Ensemble Learning Approach via Kalman Filtering for a Passive Wearable Respiratory Monitor

Sayandeep Acharya, *Member, IEEE*, William M. Mongan, *Member, IEEE*, Ilhaan Rasheed, *Student Member, IEEE*, Yuqiao Liu, *Student Member, IEEE*, Endla Anday, Genevieve Dion, Adam Fontecchio, *Senior Member, IEEE*, Timothy Kurzweg, *Senior Member, IEEE*, and Kapil R. Dandekar, *Senior Member, IEEE*

***Abstract*—*Objective:* Utilizing passive radio frequency identification (RFID) tags embedded in knitted smart-garment devices, we wirelessly detect the respiratory state of a subject using an ensemble-based learning approach over an augmented Kalman-filtered time series of RF properties. *Methods:* We propose a novel approach for noise modeling using a "reference tag," a second RFID tag worn on the body in a location not subject to perturbations due to respiratory motions that are detected via the primary RFID tag. The reference tag enables modeling of noise artifacts yielding significant improvement in detection accuracy. The noise is modeled using autoregressive moving average (ARMA) processes and filtered using state-augmented Kalman filters. The filtered measurements are passed through multiple classification algorithms (naive Bayes, logistic regression, decision trees) and a new similarity classifier that generates binary decisions based on current measurements and past decisions. *Results:* Our findings demonstrate that state-augmented Kalman filters for noise modeling improves classification accuracy drastically by over 7.7% over the standard filter performance. Furthermore, the fusion framework used to combine local classifier decisions was able to predict the presence or absence of respiratory activity with over 86% accuracy. *Conclusion:* The work presented here strongly indicates the usefulness of processing passive RFID tag measurements for remote respiration activity monitoring. The proposed fusion framework is a robust and versatile scheme that once deployed can achieve high detection accuracy with minimal human intervention. *Significance:* The proposed system can be useful in remote noninvasive breathing state monitoring and sleep apnea detection.**

***Index Terms*—Sensor fusion, wearable sensors, kalman filtering, activity recognition, binary classification.**

Manuscript received December 9, 2017; revised May 17, 2018 and June 25, 2018; accepted July 14, 2018. Date of publication July 20, 2018; date of current version May 6, 2019. This work was supported in part by the National Science Foundation Partnerships for Innovation: Building Innovation Capacity subprogram under Grant 1430212, in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award number U01EB023035, and in part by the Commonwealth of Pennsylvania through the Commonwealth Universal Research Enhancement program. *(Corresponding author: William M. Mongan.)*

S. Acharya, I. Rasheed, Y. Liu, A. Fontecchio, and K. R. Dandekar are with the Department of Electrical and Computer Engineering, Drexel University College of Engineering, Philadelphia, PA 19104 USA (e-mail: sa427@drexel.edu; ir54@drexel.edu; yl636@drexel.edu; krd26@drexel.edu).

W. M. Mongan is with the Department of Computer Science, Drexel University College of Computing and Informatics, Philadelphia, PA 19104 USA (e-mail: wmm24@drexel.edu).

E. Anday is with the Department of Pediatrics, Drexel University College of Medicine, Philadelphia, PA 19129 USA (e-mail: ea32@drexel.edu).

G. Dion is with the Westphal College of Media Arts and Design, Drexel University, Philadelphia, PA 19104 USA (e-mail: gd63@drexel.edu; af63@drexel.edu).

T. Kurzweg is with the School of Engineering, Penn State Behrend, Erie, PA 16563 USA (e-mail: tpk103@psu.edu).

Digital Object Identifier 10.1109/JBHI.2018.2857924

I. Introduction

KNITTED antennas using Radio Frequency Identification (RFID) tags have been utilized as wearable smart-garment devices [1] capable of wirelessly monitoring strain-gauge and movement-based activities such as limb movement for Deep Venous Thrombosis (DVT) monitoring, respiratory monitoring, and uterine monitoring during labor and delivery [2], as well as electrical activity such as ECG heart monitoring [3]. Each of these applications presents challenges beyond their wired, tethered counterparts. Strain-gauge movements, such as respiratory monitoring, are challenging because raw non-strain gauge movements (i.e., moving the body or walking about) are detected in the same band as strain-gauge stretches that would occur from respiratory activity. Sleep apnea is defined as a reduction of respiratory activity by 95% for a period of 10 seconds [4]. Thus, it is possible to exhibit some respiratory activity during this "boundary period" while experiencing the onset of apnea conditions. By considering classifier *inertia*, classifications made at the boundary between two states (i.e., respiration and apnea) consider recent prior classifications when determining a final prediction and the degree confidence in that prediction.

In this study, we use multiple binary classifiers to develop a framework that can detect in real-time the *breathing* or *non-breathing* state of a person. A programmable mannequin simulator was used to collect data from and test our framework. Each classifier was trained on features extracted from RFID measurements of two spatially separated tags. One tag was placed on the mannequin belly to capture breathing movements, and a

secondary reference tag was placed on the shoulder to capture signal artifacts unrelated to breathing. Towards this end, we propose a novel similarity classfier that uses current measurements from the reference tag and the past decisions for classification. A fusion center uses the local classifier confidences when weighting a final prediction of the breathing state at each sampling instant.

Using a reference tag placed elsewhere on the body that does not exhibit strain-gauge motion, noise artifacts that were once in-band with the strain-gauge signal can be observed and modelled. We observe that the noise is not white but exhibits temporal correlation; by augmenting the Kalman Filter with an Autoregressive Moving Average (ARMA) process, the signal from which our features (discussed in Section III-A) are extracted is improved, significantly increasing classification accuracy as compared to a system that used a standard Kalman filter (with white Gaussian meaurement noise). The rest of this paper is organized as follows. We survey related work in Section II. Our data collection, feature extraction, and experimental protocol are summarized in Section III. We detail our fusion system setup and approach in Section IV, and summarize results in Section V. Finally, we conclude with a discussion of future work in Section VI.

II. Related Work

Unobtrusive and wireless techniques to obtain respiration data have several advantages and is hence highly desired. Judging by the high level of interest in this research area. Vital-Radio is one such system that can accurately track a patient's breathing without body contact [5]. This system measures the reflection time of low power transmissions to determine human respiration rate. While the system provides a convenient method to obtain vital signs, it is susceptible to inaccuracies due to non-respiration related motion.

