individuals. Each subject had two nights of sleep recorded and the classification model was trained on 50% of the epochs from the first night. Accuracy ranged from 21% to 77% from the two nights of sleep. Lisenby *et al.* have shown that analyzing heart rate in the heart beat domain and the Fourier transform of the heart beat domain one can classify REM and NREM sleep to an average of 80% accuracy using correlation coefficients [25]. The study used nine normal adults from the same study as Welch. Nason *et al.* used wavelet packet modeling to achieve 75%–90% correct sleep/wake classification on one infant using different classifiers for two, three, four, and five months of age [28]. The previous work shows that sleep classification can achieve results similar to sleep scoring by technicians using EEG. The difficulty with the previous studies is that only very small populations are used and models are created at different age groups. Also, the methods utilize linear classification methods which work well on linearly separable data, but do not work well on nonlinearly separable data or nonseparable data. This study tested the ability of nonlinear classifiers to classify infant sleep/wake from only heart rate variability over a large general population. The algorithm will allow analysis of large data sets that do not contain standard sleep scoring signals (EEG, EMG, and EOG).

Data used in this study are part of the CHIME National Institute of Health (NIH) study, which studied home infant monitors for apnea and bradycardia for over 1000 infants [18], [29], [7]. Sleep state information would be helpful in analyzing the over 700 000 h of data recorded on the home monitor where traditional polysomnogram (PSG) information is not available [7]. Because of the amount of data available, it is not required that a decision be made for every observation. Therefore, the reliability method can be used to reject observations by which a confident sleep classification could not be made.

Special attention is given to the reliability of sleep scoring, since reliability of sleep states may be important for further sleep studies. Reliability is improved by incorporating rejection into the model. By rejecting data points systematically, reliability can be improved by designing a model that rejects data points in the overlap region between two classes. This method is an automated way of determining which regions should be rejected based on training and validation sets. The proposed approach is valid when a significant amount of data is available for analysis. Previous work in rejection began with probabilistic analysis for finding theoretical optimal rejection rates versus error [16], [4]. Other researchers have applied rejection to different classifiers. De Stefano *et al.* have developed rejection techniques for both the learning vector quantization (LVQ) and multilayer perceptron (MLP) neural networks [8], Madevska-Bogdanova *et al.* modify SVM classifier outputs to construct a posterior probability from the distance of an epoch to the separating hyperplane [26].

III. Methods

A. Data

Data used in this study was collected from infants as part of the CHIME study [18]. An 8-h PSG from the CHIME data set was performed which includes simultaneous manually scored sleep/wake (30-s epoch) files. In addition to a full PSG, simultaneous recordings were made with the CHIME home monitor system (NIMS, Inc., Miami Beach, FL). The CHIME home monitor included an ECG (sampled at 1000 Hz to store RR intervals while raw ECG was stored at 200 Hz), respiratory inductance plethsmography sensor (Respitrace Plus, sampled at 50 Hz), a pulse oximeter (Ohmeda type MINX, Ohmeda Corp., Madison, WI, pulse sampled at 50 Hz and oxygen saturation sampled at 1 Hz), accelerometer-position sensor (sampled at 50 Hz), and transthoracic electrical impedance (Aequitron Inc., Plymouth, MN, sampled at 50 Hz) [29]. Trained technicians performed PSG-based sleep state identification for the infants at 30-s intervals (epochs) [7]. The PSG-identified sleep states (awake, active sleep, quiet sleep, and indeterminate) were used as response variables for training the classifiers. For this study, the active and quiet sleep states were combined into "sleep" state, the awake state remained "wake," and the indeterminate states were discarded during the model-training phase. Once the models were created, a data set of only indeterminate state epochs was tested to see what percentage of the indeterminate epochs would be rejected and how many errors would be added to the overall classification result.

A total of 190 infants, with available sleep state files and ECG recordings, were divided evenly to create training (64 infants), validation (63 infants), and testing (63 infants) sets, each with ∼57 000 epochs. The infant's anthropomorphic data from each data set is compared in Table I. PCA (weeks) and birth weight (grams) are given in terms of mean(std). The male, female, term, and preterm distributions are given in terms of percent of the data set.

Three classification models were compared: MLP neural network, LVQ neural network, and support vector machine (SVM). Each classifier had reliability incorporated into the model through rejection of epochs where sleep/wake cannot be determined with confidence [8], [26]. The neural networks and rejection thresholds are designed using the training and validation sets, while the SVM model used tenfold cross validation of the training set to decide the training parameters and then

TABLE I
ANTHROPOMORPHIC DATA OF INFANTS DISTRIBUTED AMONG TRAINING, VALIDATION, AND TESTING DATA SETS

	Training	Validation	Testing
PCA [weeks]	45.6 (7.4)	45.6 (7.3)	45.4 (7.2)
Birth Weight [grams]	2447.5 (1152.1)	2598.8 (1138.4)	2344.7 (1082.8)
Male [%]	59.4	49.2	50.8
Female [%]	40.6	50.8	49.2
Term [%]	42.2	58.7	57.1
Preterm [%]	57.8	41.3	42.8

included the validation set to decide the rejection threshold. All results were confirmed by an untouched testing data set. Details of the three classification models used are described in the next sections.

The training set had a total of 57 569 epochs. This is a very large number of data points and not all points were needed to create accurate models. A 95% confidence interval table was used to determine the size of the training set that would represent all of the data. In the 95% confidence interval table, only 55 infants were needed to represent the 64 in the training set. Also, each infant has an average of 900 epochs from which only 269 are needed. To ensure the models were unbiased towards a single class, it was necessary to have an equal number of epochs from wake and sleep states. Therefore, the training data set used in this study contained a total of 14 795 epochs with 7398 epochs from each class.

The PSG records of sleep or wake represent the response variable during each 30-s epoch. Utilization of lagged metrics as predictors was based on a simple reasoning that there could be no sudden change in the wake/sleep state. Similar lagged models were used in related studies on adults [5], [31].

The RR interval signal was preprocessed to create reliable data as follows.

