2-24-2020

Cyber-Physical Security with RF Fingerprint Classification through Distance Measure Extensions of Generalized Relevance Learning Vector Quantization

Trevor J. Bihl
Air Force Research Laboratory

Todd J. Paciencia

Kenneth W. Bauer Jr.
Air Force Institute of Technology

Michael A. Temple
Air Force Institute of Technology

Recommended Citation

Bihl, T. J., Paciencia, T. J., Bauer, K. W., & Temple, M. A. (2020). Cyber-Physical Security with RF Fingerprint Classification through Distance Measure Extensions of Generalized Relevance Learning Vector Quantization. Security and Communication Networks, 2020, 1–12. https://doi.org/10.1155/2020/3909763

Hindawi
Security and Communication Networks
Volume 2020, Article ID 3909763, 12 pages
https://doi.org/10.1155/2020/3909763

Research Article

Cyber-Physical Security with RF Fingerprint Classification through Distance Measure Extensions of Generalized Relevance Learning Vector Quantization

Trevor J. Bihl,[1] **Todd J. Paciencia**,[2] **Kenneth W. Bauer Jr.**,[2] **and Michael A. Temple**[2]

[1]*Air Force Research Laboratory (AFRL) Sensors Directorate, Wright-Patterson AFB, Dayton, OH 45433, USA*
[2]*Air Force Institute of Technology (AFIT), Wright-Patterson AFB, Dayton, OH 45433, USA*

Correspondence should be addressed to Trevor J. Bihl; trevor.bihl@gmail.com

Received 6 March 2019; Revised 18 June 2019; Accepted 9 August 2019; Published 24 February 2020

Academic Editor: Sherif Abdelwahed

Radio frequency (RF) fingerprinting extracts fingerprint features from RF signals to protect against masquerade attacks by enabling reliable authentication of communication devices at the "serial number" level. Facilitating the reliable authentication of communication devices are machine learning (ML) algorithms which find meaningful statistical differences between measured data. The Generalized Relevance Learning Vector Quantization-Improved (GRLVQI) classifier is one ML algorithm which has shown efficacy for RF fingerprinting device discrimination. GRLVQI extends the Learning Vector Quantization (LVQ) family of "winner take all" classifiers that develop prototype vectors (PVs) which represent data. In LVQ algorithms, distances are computed between exemplars and PVs, and PVs are iteratively moved to accurately represent the data. GRLVQI extends LVQ with a sigmoidal cost function, relevance learning, and PV update logic improvements. However, both LVQ and GRLVQI are limited due to a reliance on squared Euclidean distance measures and a seemingly complex algorithm structure if changes are made to the underlying distance measure. Herein, the authors (1) develop GRLVQI-D (distance), an extension of GRLVQI to consider alternative distance measures and (2) present the Cosine GRLVQI classifier using this framework. To evaluate this framework, the authors consider experimentally collected *Z*-wave RF signals and develop RF fingerprints to identify devices. *Z*-wave devices are low-cost, low-power communication technologies seen increasingly in critical infrastructure. Both classification and verification, claimed identity, and performance comparisons are made with the new Cosine GRLVQI algorithm. The results show more robust performance when using the Cosine GRLVQI algorithm when compared with four algorithms in the literature. Additionally, the methodology used to create Cosine GRLVQI is generalizable to alternative measures.

1. Introduction

Cyber-physical systems (CPSs) are increasingly found in critical infrastructure (CI) applications to enable the industrial Internet of Things (IoT), with ever-increasing security implications (e.g., [1]). CPS in industrial uses, e.g., energy systems, integrates computing, communications, and control and must be dependable, safe, and secure and enable real-time operations [2]. Due to the gravity of CI systems, accurate identification and authentication of communication devices is important. Radio frequency (RF) fingerprinting extracts RF signals and develops classifier models to provide discrimination between communication devices [3, 4]. RF Distinct Native Attributes (RF-DNA) fingerprinting extends this general process by developing statistical features from regions of RF signals and has been shown to robustly enable biometric-like identification at the serial number level [4]. For this task, one needs classifier algorithms from which one can build models to discriminate between classes based on minute differences, as well as providing reliable authentication during masquerade attacks [4, 5].