A similar system called WiBreathe is presented in [6] that uses wireless signals in the same frequency range of commercial Wi-Fi devices. While this may be advantageous in reducing system costs and encouraging fast adoption, the system is not capable of uniquely identifying signals from patients. This issue is also present in the Vital-Radio system mentioned earlier and is a result of the two systems relying on signal reflections to carry out respiration detection.

NightCare is a system that uses Ultra High Frequency (UHF) RFID for patient sleep monitoring. The system consists of flexible RFID tags integrated into clothing, bedding and surrounding areas to determine patient motion patterns, falls and interaction with objects [7]. The use of passive UHF RFID allows the system to deploy sensors in patient surroundings without the need for batteries or other power sources. However, NightCare is not capable of obtaining vital signs such as respiration and heart rate, as is possible with our system.

The textile respiration sensor presented in [8] operates in the Industrial, Scientific and Medical (ISM) band (2.45 GHz) and uses a spiral antenna structure. This sensor has been used to detect breathing using two mechanisms: change in the physical properties of the spiral antenna due to stretch and change in dielectric properties of the thoracic cage during respiration. The sensor has not been operated wirelessly and a Vector Network Analyzer (VNA) was used to prove its ability to detect respiration using the two methods mentioned. The authors intend to use a wireless standard such as Bluetooth to enable wireless capabilities to this sensor. However, using such a standard would require the addition of a power source, something that the Bellyband does not require due to its use of UHF RFID.

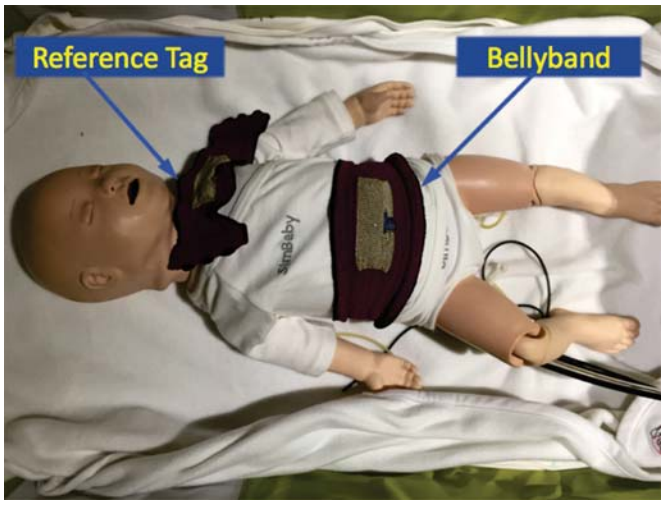

Fig. 1. The Laerdal SimBaby programmable mannequin with Bellyband tag and Reference Tag.

Since our Bellyband uses the commercial EPC Gen2 RFID protocol for wireless operation, subjects can be identified using a unique Tag ID (TID) number stored in the RFID tag embedded in the textile antenna. This enables easy identification and separation of signals from multiple patients. The use of passive UHF RFID also allows Bellyband to operate battery-free, enabling seamless integration of the sensor in clothing for vital sign monitoring.

III. Data Collection

In order to establish a repeatable ground truth (true state of breathing or non-breathing), a Laerdal SimBaby programmable mannequin simulator (see Figure 1) was used to actuate the abdominal area to emulate respiratory activity. The mannequin was programmed to perform respiration at a predefined rate (30 breaths per minute) for one minute, followed by a one minute period of non-breathing; this behavior was repeated throughout the data collection process. Two tags were used: a Bellyband tag that was placed on the abdomen of the SimBaby and a reference tag that was placed on its shoulder (see Figure 2). The positioning of the reference tag on the shoulder makes it less sensistive to movements related to breathing. We use a filtered version of the meaurements from the reference tag as a representative collection of *non-breathing* state data points.

The Impinj Speedway R420 RFID interrogator was used to poll the Bellyband wearable RFID tag and antenna with a 900 MHz band RFID signal, and to measure properties of the backscattered signal reflected from the RFID tag. Although a

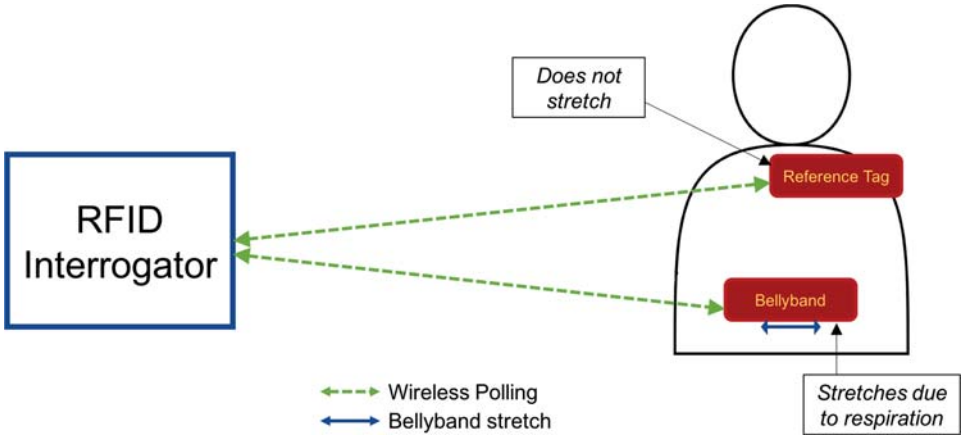

Fig. 2. System diagram illustrating Bellyband and Reference tag interrogation.

fixed polling interval is not guaranteed, interrogations are typically performed with a frequency of approximately 90 Hz. RFID properties considered include the Received Signal Strength Indicator (RSSI), phase, Doppler shift, interrogation frequency, and timestamp. Features extracted from these properties, which include a channel-normalized RSSI (ζ) and a channel-normalized change in phase or Doppler shift (the tag "velocity" v), are discussed in more detail in Section III-A. In Section IV-B, we classify the features in real-time into "breathing" and "non-breathing" states, and in Section IV-D, we fuse the classification results into a final predicted states. Data and features are accessed from a database via a web service, which encrypts and decrypts the data using an evolving key suitable for use in a medical environment under Health Insurance Portability and Accountability Act (HIPAA) regulations.