1) An ECG artifact rejection routine was run to find artifactual heartbeats (noisy, missing, or extra beats) and correct the data. The artifact rejection routine compares each RR interval with the median of the 25 surrounding intervals and last accepted interval. If both differences are outside of $+/-20\%$ the interval is detected as an artifact. Corrections of missing intervals were made by cubic spline interpolation [38].
2) Both the beginning and end of the RR interval recording were synchronized in time with the PSG data on an epoch (30-s) boundary.
3) Standard HRV measures were extracted from the signal for each epoch. A total of 20 various metrics were computed including time domain metrics (mean, median, standard deviation, coefficient of variation, etc.—a total of ten different metrics), frequency domain metrics (power in the low- and high-frequency range, power ratio, and total power), time-frequency metrics (high- and low-frequency wavelet power and power ratio), and nonlinear metrics (scatter, regularity, and space-filling propensity).
4) A fuzzy C-means (FCM) clustering algorithm [39] was run to find which of the HRV metrics best separated the data into clusters of sleep and wake. The FCM algorithm utilizes unsupervised learning to create two clusters of similar data. Each data sample receives a membership value for each of the two output classes (sleep, wake). The sample was said to belong to the output class with the highest associated membership value. Fuzzy clustering was carried out through an iterative update of membership u_{ij} of each data point X_i and the cluster centers c_j by
$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|X_i-c_j\|}{\|X_i-c_k\|}\right)^{\frac{2}{m-1}}}$$
$$c_j = \frac{\displaystyle\sum_{i=1}^{N} u_{ij}^m X_i}{\sum_{i=1}^{N} u_{ij}^m}$$
where N is the number of data points and $m = 2$ defines the sharpness of the membership function. Specifics of the iterative algorithm for FCM clustering can be found in [39].
Many of the HRV measures were not expected to be good features because the epoch length used in sleep scoring was too short for the measures to develop. The best HRV measure from the FCM experiment was mean. For the initial classification experiments, the mean for nine consecutive epochs (30 s) was chosen as the input features. Utilizing lagged inputs gives the algorithm past information since it is not expected that the change is sleep/wake state will happen abruptly. Further work is needed to explore the usefulness of other epoch durations and HRV measures for classification of sleep/wake. The second best HRV measure is the standard deviation of normal-to-normal (SDNN) RR interval values. Combining SDNN with mean as input features is explored in Section IV-F.
5) The mean RR interval for the last nine consecutive 30-s epochs, formed $M_{-8}, \ldots, M_{-2}, M_{-1}, M_0$ as separate predictors. Thus, the input to the neural networks was a nine-element vector of epoch mean values [24], [23].

B. Multilayer Perceptron

An MLP neural network is a powerful classifier that uses hyperplanes to separate the data into different classes. The structure used in this experiment consists of three parts: input layer, hidden layer, and output layer.

The inputs to the system are the lagged metrics $M_{-8}, \ldots, M_{-2}, M_{-1}, M_0$ described in the Section III-A. The network weights and biases are then adjusted to minimize a cost function by a training algorithm [9]. In this case, we used the optimized Levenberg–Marquardt with adaptive momentum (OLMAM) training algorithm. This is the traditional LM algorithm with an additional adaptive momentum term that offers excellent convergence [2]. We have used a toolbox provided by Ampazis *et al.* [2] in conjunction with Matlab's neural network toolbox (The Mathworks Inc., Natick, MA).

Each MLP network was trained and validated 100 times starting with randomly selected initial weights and biases. All results are given by the mean percent correct classification or by kappa statistic from the 100 experiments. The number of hidden neurons in the MLP was determined by a performance curve where the percent correct classification was plotted versus the number of hidden neurons. The point where the validation set starts to decrease in performance while the training set continues to increase is the point that gives the maximum training and validation performance, and determines the number of hidden neurons.

C. Learning Vector Quantization

An LVQ neural network [21], a subclass of the so-called Kohonen networks, was also tested. The inputs to the system are the lagged metrics $M_{-8}, \ldots, M_{-2}, M_{-1}, M_0$ described in Section III-A. The neural predictor was initially trained utilizing the LVQ-1 algorithm and fine-tuned with the LVQ-3 algorithm. The software used was the LVQ_PAK (Helsinki University of Technology).

Classification by both LVQ-1 and LVQ-3 is based on codebook of vectors m_i, where each codebook vector belongs to a certain class. The input vector X is compared to each codebook vector m_i and assigned to the same class to which the closest codebook vector belongs. The distance d between X and m_i is calculated according to the following formula [20]:

$$d = \min_i \parallel X - m_i \parallel .$$

Similarly to MLP, the LVQ model was trained and validated 100 times with random initial codebook vectors, and all results were given by the mean percent correct classification or by kappa statistic from the 100 experiments. The number of codebook vectors in the LVQ classifiers was determined by a performance curve where the percent correct classification or kappa statistic was plotted versus the number of codebook vectors.

D. Support Vector Machines

SVMs use kernel functions to map the input space to a higher dimensional feature space. Optimization techniques are then applied to find the separating hyperplane that maximizes the margin between two classes in the feature space. This creates an arbitrarily complex decision boundary ideal for nonlinearly nonseparable data [26], [6]. The support vector classifier is of the form

$$y = \text{sign}(f(x))$$
$$f(x) = \sum_{i=1}^{N_{SV}} \alpha_i y_i k(x, x_i) + b$$

where x_i are the input data points, y_i are the class targets $(-1, 1)$, b is the bias, $k(x, x_i)$ is a kernel function, and α_i are the Lagrange multipliers from solving a quadratic optimization problem. The kernel used in this work is the Gaussian radial basis function [26], [6]

$$k(x_i, x_j) = \exp\left(-\gamma \|x_i - xj\|^2\right).$$

An SVM tool, R package e1071, was used to create the SVM model [27]. A radial basis function (RBF) kernel is used, and the optimal model parameters, cost and γ (width of Gaussian RBF), are determined using tenfold cross validation of the training set.

E. Rejection

Reliability of classification models is achieved by training the model to reject samples that are evaluated as "unreliable." Probabilistic work has been done for finding theoretical optimal rejection rates versus error [16], [4], but we are concerned with the implementation and application of rejection for reliability. De Stefano *et al.* have developed rejection techniques for both the LVQ and MLP neural networks [8].

The variable Ψ_b is a measure of how close samples are to the decision boundary (in the MLP case), or equidistant between two codebook vectors (in the LVQ case). For LVQ, Ψ_b is given by

$$\Psi_b = 1 - \frac{O_{\text{WIN}}}{O_{2\text{WIN}}}$$

where O_{WIN} is the distance between epoch x and the closest codebook vector, and $O_{2\text{WIN}}$ is the distance between the epoch x and the next closest codebook vector that belongs to a different class.