Recent advances in classifier applications for RF fingerprinting include (1) discriminant analysis, (2) Generalized Relevance Learning Vector Quantization Improved (GRLVQI), (3) learning from signals (LFS), and (4) random

forests [6, 7]. Of these methods, the GRLVQI is one of the most robust methods, but algorithmically, GRLVQI has limitations that need to be addressed when considering the minute variations inherent when discriminating devices as the serial number level.

GRLVQI and the LVQ family of algorithms, in general, are gradient-descent-based algorithms that compute a distance from each exemplar to the nodes, termed prototype vectors (PVs), and then find the nearest PV to the exemplar [8]. The standard distance measure in GRLVQI, and all LVQ algorithms, is the linear squared Euclidean distance measure [9]. However, Euclidean distances present limitation because they are adversely affected by high levels of dimensionality [10]. In addition to this, Euclidean distances are scale variant while being translational invariant [11]; alternatively, for example, a cosine distance is translational variant but scale invariant [11]. A measure that is scale invariant but translational variant provides a potential to enable better classification of groups which differ based on minute, fingerprint-like variations.

This work extends GRLVQI, and LVQ in general, to use a cosine distance measure. This is a nontrivial modification since the distance measure is an implicit part of the cost function in all LVQ algorithms. Updating LVQ algorithms for alternative distance measures thus requires computing the first derivative (gradient) of the new distance measure to appropriately incorporate it into a revised cost function. This is an important matter to consider and often neglecting, e.g., [12–15]. This paper addresses this limitation by presenting GRLVQI-D (distance) which is a straightforward framework for incorporating distance measure extensions of GRLVQI and LVQ in general, with an example using a cosine distance measure. Modifying GRLVQI is notably complex since it includes multiple embellishments, e.g., both a sigmoid-based cost function and a relevance learning approach. The new Cosine GRLVQI classifier is then applied to an example CPS application in the form of RF fingerprinting an experimentally collected *Z*-wave wireless personal area network (WPAN) dataset. Classification results show that Cosine GRLVQI offers a distinct performance advantage over the original GRLVQI, as well as over (1) the original GRLVQI with optimized settings and (2) MDA.

This paper is organized as follows: Section 2 discusses the CPS environment and the need for reliable CPS identification methods. Section 3 discusses LVQ algorithms in general and GRLVQI in specific. Section 4 develops a straightforward approach to changing the distance measures in GRLVQI to any differentiable measure, with a specific example presented using cosine distance. Finally, results are presented in Section 5 showing a distinct classification performance advantage when using cosine GRLVQI over the baseline squared Euclidean distance method. Section 6 then concludes the paper.

2. Cyber-Physical System (CPS) Device Identification

CPS serve as a backbone for IoT connectivity ranging from critical infrastructure to home automation. Of interest is adopting a biometric-inspired approach, which involves three important steps: library creation, classifier model development, and classifier model verification [16]. Library creation involves selecting and measuring appropriate signatures, classifier model development involves selecting appropriate algorithms which can discriminate between signatures, and classifier model verification involves the robustness of the trained classifier to a claimed identity. Assembled effectively, a library and quality classifier model facilitate characterizing the system and understanding normal operations, while the verification approach enables monitoring for intrusion detection and thus fits into general autonomic visions of self-protecting IoT systems [17].

Security in CPS largely focuses on bit-centric Network (NKW) layer and Media Access Control (MAC) sublayer improvements [18]. One can view the various security measures as [18]

(1) "Something you know" (NWK—encryption keys)

(2) "Something you have" (MAC—MAC address)

(3) "Something you are" (PHY—RF fingerprints)

Such commonsense understandings relate how bit-level device identification credentials are easily spoofed and exploited by hackers. In a biometric understanding, one can consider a MAC address as the claimed identity to be verified using PHY-layer knowledge.

PHY-based security measures provide one remedy to these deficiencies and operate by either (1) incorporating physically traceable components to devices [19] or (2) RF fingerprinting which exploits inherent characteristics of the signal [20, 21]. Herein, RF fingerprinting is primarily of interest since it does not require retrofitting CPS devices or changing well-established CPS manufacturing approaches to include physically traceable components.