A. Features

Three measures from the RFID interrogator are considered for subject state tracking: the RSSI of the backscatter from the interrogation of the passive RFID tag, the phase of the reflected signal observed by the interrogator, and the Doppler shift of the reflected signal observed by the interrogator. Each of these metrics is affected in-band by the frequency of the original signal emitted by the interrogator. Under United States Federal Communications Commission (FCC) regulations, RFID interrogations must iterate (or "channel hop") over 50 frequency channels in the 900 MHz band. In addition to perturbing the raw measurement observations at the interrogator, channel hopping poses challenges in computing higher order features from changes in the observed phase, because these features depend on observing changes in successive values of the phase under the assumption that they were observed from the same interrogation frequency [9]. As a result, the observed Doppler shift is used to identify fine movements of the RFID tag, either in space or as a result of a strain force applied to the surrounding knit antenna. The tag "velocity" is computed as a feature on the Doppler measure, which is proportional to the interrogation wavelength $\lambda = \frac{c}{f}$ (where c is the speed of light in a vacuum, and f is the interrogation frequency), the observed Doppler measure, and the interrogation angle, which we assume to be constant. A higher order feature ζ is computed from the RSSI measure by accounting for the interrogation frequency. The Radar Cross Section (RCS) relates changes in received signal power (P_{Rx}) to the interrogation power (P_{Tx}), the reader and tag gains (G_{reader} and G_{tag}, respectively), the return loss (R), and the interrogation wavelength λ [10]. If we assume that the interrogation power P_{Tx} and the reader gain G_{reader} are constant, we observe that the changes in gain of the tag (resulting from movement or a strain force on the antenna), the distance between the interrogator and the tag (resulting from movement), and return loss (resulting from movement, strain force, fading or multipath interference) are proportaional to the interrogation wavelength and observed RSSI measure as defined in (1). We define ζ as those terms of the RCS model that indicate changes at the tag, such as movement, strain force, or interference.

$$\begin{aligned}\zeta &= G_{tag}^{-2} \times r^4 R^{-1} \\ &= P_{Tx,reader} \times G_{reader}^2 \times P_{Rx,reader}^{-1} \times \left(\frac{\lambda}{4\pi}\right)^4, \end{aligned} \quad (1)$$

where $P_{Tx,reader}$ is the RF transmit power from the interrogator (1 Watt, or 30 dBm), $P_{Rx,reader}$ is the observed backscattered power at the interrogator that is reflected from the RFID tag, measured in dBm; G_{tag} and G_{reader} are the tag and reader gains, respectively, r is the distance between the interrogator and tag, and R is the return loss. The proportionality defined by ζ in (1) has a residual of $\delta = -10log_{10}\frac{f^4}{(f-(0.5*10^6 MHz))^4} \approx -0.00941$, as the frequency changes by 500 kHz on each change in channel. This constant linear oscillatory pattern is mitigated by computing $\zeta = \zeta - \delta(50 - \omega)$, where ω is the channel number in [0, 50) that denotes the frequency iteration with respect to the frequency f. The channel number [0, 50) is obtained given the interrogation frequency by computing $\omega = \frac{f-902.25MHz}{500kHz}$ to determine which of the 500 kHz offsets from the start of the 900 MHz RFID band is in use.

In summary, we chose the following features for consideration during wireless respiratory state classification:

- Feature 1: Reflected signal strength as measured at the interrogator (ζ)
- Feature 2: Phase angle of the reflected signal

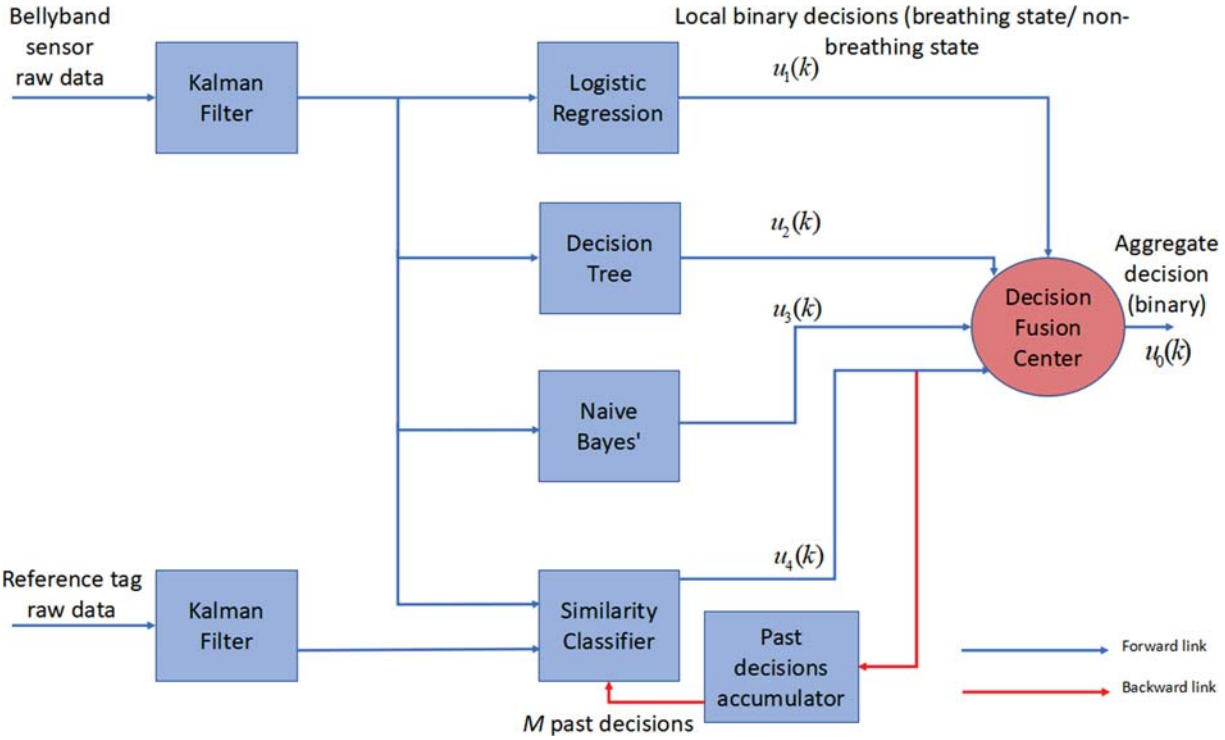

Fig. 3. Decision fusion of Bellyband and reference tag measurements.

- Feature 3: Tag velocity as a function of the observed Doppler shift at the interrogator
- Feature 4: Tag velocity as a function of the change in phase of successive interrogations

The entire dataset consisted of approximately 1200 data points from each tag (Bellyband and reference).