In the MLP case, the difference is calculated for the winning neuron and the second winning neuron

$$\Psi_b = O_{\text{WIN}} - O_{2WIN}.$$

The values of Ψ_b are between zero and the maximum network output value (depending on the output neuron type) with zero being considered unreliable. Samples are rejected if they are below the desired threshold. The reliability tool developed by De Stefano *et al.* provides an intuitive tool that allows the model designer to reject samples systematically.

Fig. 1 is an example plot of the current epoch value and the resulting reliability measure output from the MLP neural network. In the reliable range, that is, above 0.4 for the reliability measure, it is seen that based on the current epoch mean, sleep and wake are separable, while below 0.4, they are not.

For the SVM classifier, outputs are modified to construct a posterior probability from the distance of an epoch to the separating hyperplane [26]. The outputs of an SVM classifier are of the form

$$y = \text{sign}(f(x)).$$

The output $f(x)$ is then modified to a posterior probability by

$$p(x) = \frac{1}{1 + \exp\left(k * \left(\frac{1 - |f(x)|}{\|w\|}\right)\right)} = \frac{1}{1 + \exp(k * (d_{sv} - d_x))}$$

where $|f(x)|$ is the absolute value of $f(x), k$ is a scaling factor decided from a validation set, and $\|w\|$ is the norm of the weight vector. In terms of distance measures, $d_{sv} - d_x$ is the distance between an input epoch and a support vector. A threshold is applied to $p(x)$ to reject samples that lie closer to the decision

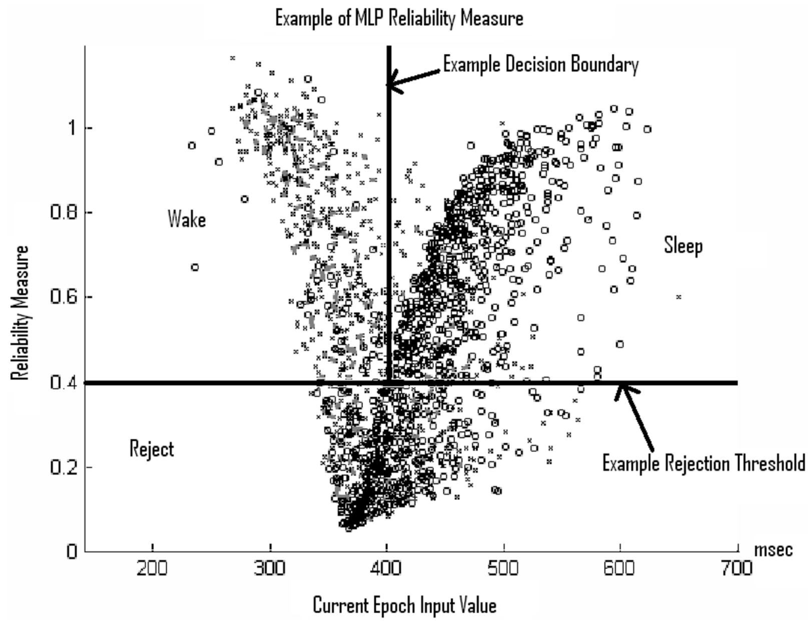

Fig. 1. Simplified example of increasing reliability through rejection. The o's on the graph indicate sleep while x's indicate wake states. The current epoch input value is plotted against the resulting MLP reliability measure. The epochs below the example rejection threshold are discarded. The epochs to the left of the decision boundary will be classified as wake while those to the right of the decision boundary will be classified as sleep. Note the separation between the two sleep states as the reliability measure increases.

boundary than the support vectors. Different values of k are tested to find a value that spread the probability across the range (0,1); the value of $p = 0.5$ is the special case when the sample is a support vector.

IV. Results

A. Multilayer Perceptron

The MLP network was trained by the data from 64 infants combined into a single data set with an equal number of epochs from wake and sleep selected at random. Nine input features were used (the mean of the current epoch and the mean of the previous eight epochs), and the network was trained for 1000 iterations. Sixty three infants were used for validation and 63 infants were used for testing.

The performance curve for graphically showing overtraining can be seen in Fig. 2. Overtraining occurs when there is a decrease in the validation set accuracy while the training set continually increases in accuracy. For the MLP, the structure determined from the training and validation sets was nine inputs, 32 hidden neurons, and two outputs.

A threshold was applied to the output reliability measure. A plot of threshold values against percent correct classification for the validation data set can be seen in Fig. 4. As can be seen in the plot, the point at which maximum percent correct classification occurs only leaves about 25% of the original data. A rejection threshold of 0.30 was set to remove 29.9% of the training data and applied to the validation and test data. This was the maximum amount of data we were willing to reject based on the 30% estimate provided by consultants for our application. Since the distribution of data is very similar between sets, the MLP rejected 27.2% in the validation set and rejected 28.0% in the testing set. The performance of the MLP comparing no rejection to approximately 30% rejection is summarized in Table II.

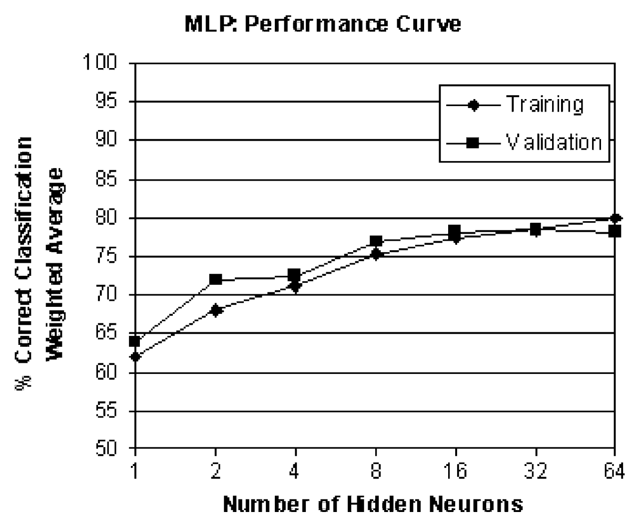

Fig. 2. Performance curve for the MLP neural network. The number of hidden neurons is plotted against the percent correct classification of the weighted average of sleep and wake epochs.

The manual PSG scored test set has 38 121 (67.8%) sleep and 18 076 (32.2%) wake epochs for a total of 56 197 epochs. The MLP classification of the entire test set resulted in 77.5% agreement with the PSG sleep epochs and 79.0% with wake. The rejection scheme applied to the MLP resulted in 28.0% of sleep and wake epochs meeting the rejection criterion. Of the remaining 40 474 epochs 85.7% are in agreement with the PSG sleep epochs and 85.3% with wake.