2.1. Radio Frequency (RF) Fingerprinting. Fingerprinting in communication systems considers physical layer (PHY) attributes, which are intrinsic to a specific device. RF fingerprinting extracts statistical features from RF signals and enables biometric-like identification of communication devices [3, 4]. Thus, principles from biometrics (see [16, 22]) and digital forensics (see [23]) are leveraged to develop, select, and identify RF features which have the general biometric qualities of universality, distinctiveness, permanence, and collectability [24]. A variety of RF fingerprinting approaches exist to accomplish this, which includes both transient and steady-state methods [24, 25].

Steady-state methods are of interest herein and consider specific segments of RF emissions. Of interest in steady-state RF fingerprinting is comparing emissions from predefined signal characteristics, e.g., preambles, to discriminate devices via unique features, i.e. from production variations [4, 26]. Thus, one can be divorced from attack modes, e.g., the types/volumes of data being transmitted, and focus on identifying individual devices based on minute variations in the selected region [24].

Of interest, herein, is employing the systematic RF Distinct Native Attributes (RF-DNA) approach [4], which

adopts the steady-state methodology. RF-DNA fingerprinting, conceptualized in Figure 1, works by (1) extracting the region of interest (ROI) of a signal, (2) performing signal processing to extract features (instantaneous amplitude (a), phase (ϕ), and frequency (f)) from the ROI, (3) computing statistical fingerprint features (variance (σ^2), skewness (γ), and kurtosis (κ)) from the signal processing features, and (4) developing and deploying classifier models with these features [24].

In the RF-DNA fingerprinting process conceptualized in Figure 1, RF fingerprints are computed as statistical features from time-domain responses of instantaneous amplitude (a), phase (ϕ), and frequency (f) [4]. Each response is then divided into N_R contiguous, and equal length, intervals [4]. Within each interval statistics of variance (σ^2), skewness (γ) and kurtosis (κ) are computed along with additional features for the entire response [4].

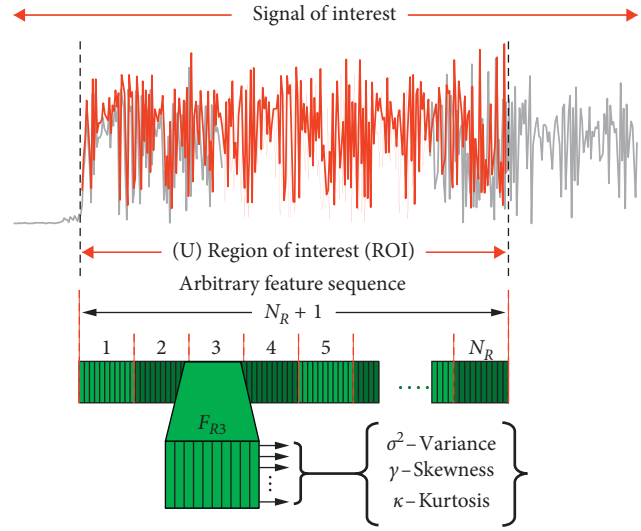

Figure 1: Conceptualization of RF fingerprinting from a given signal's region of interest (ROI).

2.2. Classification Algorithms for RF Fingerprinting. To discriminate between RF fingerprints and provide accurate identification of individual CPS devices, one needs to develop and train an effective classifier model. Herein, supervised classification is considered to develop a classifier model that takes labelled data, i.e. RF signatures and known identities, from the library of authenticated devices. From here, pattern recognition algorithms are employed to develop a mapping that separates the known identities (or groups) [27].

A variety of classifier algorithms have been considered for RF fingerprinting, see [6] for one example. Both GRLVQI and Multiple Discriminant Analysis (MDA) have seen consistent and successful use in RF fingerprinting discrimination [6]. MDA is considered, consistent with [6], to evaluate baseline performance. MDA operates via an eigenspace projection to find optimal linear separation between groups, where the underlying process extends Fisher's discriminant analysis to multiple classes [6]. MDA is computationally inexpensive, easy to interpret, and competitive with more complex algorithms [6]. Conversely, GRLVQI is more computationally intensive, but various applications can benefit from the nonlinear mappings inherent in GRLVQI and thus GRLVQI outperforms MDA depending on application [6].