B. Dataset Labeling

As shown in Figure 7, we label each feature tuple with the known state established from the ground truth ("breathing" or "non-breathing"). This is denoted in Figure 7 via vertical lines in time at each minute interval. Each minute of data corresponds to approximately 400 "sampling instants," which consists of a moving average of 6 data points. This indicates an RFID interrogation frequency of 40 Hz per RFID tag; as a primary Bellyband tag and a reference tag were used in the system, an interrogation frequency of roughly 80 Hz was observed. These labels were used to determine accuracy, precision and recall metrics of the classifiers and fusion algorithm, as well as to train the individual classifiers.

IV. System Setup

Figure 3 shows the system setup used for filtering and fusion of raw observations from the Bellyband and reference tags. The system consists of a bank of classifiers; namely logistic regression, decision tree and naive Bayes' for the Bellyband tag observations and one similarity classifier that takes in both the reference tag and Bellyband tag measurements as input. Each classifier generates local binary decisions $u_i(k), i = 1, ..., N$ ($N = 4$ is the number of classifiers) supporting either of the two hypotheses (H_0: breathing state, H_1: non-breathing state). The decision fusion center aggregates the local decisions to compute a global decision $u_0(k)$ supporting the same set of hypotheses. Each Kalman filter has a state augmented form (see Section IV-A) and is tuned individually to the incoming observation source. The 3 binary classifiers excluding the similarity classifier are trained on data sets obtained from the Bellyband sensor through bootstrap aggregation [11]. At each time instant, the similarity classifier computes a binary decision based on the current Bellyband and reference tag measurements and also M past decisions. For this study, the parameter M was fixed to 7 (see Section IV-C) but in general it can be changed during run time based on some chosen performance metric.

A. Kalman Filtering

The standard model of an autonomous (with no control input) discrete dynamic system as a first order linear difference equation is

$$x(k) = F(k)x(k-1) + G(k)w(k), \tag{2}$$

where $x(k) \in \mathbb{R}^n$ is the state of interest at time instant k, and $w(k) \in \mathbb{R}^l$ is a random vector referred to as the process noise. $F \in \mathbb{R}^{n\times n}$ is the system matrix relating past state to the state at time instant k. The matrix $G \in \mathbb{R}^{n\times l}$ relates the process noise to the state at time k. The state observation model is defined as

$$z(k) = H(k)x(k) + v(k), \tag{3}$$

where $z(k) \in \mathbb{R}^d (d \leq n)$ is the observation vector, $H(k) \in \mathbb{R}^{d\times n}$ is the observation matrix at time instant k, and $v(k) \in \mathbb{R}^d$ is the measurement noise. In the model described above, the

TABLE I
ARMA MODEL PARAMETERS BASED ON GRID SEARCH WITH AIC MINIMIZATION FOR BOTH BELLYBAND AND REFERENCE TAG MEASUREMENTS

Feature	AR process order (p)		MA process order (q)	
	Bellyband	Reference	Bellyband	Reference
1	3	1	1	0
2	3	1	1	1
3	1	0	1	0
4	1	0	1	0

noise processes $w(k)$ and $v(k)$ are assumed to be white, zero-mean, uncorrelated, with covariance matrices $Q \in \mathbb{R}^{l\times l}$ and $R \in \mathbb{R}^{d\times d}$ respectively. However initial inspection of the measurements from the tags (Bellyband and reference) showed that the measurement noise is colored (temporally correlated). Colored measurement noise can be handled in multiple ways in the context of Kalman filtering. Two of the popular methods are the *state augmentation approach* ([12], section 7.2) and the *measurement differencing approach* ([13], section 11.2). Here we follow the state augmentation approach. It augments the actual state vector $x(k)$ in (2) with colored noise samples $v(k)$ which are output of an Autoregressive Moving Average (ARMA) process driven by current and delayed samples of zero mean white Gaussian noise. The ARMA(p, q) model of the measurement noise $v(k)$ is

$$v(k) = \sum_{i=1}^{p} a_i v(k-i) + \sum_{j=1}^{q} b_j e(k-j) + e(k), \qquad (4)$$

where a_i for $i = 1, .., p$, b_j for $j = 1, .., q$ are the coefficients of the AR and MA models respectively. p, q are the orders of the AR and MA processes. $e(k) \sim \mathcal{N}(0, \sigma_e^2(k))$ is zero mean Gaussian noise input to the ARMA system. In this study $\sigma_e^2(k)$ was fixed at 0.1. The ARMA model orders p, q were estimated through grid search and the pair with the minimum Akaike Information Criterion (AIC) [14] value was chosen. Table I shows the the ARMA model orders for each of the features (see Section III-A) from both Bellyband and reference tags.

Each measurement obtained from the tag is independently modeled as a scalar state x defined by a first order ($n = 1$) system as

$$x(k) = Fx(k-1) + w(k), \qquad (5)$$

where $x(k)$ is the actual measurement (such as phase in radians or velocity by doppler) at time instant k. The system matrix F is considered to be constant, equal to 1. $w(k)$ is the zero mean Gaussian process noise with variance Q. In other words, we model the state dynamics as a simple random walk. There are various methods (some being computationally demanding) such as [15], [16] of choosing Q. However, such methods were not employed in this study and the process noise variance Q was fixed at 0.05.

Modeling the measurement noise as an ARMA process results in its own dynamic system given by (6).

$$\begin{bmatrix} v(k) \\ v(k-1) \\ \vdots \\ v(k-p+1) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} v(k-1) \\ v(k-2) \\ \vdots \\ v(k-p) \end{bmatrix} + \begin{bmatrix} 1 & b_1 & \cdots & b_q \\ & & & \\ & & 0 & \end{bmatrix} \begin{bmatrix} e(k) \\ e(k-1) \\ \vdots \\ e(k-q) \end{bmatrix} \qquad (6)$$

The state dynamics represented in (6) is included in the state space model of the Kalman filter shown in (2) as a block diagonal element. The augmented state vector becomes $x_{aug}(k) = [x(k), v(k), v(k-1), \ldots, v(k-p+1)]^T \in \mathbb{R}^{(p+1)\times 1}$. The augmented matrices are as follows:

$$F_{aug} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & a_1 & a_2 & \cdots & a_p \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \qquad (7)$$

with $F_{aug} \in \mathbb{R}^{(p+1)\times(p+1)}$ and

$$G_{aug} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & b_1 & \cdots & b_q \\ & & & & \\ & & 0 & & \end{bmatrix} \qquad (8)$$

with $G_{aug} \in \mathbb{R}^{(p+1)\times(q+2)}$. The augmented process noise vector is $w_{aug}(k) = [w(k), e(k), e(k-1), \ldots, e(k-q)]^T \in \mathbb{R}^{(q+2)\times 1}$. The augmented measurement equation becomes

$$z_{aug}(k) = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ v(k) \\ v(k-1) \\ \vdots \\ v(k-p+1) \end{bmatrix} \qquad (9)$$

where the augmented observation matrix $H_{aug} \in \mathbb{R}^{1\times(p+1)}$ is given by $\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix}$. Note that the augmented measurement equation in (9) no longer has any direct measurement noise. The above formulation transforms the single order system in (5) to a $(p+1)$ order system. Using the transformed Kalman matrices, it is shown through (9) that the observation at any instant is the sum of the variable $x(k)$ which from (5) is the RFID signal under investigation and the original measurement noise; namely $z(k) = x(k) + v(k)$. The other elements of the state vector are not observed. Essentially the steps of augmenting the Kalman matrices retained the original observation $z(k)$ (a noisy version

of the RFID signal) but got rid of the colored measurement noise so that a standard Kalman filter can be applied. The predict and update steps are provided:

Predict Step:

$$\hat{x}^{-}(k) = F_{aug}\hat{x}^{+}(k-1), \tag{10}$$

$$\hat{P}^{-}(k) = F_{aug}\hat{P}^{+}(k-1)F_{aug}^{T} + G_{aug}D(w_{aug})G_{aug}^{T}. \tag{11}$$

Update Step:

$$K(k) = \hat{P}^{-}(k)H_{aug}^{T}(H_{aug}\hat{P}^{-}(k)H_{aug}^{T})^{-1}, \tag{12}$$

$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k)(z_{aug}(k) - H_{aug}\hat{x}^{-}(k)), \tag{13}$$

$$\hat{P}^{+}(k) = (I - K(k)H_{aug})\hat{P}^{-}(k). \tag{14}$$

In the above equations $\hat{x}^{-}(k)$ is the apriori estimate of state vector (before observation at instant k has been processed). $\hat{P}^{-}(k) \in \mathbb{R}^{(p+1)\times(p+1)}$ is the covariance of the apriori state estimation error. $\hat{x}^{+}(k)$ is the posteriori state estimate (after observation at instant k has been processed). $\hat{P}^{+}(k)$ is the covariance of the posteriori state estimation error. $K(k)$ is the Kalman gain at instant k. $I \in \mathbb{R}^{(p+1)\times(p+1)}$ is the identity matrix. $D(w_{aug}) \in \mathbb{R}^{(q+2)\times(q+2)}$ is a diagonal matrix with w_{aug} filling up the diagonal elements. For further details in state augmented Kalman filters see [12].

Under ideal conditions, for a perfectly working Kalman filter, the autocorrelation of the innovations sequence has the statistical characteristics of white noise (non-zero autocorrelation only at zero lag). To test the performance of the state augmented Kalman filter we used the second order whiteness test as defined in ([17], page 16-8, [18]). The test involves using the biased estimate of autocorrelation defined as

$$c_y(\tau) = \frac{1}{S}\sum_{t=1}^{S-t} y(t+\tau)y(t), \tau \geq 0 \tag{15}$$

to form the test statistic

$$T = \frac{S}{c_y^2(0)}\sum_{i=1}^{L} c_y^2(i). \tag{16}$$

Here the sequence y is zero mean with length S. The parameter L is a chosen maximum lag for whiteness testing. The test statistic T is distributed chi-squared, $\chi^2(L)$ (χ^2 with L degrees of freedom) if the sequence y is zero mean white. For a fixed significance level, the test statistic can be compared to a threshold to validate the second order whiteness. If T is greater than the threshold, the sequence y is declared non-white at the chosen significance level.

Figure 4 shows the performance of the state augmented Kalman filter on the ζ measurements from the Bellyband tag. The top left trace shows the raw observations (magenta) and filtered estimates (black). The innovation sequence is shown in the bottom left trace with the autocorrelation of the innovation presented in the top right trace. The autocorrelation does resemble that of white noise and was further corroborated as the innovation sequence passed the second order whiteness test mentioned above at 0.05 significance level.

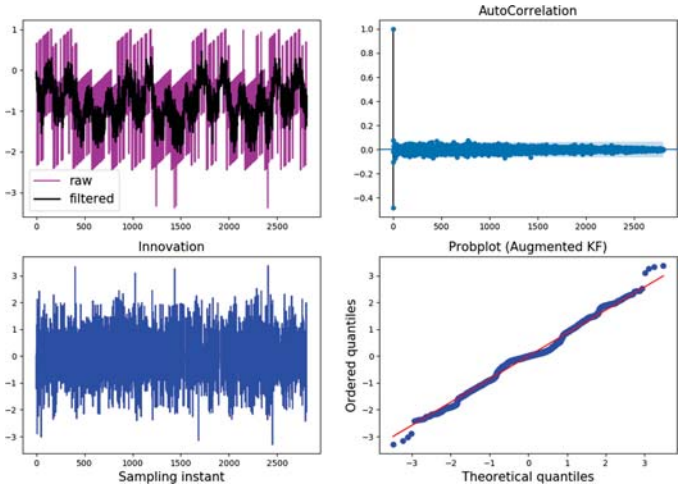

Fig. 4. Performance of state augmented Kalman filter on feature 1 : ζ. Clockwise from top left: raw observation (magenta) and filtered (black) observations, autocorrelation of the innovation sequence, Normal QQ plot of the innovation sequence, residuals from the filtering operation.

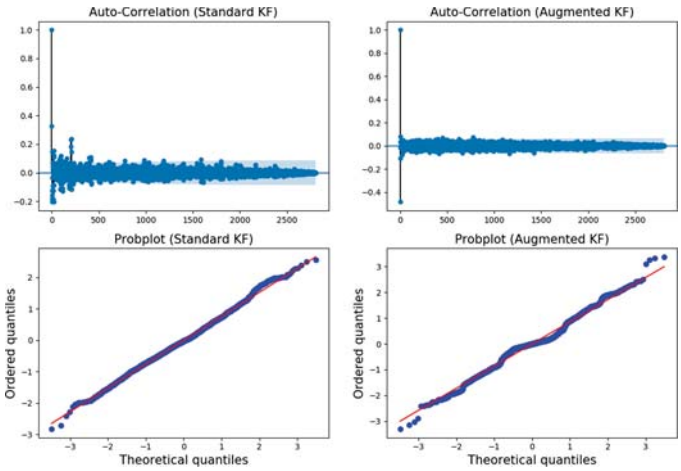

Fig. 5. Performance comparison of standard and state augmented Kalman filter on Bellyband tag measurements.