TABLE II
MLP: TESTING PERFORMANCE—ZERO REJECTION COMPARED TO APPROXIMATELY 30% REJECTION

		Training		Validation		Testing	
		Zero Rejection	30% Rejection	Zero Rejection	30% Rejection	Zero Rejection	30% Rejection
Sleep	Epochs	39827	27607(69.3%)	37840	27350(72.3%)	38121	26790(70.3%)
	% Correct	77.2	85.1	78.5	86.1	77.5	85.7
Wake	Epochs	17742	12728(71.7%)	18957	13981(73.8%)	18076	13684(75.7%)
	% Correct	76.1	83.3	78.1	84.9	79.0	85.3
Total	Epochs	57569	40335(70.1%)	56797	41331(72.8%)	56197	40474(72.0%)
	% Correct	76.9	84.6	78.4	85.7	77.9	85.6

TABLE III
LVQ: TESTING PERFORMANCE—ZERO REJECTION COMPARED TO APPROXIMATELY 30% REJECTION

		Training		Validation		Testing	
		Zero Rejection	30% Rejection	Zero Rejection	30% Rejection	Zero Rejection	30% Rejection
Sleep	Epochs	39827	28929(72.6%)	37840	27548(72.8%)	38121	28157(73.9%)
	% Correct	80.3	88.2	81.1	89.3	79.8	87.2
Wake	Epochs	17742	11162(62.9%)	18958	11601(61.2%)	18076	11572(64.0%)
	% Correct	71.9	76.8	73.2	79.6	74.5	80.9
Total	Epochs	57569	40091(69.6%)	56798	39149(68.9%)	56197	39727(70.7%)
	% Correct	77.7	85.1	78.4	86.5	78.1	85.4

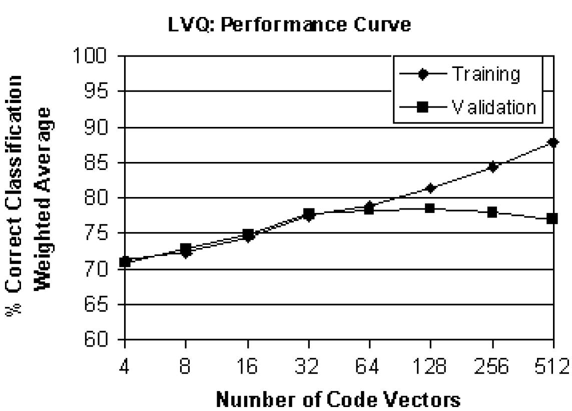

Fig. 3. Performance curve for the LVQ neural network. The number of codebook vectors are plotted against the percent correct classification of the weighted average of sleep and wake epochs.

B. Learning Vector Quantization

The LVQ network of 128 codebook vectors (64 sleep and 64 wake) was trained by the data from 64 infants combined into a single data set. Nine input features were used (the mean of the current epoch and the mean of the previous eight epochs). The network was trained for 70 960 iterations by the LVQ_1 algorithm and fine-tuned by LVQ_3 algorithm for 70 960 iterations. The value of 128 codebook vectors was decided from the performance curve for LVQ in Fig. 3 that shows when overtraining occurs.

A threshold was applied to the output reliability measure. In Fig. 4, it can be seen that the results are quite similar to the MLP in the sense that as more data was rejected, the better the overall accuracy of classification became. A rejection threshold of 0.29 was set to remove 30.4% of the training data. Since the distribution of data is very similar between sets, the LVQ rejected 31.1% in the validation set and rejected 29.3% in the testing set. The performance of the LVQ comparing no rejection to 30% rejection is summarized in Table III.

The manual PSG scored test set has 38 121 (67.8%) sleep and 18 076 (32.2%) wake epochs for a total of 56 197 epochs. The LVQ classification of the entire test set resulted in 79.8% agreement with the PSG sleep epochs and 74.5% with wake. The rejection scheme applied to the LVQ resulted in 29.3% of sleep and wake epochs meeting the rejection criterion. Of the remaining 39 729 epochs 87.2% are in agreement with the PSG sleep epochs and 80.9% with wake.

C. Support Vector Machine

The SVM parameters determined from the tenfold cross validation are a cost of two and a γ of four. When modifying outputs for rejection, a value for k of 1300 was chosen to spread the probability function across the range of zero to one. The results are quite similar to the other classifiers as seen in Fig. 4. The rejection threshold of 0.22 was set to remove 29.8% of the training data and applied to the validation and test data. Since the distribution of data is very similar between sets, 36.1% of the data was rejected in the validation set and 34.5% in the testing set. The performance of the SVM comparing no rejection to approximately 30% rejection is summarized in Table IV.

The manual PSG scored test set has 38 121 (67.8%) sleep and 18 076 (32.2%) wake epochs for a total of 56 197 epochs. The SVM classification of the entire test set resulted in 78.0% agreement with the PSG sleep epochs and 78.5% with wake. The rejection scheme applied to the SVM resulted in 34.5% of sleep and wake epochs meeting the rejection criterion. Of the remaining 36 830 epochs 89.1% are in agreement with the PSG sleep epochs and 80.2% with wake.

D. Comparison of Classifiers

In this section, we compare plots and tables that summarize the results for all three classifiers: MLP, LVQ, and SVM. Most results will be given in percent correct classification because this is the most intuitive. Fig. 4 compares the percent of data rejected to the percent correct classification of the three classifiers.

TABLE IV
SVM: TESTING PERFORMANCE—ZERO REJECTION COMPARED TO APPROXIMATELY 30% REJECTION

		Training		Validation		Testing	
		Zero Rejection	30% Rejection	Zero Rejection	30% Rejection	Zero Rejection	30% Rejection
Sleep	Epochs	39827	28152(70.7%)	37840	25540(67.5%)	38121	26062(68.4%)
	% Correct	80.2	90.9	78.4	90.2	78.0	89.1
Wake	Epochs	17742	12267(69.1%)	18958	10766(56.8%)	18076	10768(59.6%)
	% Correct	78.4	82.2	77.1	78.8	78.5	80.2
Total	Epochs	57569	40419(70.2%)	56798	36306(63.9%)	56197	36830(65.5%)
	% Correct	79.7	88.3	78.0	86.8	78.2	86.5