Notably, GRLVQI and machine learning algorithms in general are highly sensitive to hyperparameter settings, such as learning rates and architecture size [28]. Although work has considered finding optimal settings for such algorithms, e.g., GRLVQI-SD, GRLVQI with optimized hyperparameters for Stochastic Optimization via Sequential Design of [28], such approaches are computationally costly with dozens of iterations needed to obtain improved algorithm settings. Additionally, such highly tuned hyperparameter values are often specific to the scope of the data and thus not useable on other datasets. Of interest herein is considering the well-known GRLVQI classifier in this domain and improving them to be a better fit to data that varies in only small, minute, dimensions, e.g., RF fingerprinting data. This extended classifier algorithm will then be compared against MDA, the baseline GRLVQI algorithm, and the optimized GRLVQI-SD.

2.3. Classification and Verification Performance. Assessing classifier performance involves using the appropriate performance measures. In RF fingerprinting and biometrics, in general, classification considers authorized fingerprints, which best discriminates them in a "1-vs-many" situation [24]. In contrast to this, verification takes the trained classifier model and a claimed identity, e.g., from a MAC address, and evaluates that claim as a device attempts to gain network access, e.g., a "1-vs-1" assessment [24]. Classification accuracy is considered as average percent correctly classified versus SNR (dB) operative points. Verification is considered as percentage correctly authorized in a one vs one claimed MAC address identity scenario [29]. Within both performance evaluation paradigms, a few measures are considered.

2.3.1. Classification Accuracy Measures. To evaluate classification performance, a plot of average percent correct classification ($\%C$) versus SNR is considered [29]. At each discrete SNR point on the plot, a classifier model was developed and trained for data at that SNR level. To provide for assessment and comparison of methods, both a gain measure and a Relative Accuracy Percentage (RAP) measure can be used [29].

Gain is defined, per [24, 29] as the reduction in required SNR, in dB, for a method to achieve the same $\%C$ as a reference method. Generally, *gain* is evaluated at an arbitrary benchmark of $\%C = 90\%$ accuracy [24, 29]. As stated in the study by Bihl et al. [29], gain values, G_{SNR} are interpreted as follows:

(1) $G_{\text{SNR}} < 0.0$ (negative gain), wherein a given method underperforms a baseline method by achieving the same $\%C$ as the baseline at a higher SNR

(2) $G_{SNR} = 0.0$, wherein a given method is indistinguishable in performance to the baseline method by achieving the same *%C* at the same SNR

(3) $G_{SNR} > 0.0$ (positive), wherein a given method outperforms a baseline method by achieving the same *%C* at a lower SNR

However, G_{SNR} can be insufficient for relative performance comparisons because it only considers one part of the *%C* vs. SNR curve [29]. To alleviate this deficiency, the authors introduced the RAP measure in the study by Bihl et al. [29] to provide classifier assessment over the entire curve.

RAP measures are computed by first taking the *%C* vs. SNR curve and computing the area under this curve via trapezoidal approximation [29]. This is known as the Area Under Classification Curve (*AUCC*). Since the *x*-axis is not bounded on a simple 0 to 1 interval, it can be nonintuitive to interpret raw AUCC values. Thus, the RAP measure enables relative comparisons by considering

$$\mathrm{RAP}_{\mathrm{method}} = \frac{\mathrm{AUCC}_{\mathrm{method}}}{\mathrm{AUCC}_{\mathrm{baseline}}}, \quad (1)$$

where $\mathrm{AUCC}_{\mathrm{method}}$ is the AUCC of a given method and $\mathrm{AUCC}_{\mathrm{baseline}}$ is the AUCC of the baseline algorithm [29]. As developed in the study by Bihl et al. [29], RAP is interpreted as follows:

(1) RAP < 1.0 indicates that a given method achieves overall lower *%C* than the baseline across all SNR

(2) RAP = 1.0, a given method achieves an overall *%C* comparable to the baseline

(3) RAP > 1.0 indicates that a given method achieves overall better *%C* than the baseline across all SNR

2.3.2. Verification Accuracy Measures. For verification, signatures are considered in a claimed identity scenario, e.g., authorized user attempting to access the network or a masquerade attack [5], where an "unknown" device claims the bit level credentials (e.g., MAC address). The RF fingerprint features are computed for this signature and the classifier model is used to compare these RF fingerprints against trained model. Verification performance is evaluated at a specified SNR, typically at the lowest SNR a $\%C = 90\%$ accuracy threshold, by taking the trained classifier model and querying it [29]. For evaluation, Receiver Operating Characteristic (ROC) curves are used, as described in Bihl et al. [24] and Dubendorfer et al. [30]. For authorized devices, ROC curves are plotted as True Verification Rate (TVR) versus False Verification Rate (FVR).