A comparison of the standard and augmented Kalman filter is shown in Figure 5. The top two traces show the autocorrelation of the innovations for standard (left) and augmented (right) Kalman filter. In case of the standard Kalman filter, multiple spikes at non-zero lags are observed which clearly shows that the innovation was not white (standard Kalman filter formulation failed the second order whiteness test at 0.05 significance level). The normal QQ plot for the augmented Kalman filter shows that the innovation sequence exhibits *Gaussian* characteristics.

B. Binary Classifiers

The data set collected from the Bellyband tag for the 4 selected features was used for training three binary classifiers namely logistic regression, decision tree and naive Bayes' (see [11] for details on individual classification algorithms). Each classifier was trained on a randomly sampled (without replacement) subset (85%) of the total training set. The random

TABLE II
BINARY CLASSIFIER PARAMETERS CHOSEN AFTER GRID SEACRH

Classifier	Parameters grid searched	Chosen parameters
Logistic Regression	Regularization weight (C)	0.1
	Regularization penalty	L2 (Ridge)
Decision Tree	Maximum depth	20
	Regularization penalty	L2 (Ridge)

sampling is a common process in ensemble learning and is used to introduce diversity in the samples on which the individual classifiers are trained on [19]. The entire dataset was of size 2805×4 with rows indicating the number of samples and columns indicating the 4 chosen features. Each classifier was trained on 2384 samples (85% of dataset). There were no separate validation sets as the classifier performances were verified with 5-fold cross-validation. Four (4) criteria, namely the classification accuracy, precision, recall [11], and the probability of error [20] were used to assess performance.

The classifier parameters were chosen through grid search. Table II shows the hyper parameters that were tuned.

For details about ridge regularization see [11]. The regularization weight C in logistic regression indicates the degree to which regularization is enforced. Lower the value of C, more the regularization. For the decision tree classifier, lower the tree depth, higher the bias, lower the variance and therefore lower the chances of overfitting. For the naive Bayes' classifier, the prior probablities of each class were estimated from the labeled dataset and the likelihoods were assumed to be Gaussian distributed.

C. Similarity Classifier

The assumption in this study is that the reference tag owing to the location (shoulder) is not significantly affected by breathing movements and hence the tag measurements are representative data points for the *non-breathing* state. Therefore, more similar the Bellyband tag measurement is to the reference tag measurement at a given time instant, the higher the confidence that hypothesis H_1 (non-breathing state) is true. Based on this rationale, the similarity classifier is designed to compute a binary decision on the set of hypotheses based on a weighted sum of how similar the Bellyband tag observation is to the reference tag measurement at the same time instant and the average of the past M local decisions (see Figure 3). Let us denote the measurement vector obtained from the belly-band tag at time instant k as $x_d(k)$ and the corresponding measurement from the reference tag as $x_r(k)$. We form a weighted sum $y(k)$ of two factors namely the similarity between x_d and x_r denoted by $sim(x_d, x_r)$ (we chose the cosine similarity measure [21] in this study) and the average of the last M decisions from the similarity classifier. The formulation is given as:

$$y(k) = \alpha sim(x_d, x_r) + (1-\alpha)\left(\frac{\sum_{i=1}^{M} u_4(k-i)}{M}\right). \quad (17)$$

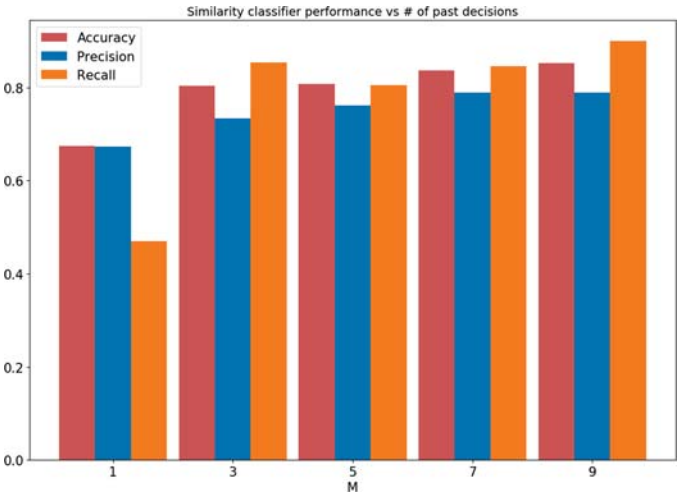

Fig. 6. Similarity classifier performance vs. number of past decisions (M).

Here $\alpha \in [0, 1]$ is a weighting factor that decides how to distribute support between the current measurement and past decisions. Since the cosine similarity $sim() \in [-1, 1]$, a scaled version of the similarity measure is used in (17) so that $y(k) \in [0, 1]$. We interpret $y(k)$ as the amount of confidence the decision maker accumulates based on the similarity between the current Bellyband and reference tag measurements, as well as the past decisions.

At every time instant the similarity classifier compares $y(k)$ to a threshold t to generate a binary decision $u_4(k)$. The fusion rule is given as

$$u_4(k) = \begin{cases} 1, & \text{if } y(k) \geq t \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

The weighting factor α and threshold $t \in [0, 1]$ were computed using grid search such that the similarity classifier produced the highest classification accuracy on the data sets.

The number of past decisions M to use in (17) is not obvious. The larger M is, more smoothed the classifier decisions will be. In other words, for large M, new sudden changes that are actually true and not due to noise may be ignored. On the other hand a small M may force the classifier to trust the most recent observations more and as a result make decisions influenced by random fluctiations possibly due to noise. We tested the classifier performance based on three (3) criteria namely, accuracy, precision, and recall for different values of M. Only odd values were chosen so as to break symmetry among past decisions. Figure 6 shows the performance variation as M was changed. It can be observed that there is only marginal improvement in accuracy from $M = 7$ to $M = 9$, whereas the precision remains same. As increasing M beyond 7 improves classification perfomance only marginally, for this study M was fixed at 7.