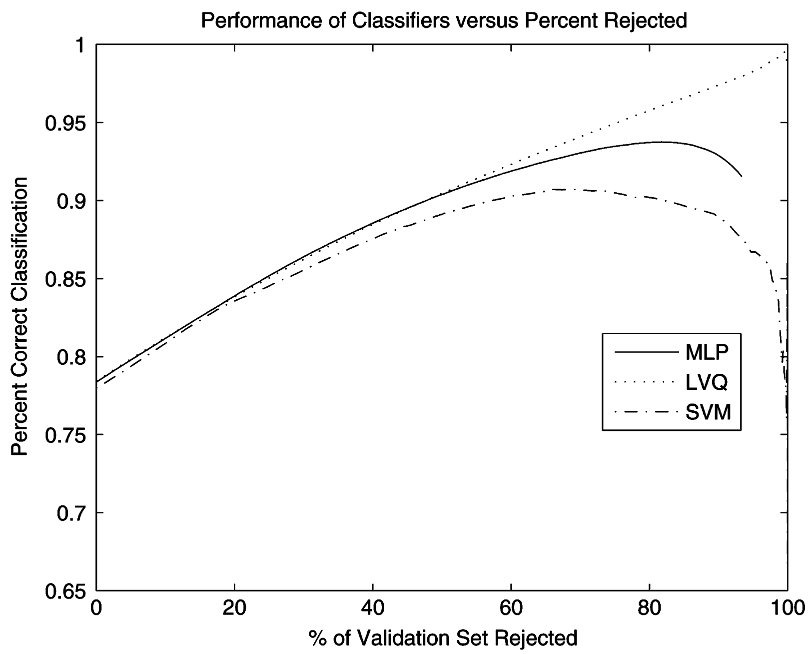

Fig. 4. Rejection performance curve for the LVQ neural network in terms of percent correct classification. The percent of the data rejected is plotted against the percent correct classification of the weighted average of sleep and wake epochs.

The difference in classification between the training, validation, and testing sets is very small suggesting that the data is similarly distributed across infants in the three data sets. Also, the classifiers all offer similar results.

To better understand the results from the classifiers, results are also given by κ statistics. The κ statistics are used for interrater reliability to assess agreement between the manual PSG classification and the sleep classification algorithms. The κ statistic is used to find the consensus between raters after being corrected for chance. This provides the baseline to see which automated method can achieve the best performance after correcting for chance. Equation (1) shows how to calculate the κ statistic [11], [36], [35]

$$\kappa = \frac{P_A - P_C}{1 - P_C} \tag{1}$$

where P_A is the percent agreement between the raters and P_C is the percent agreement expected by chance. Table V shows the strength of agreement for various κ values [35]. A κ value of zero simply shows that the raters have not done any better than what is expected by chance.

A graph of κ values versus percentage of validation data set rejected can be seen in Fig. 5. The plots are very similar in shape to the plots in Fig. 4. The κ results move from moderate to substantial agreement when rejection is added.

TABLE V
κ STATISTIC VALUES AND THE CORRESPONDING STRENGTH OF AGREEMENT

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

To compare all three classifiers with and without rejection, the results are summarized in Table VI for no rejection, rejection of 10%, rejection of 20%, and rejection of 30% of the testing data set.

E. Indeterminate State

The final experiment performed was to study what the classifiers would do when presented with an epoch that had a label of "indeterminate." Indeterminate is when the technicians scoring the data could not place an epoch in either sleep or wake class. From the 190 infants, there was a total of 10 115 indeterminate epochs averaging 53 per infant. Assuming each infant has 960

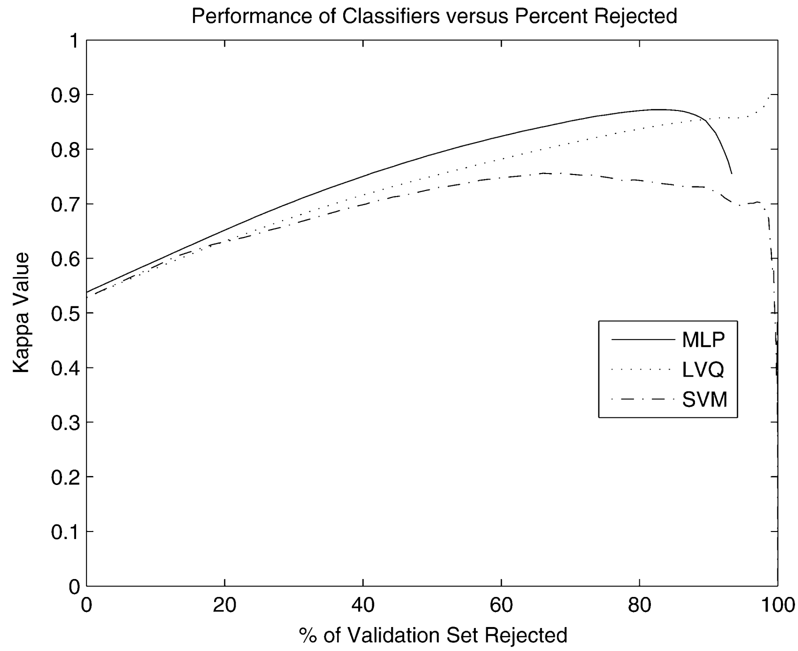

Fig. 5. Rejection performance curve for the LVQ neural network in terms of κ. The percent of the data rejected is plotted against the κ statistic of comparing the gold standard to the algorithms' output.

TABLE VI
COMPARISON OF CLASSIFIERS BY κ AND PERCENT CORRECT CLASSIFICATION FROM THE TESTING DATA SET

	Zero Rejection		10% Rejection		20% Rejection		30% Rejection	
	Kappa	% Correct	Kappa	% Correct	Kappa	% Correct	Kappa	% Correct
MLP	0.528	78.0	0.576	80.3	0.631	82.9	0.688	85.6
LVQ	0.521	78.1	0.569	80.7	0.618	83.2	0.659	85.4
SVM	0.531	78.2	0.593	81.2	0.643	84.3	0.680	86.5

TABLE VII
ANALYSIS OF INDETERMINATE SLEEP REJECTION BY DIFFERENT CLASSIFIERS

Classifier	% Indeterminate State Rejected	Average Errors/Infant	% Errors/Total Epochs
MLP	36.4	33.7	3.6
LVQ	38.8	32.4	3.4
SVM	45.0	29.1	3.1

epochs (30-s epochs * 8 h) that would impact classification accuracy by adding 5.5% error since the zero rejection classifier will call each indeterminate either sleep or wake.