From the ROC curves, two approaches are used to evaluate performance, per the study by Bihl et al. [29]:

(1) The percentage authorized (*%Aut*) at an arbitrary $\mathrm{TVR} \geq 90\%$ at $\mathrm{FVR} \leq 10\%$ threshold

(2) The mean area of the ROC curves ($\mathrm{AUC}_{\mathrm{M}}$)

The $\mathrm{AUC}_{\mathrm{M}}$ approach was developed in the study by Bihl et al. [29] to avoid dichotomous performance results since *%Aut* reflects coarse sampling.

3. Learning Vector Quantization (LVQ) Classifiers

LVQ is an artificial neural network (ANN) approach that classifies data via lower-dimensionality maps. LVQ provides a supervised learning extension unsupervised self-organizing maps (SOMs) or vector quantization (VQ). Epistemologically, SOM algorithms are self-organizing ANNs [31], and a general example is conceptualized in Figure 2, which compares a typical three layer (input, hidden, and output) ANN in Figure 2(a) with a typical LVQ network in Figure 2(b). Of note is that the LVQ network does not have an outer layer, which maps the response of a typical ANN to the output class; LVQ networks operate differently and work to move the nodes, prototype vectors (PVs), to represent the underlying data through an iterative training process [8, 33]. LVQ considers each PV as associated with a specific class resulting in a "winner take all" approach where one and only one PV will win for each exemplar [8, 34–38].

The operation of LVQ is as follows: a distance measure is used to compute the distance of a given exemplar to all PVs. The PV that is closest to the exemplar is then selected for modification. If this PV has the same class label as the exemplar, it is moved closer to the exemplar; however, if the exemplar is misclassified, the PV is moved away [8]. The update process for PVs employs a general gradient descent:

$$w(t+1) = w(t) - \varepsilon(t)\nabla C(w(t)), \quad (2)$$

to train PVs with t being the training iteration number, $\varepsilon(t)$ being the learning rate, $w(t)$ being a given PV, $C(w(t))$ being the cost function, and ∇ implying the gradient [6, 9, 39]. The cost function in LVQ is the squared Euclidean distance used to find the distance between an exemplar and the PVs. Generally, LVQ algorithms train PVs by moving correctly classified PVs closer to a given class, and incorrectly classified PVs are moved away from a given class. Thus, LVQ is considered as a nearest neighbor approach to learning, and the nearest PV is iteratively moved to characterize the data [40].

Various extensions and embellishments of LVQ have been developed, differing in cost function, update logic, and the inclusion of additional computational steps (e.g., relevance computations) [41]. Kohonen [42] first extended LVQ by creating variants (e.g., LVQ2 and LVQ2.1) that improved the PV update strategy to updates involving both in-class and nearest out-of-class PVs. Further major LVQ variations are reflected through the addition of letters to the LVQ acronym, c.f. [41, 43]. One such algorithms is GRLVQI, which is decoded as follows: G (generalized): a sigmoidal cost function [44, 45], R (relevance): relevance learning [39, 46], and *I (improved)*: PV update logic and operation [9, 47].

Relevance LVQ (RLVQ) extended LVQ by incorporating a relevance weight for each data feature, which is learned during the training process [46]. GLVQ extends LVQ by improving class boundary approximations through the incorporation of a sigmoid cost function [44]. GRLVQ of Hammer and Villmann [39] combined the innovations of both GLVQ and RLVQ to create a GLVQ algorithm that