Role of the similarity classifier: Unlike the binary classifiers that use only the measurement at the current instant to decide on a class, the similarity classifier makes a decision based on a weighted combination of the current measurements and past M decisions. Incorporating past decisions induces a form of

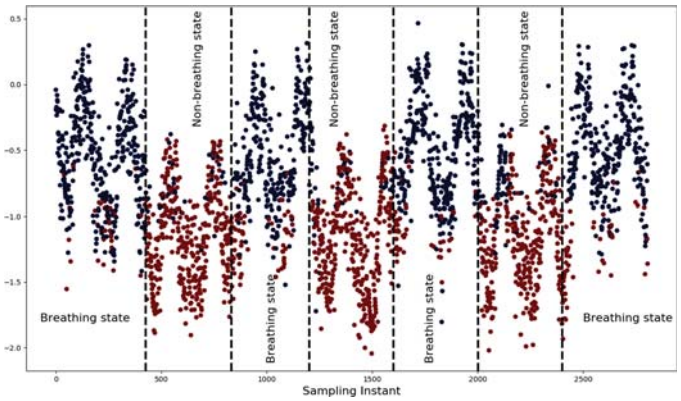

Fig. 7. Classification of data points from the Bellyband tag for the feature ζ. Red points are classified as H_1: non-breathing state, and blue points as H_0: breathing state.

inertia in the classifier which makes the classifier stick to the previous decisions unless there is very high evidence from the current measurement to decide otherwise. For example, let us consider the case when hypothesis H_0 is true and the similarity classifier have correctly classified majority of the past K measurements. At the current instant k let the measurement value be affected by unknown noise artifacts which shifts the measurement across the decision boundaries of the binary classifiers. In this case, based on the measurement at time instant k, the first three binary classifiers would decide H_1 for the current observation, whereas the similarity classifier would weigh in the past decisions (which were mostly supporting H_0) and has a higher chance of making the correct classification.

D. Decision Fusion

The local detector decisions u_i, $i = 1, ..., N$ ($N = 4$ in this study) are sent to a decision fusion center (DFC), where they are combined to generate an aggregate decision $u_0(k)$, to accept H_1 ($u_0(k) = 1$) or H_0 ($u_0(k) = 0$). Fusing local decisions not only provides a more robust classifier (the fusion center) that can overcome individual classifier deficiencies [19], [20] (fusion center decision is more aligned toward the general consensus among the individual classifiers) but also results in a modular detection framework where better classifiers can be added and inefficient classifiers removed without affecting the overall system setup. The fusion center computes a weighted sum of the local classifier decisions with each classifier's decision being weighted by the corresponding classifier's reliability given as $\frac{p}{1-p}$, where $p \in [0, 1]$ is the classifier accuracy. The class $H_i, i \in 0, 1$ that receives the highest weighted vote wins and is chosen as the aggregate decision $u_0(k)$ at sampling instant k.

V. Results

Figure 7 shows the classification performance of the ensemble system on the feature ζ. Three cycles of H_0: breathing state, H_1: non-breathing state are shown. The true classes are demarcated by the vertical lines. The points marked in red were classified as belonging to class H_1 whereas the blue points were classified as belonging to class H_0. Overall the ensemble system was able to demarcate the breathing and non-breathing state data points. Most of the mis-classifications occurred around the borders during the transition between states.

TABLE III
5-FOLD CROSS-VALIDATED CLASSIFICATION PERFORMANCE OF LOCAL CLASSIFIERS AND FUSION CENTER WITH KALMAN FILTERING

Classifier	Accuracy (%)	Precision	Recall	Probability of error
Logistic Regression	73.6	0.703	0.671	0.272
Decision Tree	85.5	0.832	0.834	0.147
Naive Bayes'	72.7	0.693	0.659	0.282
Similarity Classifier	83.7	0.79	0.846	0.162
Fusion Center	86.7	0.835	0.861	0.133

TABLE IV
5-FOLD CROSS-VALIDATED CLASSIFICATION PERFORMANCE OF LOCAL CLASSIFIERS AND FUSION CENTER WITHOUT KALMAN FILTERING

Classifier	Accuracy (%)	Precision	Recall	Probability of error
Logistic Regression	60.5	0.55	0.471	0.412
Decision Tree	83.3	0.804	0.808	0.17
Naive Bayes'	60.9	0.557	0.433	0.413
Similarity Classifier	73.2	0.686	0.693	0.273
Fusion Center	80.5	0.813	0.713	0.205

In Table III, we show the classification performance after 5-fold cross-validation based on four criteria: namely, the classification accuracy, precision, recall, and the probability of error when the input measurements where Kalman filtered. Fusion improves the classification performance with respect to all four chosen criteria. The fusion center performance is better than the best individual classifier and hence underlines the advantages of an ensemble system. However it should be noted that weighted majority vote does not guarantee improvement in performance [19] in all condtions. The ensemble system is however more robust and resilient to failures or abrupt input data inconsistencies. If the inputs to the classifiers are statistically independent conditioned on the hypotheses (this condition was not satisfied in this study), optimal decision fusion rules such as [22] can be applied that guarantees overall improvement in performance. Among the individual classifiers, the decision tree classifier appeared to perform the best.

The similarity classifier was able to correctly classify around 68% of the data points that were mis-classified by two or more of the other three classifiers (logistic regression, decision tree, and naive Bayes'). The correct classification by the similarity classifier was due to the decision making process which relied on both the current measurement and the past M decisions. It can also be observed that the similariy classifier had the best true detection rate (identifying non-breathing states correctly) as shown by the highest recall value among the individual classifiers.

Table IV shows the deterioration in 5-fold cross-validated performance when raw (unfiltered) observations were used for classification. Overall there was around 7.7% (80.5% to 86.7%) improvement in accuracy when Kalman filtered measurements were used in the classification setup. The decision tree clas-

sifier outperformed the fusion center in this case in terms of accuracy, recall, and probabiltiy of error. Although the individual classifier provides a small but useful improvement in classification, the fusion center yields constructive sensor fusion that improves upon the classification of individual sensors used independently. Because the filter improves classification of each individual sensor (see Tables III and IV) as well as the subsequent fused decision, and many classification errors occur at the transition boundary between states (i.e., from breathing to non-breathing or *vice-versa*), the use of an augmented Kalman filter as an input to the classifier is of particular benefit. The filtering mechanism would alleviate discrepancies arising when classifiers disagree due to the state transition of the wearer as well as due to noise already inherent in the signal. Although the addition of training data will always theoretically improve the classification accuracy, there are several limitations to training a physiological process such as respiration. These include environmental parameters such as the position of the wearer and possible interference from surrounding objects in the environment. Additionally, the probability of mis-classification diminishes exponentially with the number of training samples and with the square of the target accuracy tolerance [23]–[25]. This tolerance is bounded more tightly in practice than that provided by the theoretical model [26], suggesting that additional training data is of theoretical benefit that diminishes over time. Analyzing improvement in performance with additional training data may be an interesting future study, but may be inconclusive or incompatible with the practical limitation of manually training the device prior to use.