A data set was created by taking every indeterminate state epoch and the eight epochs before the indeterminate state from the training, validation, and testing data sets. When the data set containing only indeterminate states was presented to each of the classifiers, there was a marked increase in rejection. The MLP rejected 36.4% of the data set with indeterminate states (versus 28.0% of the sleep/wake data of the testing data set), LVQ rejected 38.8% (versus 29.3%), and SVM rejected 45.0% (versus 34.5%) of the indeterminate state data set. For comparison, SVM rejected an average (training, validation, and testing) of 33.5% of the sleep/wake data and rejected 45.0% of the data set with indeterminate states. Therefore, there was an 11.5% increase in rejection for indeterminate data set. If 45.0% of the indeterminate epochs are rejected, from a total of 180 678 (train + validate + test + indeterminate) epochs, this decreases the error from 5.5% without rejection to 3.1% with rejection when including indetermine sleep state. All of the classifiers have shown that indeterminate epochs were rejected more than sleep/wake epochs. Table VII is a summary of indeterminate rejection.

F. *Combining Mean and SDNN as Input Features*

To assess the impact of additional features to the classifier, the classifier of choice (SVM) was retrained with the mean and SDNN as inputs. From the feature selection routine (Section III-A), it was discovered that mean was the best input feature and SDNN was the second best input feature. Similar to the previous classifiers, the current epoch plus the previous nine epochs were used as inputs. This created an 18-element input vector for each 30-s epoch. The SVM classifier was then retrained as before (Section III-D). The results from the test set with SDNN included were much worse than using only the mean. Sleep was correctly classified 26.9% of the time; wake was correctly classified 97.5% of the time. This yields an overall correct classification of 50.0%.

V. DISCUSSION

The work presented here has developed a simple yet powerful technique to classify sleep and wake states from infant ECG. This method incorporates a unique approach which considers reliability in the development of the model. The automated techniques show significant agreement with manual PSG based sleep scoring using a reduced set of physiologic signals. After rejection of difficult to classify segments, this model can achieve 85%–87% correct classification while rejecting only approximately 30% of the data. This is an improvement of about 8% over a traditional model without rejection.

The classification results were very similar between all three classifiers, but when comparing the rejection of the indeterminate state data set, the SVM model rejected more epochs. Although MLP offers a slightly higher kappa statistic, SVM rejects more of the indeterminate state and has a higher overall classification making it the model of choice for this application to classify infant sleep/wake states from ECG recordings. The SVM success is due to the classifier's ability to create an optimal, highly complex decision boundary for the training data. The data in this problem is nonlinearly nonseparable. This means that even the most complex solutions will not be able to solve the problem completely, but we can increase the accuracy by rejecting epochs when the classifier cannot reliably make a decision.

Looking at Tables II–IV, the wake epochs are rejected more than the sleep epochs. Wake epochs may appear very similar at times to sleep epochs because the infant is not moving and may be in a physiological state very similar to sleep. These epochs are probably on the decision boundary and thus rejected. This leaves a large amount of sleep data that is far from the decision boundary and classified correctly.

When another HRV feature (SDNN) was added to the classification model, poor results were obtained. The size of the input vector doubled from nine to 18 elements. The additional inputs may have created a classifier with less predictive dimensions in the feature space and led to a model that was not able to generalize effectively. Nine epochs (the current plus nine previous) were utilized per each feature to create a lagged model that was not as susceptible to sudden changes. Future work should investigate the tradeoff between epoch length, number of lagged epochs, and the number of features used for the classification model.

The results from the model utilizing only the mean plus rejection achieve accuracy levels comparable to the previous studies. Nason *et al.* achieved 75%–90% on a single infant [28]. Harper *et al.* achieved 84.8% using ECG and respiration measures on a population of 25 normal infants [14]. Haddad *et al.* and Lisenby *et al.* were able to separate QS from REM, but tested on segments of data only from sleep [12], [25]. Welch *et al.* created individual models for subjects [37]. One short coming is that all of the previous studies use very small populations to develop their algorithms. Other differences are that previous studies require respiratory measures or require the data to already be separated into sleep and wake states. Our algorithm uses HRV metrics derived from ECG data to classify sleep and wake states without the aid of other physiological measures or prescreening. This experiment utilized a very large infant data set that incorporated a large range of ages and birth weights while keeping also incorporating similar proportions of male/female and term/preterm (see Table I).

This exploratory study was performed considering many measures of HRV. It was found that mean provided the most power in determining sleep/wake. We utilized nine lagged means as inputs to the classifier, based on the idea that sleep/wake state does not change abruptly. One reason that mean provides the most power may be that many measures of HRV are not reliable for a 30-s epoch. More exploration is needed to determine the potential of adding other HRV measures to the model.

Another simplification was to use a vector of 30-s epochs as input to the classifier. The choice of 30 s is derived from standards of sleep scoring and the nature of the CHIME data set. An alternate option to this method that should be investigated further is using longer epochs in order to get a better measurement of HRV (e.g., 5-min epochs). This could make classification more difficult since a single metric is calculated from a potentially mixed epoch (multiple sleep states in a single epoch), but a rejection algorithm could prove to remove epochs that do not contain a dominate sleep state.

This algorithm used only HRV metrics derived from the ECG. Further refinements could include other physiological signals commonly recorded on ambulatory monitors such as respiration. If respiration data was used, then a longer epoch could also prove to be useful since breath intervals are on the order of seconds.

Another important note on this work is our gold standard. Typically, only a single sleep scoring technician reviews each file. Interrater reliability has been performed for a small portion of the CHIME data [7]. In this study, a κ statistic of 0.68 was achieved which indicates "substantial agreement" between scorers. However, even trained scores may not always agree. This introduces some uncertainty to the model since we are training our classifier on potentially conflicting classifications. That is to say sleep scorers could be presented the same data and score it differently. Therefore, two epochs that look exactly the same from the perspective of the classifier would have different classes. It would be interesting to study where multiple sleep scoring technicians are presented the same data. The data where the technicians agree the most could be used as a gold standard for training, and the data where the technicians do not agree could be used to aid in the rejection routine.

The model of choice from this experiment is the SVM. To recreate the exact model from the experiment R package e1071 should be used with the RBF kernel, a cost of two, a γ of four, and a value for k of 1300 (used for rejection routine).

The advantages of the automated algorithm over the manual PSG-based techniques are cost, time, resources, and noninvasiveness. The techniques presented offer a first step-to-sleep state classification using a limited physiologic data set. This will be very helpful in other studies, such as maturation, when it is too costly to record a full PSG over a long period of time.

This study demonstrates an automated approach to using rejection in training of a model. Rejection is useful for biomedical studies where accuracy of the decision is important and there is a large data set available (and/or time available). Many biomedical applications could take advantage of this approach, where the rejection option in a classifier could be used to ensure reliable classification.