VI. Conclusion

This study used multiple feature measurements obtained from a Bellyband RFID tag and reference measurements from a reference tag to assess the breathing state of a programmable mannequin simulator. The reference tag was placed on the shoulder and measurements obtained were assumed to be not significantly influenced by breathing related movements. State augmented Kalman filtering aided in handling temporally correlated measurement noise and provided smoothed estimates of the features. Three binary classifiers were used to classify an incoming Kalman filtered observation vector into one of the two classes (H_1 or H_0). A new classifier named the similarity classifier was developed to incorporate information from a reference tag. The similarity classifier used both current measurements and past decisions to generate a local decision which made it good at correctly classifying observations mis-classified by majority of the other classifiers. An aggregate decision was generated by a fusion center that combined individual classifier decisions using a weighted majority vote rule. Through the proposed architecture, we showed a framework that can aid in real-time remote monitoring of breathing rates using noninvasive RFID tags. A similar system could also be used for sleep apnea detection; this hypothesis-based classification is suitable for integration into a software monitoring framework that has been used in an IRB-approved human trial applications of our technology and algorithms for respiratory and apnea classification [27]. For example, the proposed system could be used in real-time to keep track of the time interval for which continuous "non-breathing" state is detected and if the interval is above a threshold the onset of apnea could be inferred. Provided relevant data are available for training, the system could be modified to generate decisions on multi-levels of breathing (normal, reduced, absence). Such a system may be beneficial for detecting apnea events and lead to better tracking and diagnosis. Our current effort involves generating synthetic training data to simulate the "non-breathing" class for purposes of human-based classification. We envision generating such synthetic anomaly data in real-time, providing it to the augmented Kalman filter-based classifier system, and thus enabling human apnea detection as a result of this effort.

The similarity classifier proved to be able to distinguish points that were mis-identified by majority of the other trained classifiers which did not use past decisions. In future studies, multiple reference tags may be used to consolidate the similarity classifier decisions optimally.

Instead of a binary classification framework, the proposed setup (see Figure 3) can be altered to provide decisions on levels of breathing using multi-class classifiers or by comparing the distance of a data point from the classifier decision boundary to a graded scale.

Acknowledgment

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] D. Patron *et al.*, "On the use of knitted antennas and inductively coupled RFID tags for wearable applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 6, pp. 1047–1057, Dec. 2016.

[2] W. Mongan *et al.*, "A multi-disciplinary framework for continuous biomedical monitoring using low-power passive RFID-based wireless wearable sensors," in *Proc. IEEE Int. Conf. Smart Comput.*, May 2016, pp. 1–6.

[3] S. Vora, K. Dandekar, and T. Kurzweg, "Passive RFID tag based heart rate monitoring from an ECG signal," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 4403–4406.

[4] R. Begg and M. Palaniswami, *Computational Intelligence for Movement Sciences: Neural Networks and Other Emerging Technologies*. Hershey, PA, USA: Idea Group Publishing, 2006.

[5] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 837–846. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2702123.2702200

[6] R. Ravichandran, E. Saba, K. Y. Chen, M. Goel, S. Gupta, and S. N. Patel, "WiBreathe: Estimating respiration rate using wireless signals in natural settings in the home," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2015, pp. 131–139.

[7] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, and G. Marrocco, "RFID technology for IoT-based personal healthcare in smart spaces," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 144–152, Apr. 2014.

[8] P. Guay, S. Gorgutsa, S. LaRochelle, and Y. Messaddeq, "Wearable contactless respiration sensor based on multi-material fibers integrated into textile," *Sensors*, vol. 17, no. 5, p. E1050, 2017.

[9] J. Han *et al.*, "CBID: A customer behavior identification system using passive tags," in *Proc. IEEE 22nd Int. Conf. Netw. Protocols*, Oct. 2014, pp. 47–58.

[10] Z. Su, S. C. Cheung, and K. T. Chu, "Investigation of radio link budget for UHF RFID systems," *Proc. IEEE Int. Conf. RFID-Technol. Appl.*, Jun. 2010, pp. 164–169.
[11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer Series in Statistics)*. New York, NY, USA: Springer-Verlag, 2013. [Online]. Available: https://books.google.com/books?id=yPfZBwAAQBAJ
[12] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. New York, NY, USA: Wiley, 2006.
[13] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall Inc., 1979.
[14] S. Hu, *Akaike Information Criterion*. Raleigh, NC: Center for Research in Scientific Computation, 2007.
[15] D. Dreano, P. Tandeo, M. Pulido, B. Ait-El-Fquih, T. Chonavel, and I. Hoteit, "Estimating model-error covariances in nonlinear state-space models using kalman smoothing and the expectation-maximization algorithm," *Quart. J. Roy. Meteorol. Soc.*, vol. 143, no. 705, pp. 1877–1885, 2017.
[16] X. Luo, R. J. Lorentzen, A. S. Stordal, and G. Nævdal, "Toward an enhanced Bayesian estimation framework for multiphase flow soft-sensing," *Inverse Problems*, vol. 30, no. 11, Art. no. 114012, 2014.
[17] V. Madisetti, *The Digital Signal Processing Handbook (Electrical Engineering Handbook)*. Boca Raton, FL, USA: Taylor & Francis, 1997. [Online]. Available: http://books.google.com/books?id=Zhc36gyobk0C
[18] P. Stoica, "A test for whiteness," *IEEE Trans. Autom.Control*, vol. 22, no. 6, pp. 992–993, Dec. 1977.
[19] R. Polikar, "Ensemble based systems in decision making," *Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 2006.
[20] P. K. Varshney, "Multisensor data fusion," *Electron. Commun. Eng. J.*, vol. 9, no. 6, pp. 245–253, Dec. 1997.
[21] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *City*, vol. 1, no. 4, pp. 300–307, 2007.
[22] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.
[23] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, Mar. 1963. [Online]. Available: http://www.jstor.org/stable/2282952?
[24] V. Vapnik, E. Levin, and Y. L. Cun, "Measuring the vc-dimension of a learning machine," *Neural Comput.*, vol. 6, pp. 851–876, 1994.
[25] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning From Data*. USA: AMLBook, 2012.
[26] D. Hush and B. Horne, "Progress in supervised neural networks," *IEEE Signal Process. Mag.*, vol. 10, pp. 8–39, Jan. 1993.
[27] W. Mongan *et al.*, "Data fusion of single-tag rfid measurements for respiratory rate monitoring," in *Proc. IEEE Signal Process. Med. Biol. Symp.*, Dec. 2017, pp. 1–6.