VI. Conclusion

In conclusion, rejection utilized in a classification model has been demonstrated for a biomedical application. In this study, a method has been developed for sleep/wake classification based on the electrocardiogram that incorporates reliability. After rejection of difficult to classify segments, this model can achieve 85%-87% correct classification while rejecting only 30% of the data. This is an improvement of about 8% over a model without rejection. This level of classification may be useful for physiologic data where EEG is not available. This study is a step towards sleep state classification in infants with a limited set of physiologic parameters. Models which incorporate rejection could be utilized for many biomedical applications where a large amount of data is available in order to improve classification performance.

References

[1] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1412–1423, Dec. 2001.

[2] N. Ampazis and S. Perantonis, "Two highly efficient second-order algorithms for training feedforward networks," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1064–1074, Sep. 2002.

[3] D. Barnett, A. Laposky, C. Thomas, and M. Anch, "Neural network scoring of rat sleep stages," in *Proc. IEEE 1st BMES/EMBS Conf.*, Oct. 1999, p. 389.

[4] C. K. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 1, pp. 41–46, Jan. 1970.

[5] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist actigraphy," *Sleep*, vol. 15, pp. 461–469, 1992.

[6] N. Cristianini and J. S. Taylor, *An Introduction to Support Vector Machines And Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2000.

[7] D. H. Crowell, T. T. Hoppenbrouwers, C. E. Hunt, L. E. Kapuniai, G. Lister, M. R. Neuman, M. Peucker, S. L. Ward, D. E. Weese-Mayer, and M. Willinger, "Infant polysomnograph: Reliability. Collaborative home infant monitoring evaluation (chime) steering committee," *Sleep*, vol. 20, no. 7, pp. 553–560, 1997.

[8] C. De Stefano, C. Sansone, and M. Vento, "To reject or not to reject: That is the question—An answer in case of neural classifiers," *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, vol. 30, no. 1, pp. 84–94, Feb. 2000.

[9] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[10] P. Estevez, C. Held, C. Holzmann, C. Perez, J. Perez, J. Heiss, M. Garrido, and P. Peirano, "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants," *Med. Bio. Eng. Comput.*, vol. 40, pp. 105–113, 2002.

[11] J. L. Fleiss, *Statistical Methods for Rates and Proportions*. New York: Wiley, 1973.

[12] G. G. Haddad, H. J. Jeng, T. L. Lai, and R. B. Mellins, "Determination of sleep state in infants using respiratory variability," *Pediatric Res.*, vol. 21, no. 6, pp. 556–562, 1987.

[13] R. Harper, V. Schechtman, and K. Kluge, "Machine classification of infant sleep state using cardiorespiratory measures," *Electroencephalogr. Clin. Neurophysiol.*, vol. 67, pp. 379–387, 1987.

[14] R. Harper, V. Schechtman, and K. Kluge, "Machine classification of infant sleep state using cardiorespiratory measures," *Electroencephalogr. Clin. Neurophysiol.*, vol. 67, pp. 379–387, 1987.

[15] J. Heiss, C. Held, P. Estevez, C. Perez, C. Holzmann, and J. Perez, "Classification of sleep stages in infants: A neuro fuzzy approach," *IEEE Eng. Med. Biol. Mag.*, vol. 21, no. 5, pp. 147–151, Sep./Oct. 2002.

[16] M. E. Hellman, "The nearest neighbor classification rule with a reject option," *IEEE Trans. Syst. Sci. Cybern.*, vol. 6, no. 3, pp. 179–185, Jul. 1970.

[17] C. Holzmann, C. Perez, C. Held, M. S. Martin, F. Pizarro, J. Perez, M. Garrido, and P. Peirano, "Expert-system classification of sleep/waking states in infants," *Med. Biol. Eng. Comput.*, vol. 37, pp. 466–476, 1999.

[18] T. Hoppenbrouwers, M. Neuman, M. Corwin, J. Silvestri, T. Baird, D. Crowell, C. Hunt, M. Sackner, G. Lister, M. Willinger, and The CHIME Study Group, "Multivariable cardiorespiratory monitoring at home: Collaborative home infant monitoring evaluation (chime)," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol.*, 1996, pp. 61–62.

[19] B. Kim and K. Park, "Automatic sleep stage scoring system using genetic algorithms and neural network," in *Proc. 22nd IEEE Eng. Med. Biol. Soc. (EMBS) Conf.*, Jul. 2000, pp. 849–850.

[20] T. Kohonen, "Self-organizing maps," in *Springer Series in Information Sciences*, 3rd ed. New York: Springer-Verlag, 2001, vol. 30.

[21] T. Kohonen, J. Hynninen, J. Kangas, J. Laaksonen, and K. Torkkola, LVQ PAK: The Learning Vector Quantization Program Package, Helsinki University of Technology, Helsinki, Finland, 1995.

[22] I. Koprinska, G. Pfurtscheller, and D. Flotzinger, "Sleep classification in infants by decision tree-based neural networks," *Artif. Intell. Med.*, vol. 8, pp. 387–401, 1996.

[23] A. T. Lewicke, E. S. Sazonov, M. J. Corwin, and S. A. C. Schuckers, "Reliable determination of sleep versus wake from heart rate variability using neural networks," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2005, pp. 2394–2399.

[24] A. T. Lewicke, E. S. Sazonov, and S. A. C. Schuckers, "Sleep-wake identification in infants: Heart rate variability compared to actigraphy," in *Proc. 26th IEEE Eng. Med. Biol. Soc. (EMBS) Conf.*, Sep. 2004, pp. 442–445.

[25] M. Lisenby, P. Richardson, and A. Welch, "Detection of cyclic sleep phenomena using instantaneous heart rate," *Electroencephalogr. Clin. Neurophysiol.*, vol. 40, pp. 169–177, 1976.

[26] A. Madevska-Bogdanova and D. Nikolic, "A new approach of modifying SVM outputs," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2000, pp. 395–398.

[27] D. Meyer, "Support vector machines: The interface to LINSVM in package e1071," *R-News*, vol. 1, no. 3, pp. 23–26, 2001.

[28] G. P. Nason, T. Sapatinas, and A. Sawczenko, "Wavelet packet modelling of infant sleep state using heart rate data," *Sankhya: The Indian J. Statist.*, vol. 63, pp. 199–217, 2001.

[29] M. R. Neuman, H. Watson, R. S. Mendenhall, J. T. Zoldak, J. M. Di Fiore, M. Peucker, T. M. Baird, D. H. Crowell, T. T. Hoppenbrouwers, D. Hufford, C. E. Hunt, M. J. Corwin, L. R. Tinsley, D. E. Weese-Mayer, M. A. Sackner, and The CHIME Study Group, "Cardiopulmonary monitoring at home: The chime monitor," *Physiol. Meas.*, vol. 22, no. 2, pp. 267–286, 2001.

[30] L. Quanzheng and G. Xiaorong, "Subsection approximate entropy and its application in sleep staging," in *Proc. IEEE 1st BMES/EMBS Conf.*, Oct. 1999, p. 900.

[31] E. Sazonov, N. Sazonova, S. A. C. Schuckers, and M. Neuman, "Activity-based sleep-wake identification in infants," *Physiol. Meas.*, vol. 25, no. 5, pp. 1291–1304, Oct. 2004.

[32] M. Scher, S. Dokianakis, D. Steppe, D. Banks, and R. Sclabassi, "Computer classification of state in healthy preterm neonates," *Sleep*, vol. 20, no. 2, pp. 132–141, 1997.

[33] M. Scher, S. Dokianakis, M. Sun, D. Steppe, R. Guthrie, and R. Sclabassi, "Computer classification of sleep in preterm and full-term neonates at similar postconceptional term ages," *Sleep*, vol. 19, no. 1, pp. 18–25, 1996.

[34] M. Scher, M. Sun, G. Hatzilabrou, N. Greenberg, G. Cebulka, D. Krieger, R. Guthrie, and R. Sclabassi, "Computer analyses of EEG-sleep in the neonate: Methodological considerations," *J. Clin. Neruophysiol.*, vol. 7, no. 3, pp. 417–441, 1990.

[35] S. Stemler, "An overview of content analysis," *Practical Assess. Res. Eval.*, vol. 7, 2001, ISSN 1531-7714.

[36] S. Stemler, "A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability," *Practical Assess. Res. Eval.*, vol. 9, 2004, ISSN 1531-7714.

[37] A. Welch and P. Richardson, "Computer sleep stage classification using heart rate data," *Electroencephalogr. Clin. Neurophysiol.*, vol. 34, pp. 145–152, 1973.

[38] X. Xu and S. Schuckers, "Automatic detection of artifacts in heart period data," *J. Electrocardiol.*, vol. 33, 2001.

[39] J. Yen and R. Langari, *Fuzzy Logic: Intelligence, Control, and Information*, 1st ed. Englewood Cliffs, NJ: Prentice-Hall, 1998.

Aaron Lewicke (S'00–M'07) received the B.S.E., M.E., and Ph.D. degrees in electrical engineering from Clarkson University, Potsdam, NY, in 2001, 2003, and 2005 respectively. His thesis research focused on heart rate variability of infants, including sleep state classification and analysis of cardiorespiratory events.

Currently, he is a Senior Research Scientist at Boston Scientific Cardiac Rhythm Management, St. Paul, MN.

Edward Sazonov (S'02–M'02) received the diploma of Systems Engineer from Khabarovsk State University of Technology, Khabarovsk, Russia, in 1993 and the M.S. degree in electrical engineering and the Ph.D. degree in computer engineering from West Virginia University, Morgantown, in 2002 and 1999, respectively.

Currently, he is an Assistant Professor at the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY. His research interests are focused on ambient intelligent systems with applications in bio- and rehabilitation engineering such as self-powered wearable sensors for collection of physiological data, noninvasive measurement of energy intake and expenditure in free living population, sleep state scoring from noninvasive measurements, and biofeedback devices for people with impaired proprioceptive abilities.

Michael J. Corwin received the B.A. degree in biology from the Northwestern University, Evanston, IL, in 1973, and the M.D. degree from the Wayne State University, Detroit, MI, in 1977.

Currently, he is the Executive Vice President at BattelleCRO, Inc., Newton, MA, where he is responsible for medical and regulatory affairs and the Associate Professor of Pediatrics and Epidemiology, Boston University School of Medicine, Boston, MA.

Dr. Corwin is a member of the American Academy of Pediatrics and the Society for Pediatric Research.

Michael R. Neuman received the B.S., M.S., and Ph.D. degrees in electrical engineering from Case Institute of Technology, Cleveland, OH, in 1961, 1963, and 1966, respectively, and the M.D. degree from Case Western Reserve University School of Medicine, Cleveland, OH, in 1974.

He joined the Department of Biomedical Engineering, Michigan Technological University, Houghton, in August, 2003, as the Professor and Chairman. He was with the Memphis Joint Program in Biomedical Engineering, a joint activity between the University of Tennessee Health Science Center and The University of Memphis where he held the Herbert Herff Chair of Excellence. Prior to that, he served for 32 years on the faculty at Case Western Reserve University where he held positions in the Departments of Biomedical and Electrical Engineering in the School of Engineering and the Departments of Reproductive Biology and Obstetrics and Gynecology in the School of Medicine. In 1989, he joined the National Science Foundation (NSF) Engineering Research Center for Emerging Cardiovascular Technologies, a consortium of several universities based at Duke University and became an Adjunct Professor of Biomedical Engineering at Duke. He and Prof. R. Buck of the Department of Chemistry at the University of North Carolina codirected the biomedical sensors thrust, one of three research thrusts in this ERC. He has published 152 scientific papers and holds five patents. His primary research interests have been in the application of microelectronic techniques and technology to problems in clinical medicine.

Dr. Neuman has served two terms as the President of the International Society on Biotelemetry from 1984 to 1988 and was the Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING from 1989 to 1996. Currently, he serves as the Editor-in-Chief of the international journal *Physiological Measurement*.

Stephanie Schuckers (M'95–SM'07) received the B.S.E. degree in electrical engineering from University of Iowa, Iowa City, in 1992 and the M.S.E. and Ph.D. degrees in electrical engineering: systems from University of Michigan, Ann Arbor, in 1994 and 1997, respectively.

She was an Assistant Professor at West Virginia University for five years and is currently an Associate Professor at Clarkson University, Potsdam, NY. She has published over 40 journal papers, conference papers, and book chapters. Her current research interests are signal processing and pattern recognition for a variety of biorelated applications, including biometric devices, implantable defibrillators, sudden cardiac death, and sudden infant death syndrome.

Dr. Schuckers was Region 2 representative to the administrative committee for the IEEE Engineering Medicine and Biology Society for two years.