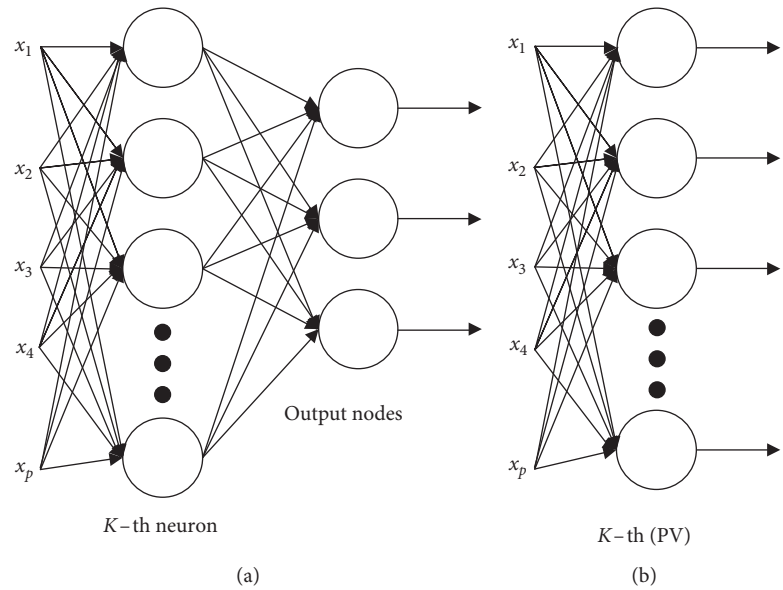

Figure 2: General conceptualization of an LVQ neural network, adapted from [32]. (a) Feedforward neural network. (b) Learning vector quantization.

learned the input dimension weights to provide relevance information regarding each feature. GRLVQ was then further extended by Mendenhall [9] through improvements resulting in the GRLVQI algorithm. In GRLVQI, GRLVQ is extended with the conscience learning of DeSieno [48], improved PV update logic, and a frequency-based maximum input update strategy [9, 47]. Despite any embellishment differences, all LVQ algorithms similarly employ the gradient-descent process seen in equation (1) to train PVs via *nearest neighbor* approaches.

3.1. Generalized Relevance Learning Vector Quantization Improved (GRLVQI). GRLVQI has been applied to RF fingerprinting due to its inherent nonlinearities and potential to better learn nonlinear data manifolds over MDA and linear methods. For GRLVQ and GRLVQI, the underlying cost function for equation (2) is

$$C(w(t)) = \sum_{m=1}^{N_{\text{Samples}}} f(\mu(x^m)), \tag{3}$$

where $f(\mu(x^m))$ is a sigmoid and ψ are the relevance scores:

$$f(\mu(x^m)) = \frac{1}{1 + e^{-\mu(x^m)}}, \tag{4}$$

and $\mu(x^m)$ is a relative distance difference:

$$\mu(x^m) = \tau \frac{(d^J - d^K)}{(d^J + d^K)}, \tag{5}$$

with τ being a GRLVQI rate (implicitly, in GRLVQ $\tau = 1$), and d^J and d^K being distances between the exemplar x^m and the *in-class* PV, w^J and *out-of-class* PV, w^K, respectively [9, 39]. Nominally, d^J and d^K are computed via a squared Euclidean distance as

$$d^{J,K} = \left\| x^m - w^{J,K} \right\|^2. \tag{6}$$

To determine the PV update expressions for equation (3), the gradient descent for GRLVQ-type algorithms is then the gradient by chain and quotient rules multiplied by the learning rate, $\varepsilon(t)$, and a differential shifting. The process yields

$$w^{J,K}(t+1) = w^{J,K}(t) \pm \frac{4\varepsilon(t)(\partial f / \partial \mu(x^m)) d^{K,J}}{(d^J + d^K)^2} \Psi \cdot \left(x^m - w^{J,K}\right), \tag{7}$$

where the superscript indicates if a positive (+) update, for *in-class* PVs, or a negative (–) update, for *out-of-class* PVs, is performed. In equation (7), the numerator includes the distance $d^{K,J}$, indicating that d^K is used for w^J and d^J is used for w^K. Relevance learning in GRLVQI then involves a further gradient descent:

$$\psi(t+1) = \psi(t) - \xi(t) \nabla C(\psi), \tag{8}$$

where for a specific *q-th* feature, the derivative is computed with respect to the relevance ψ [9] as

$$\psi_q = \psi_q - \xi(t) f' \Big|_{\mu\left(x_q^m\right)} \left(\frac{d^K}{\left(d_\lambda^J + d_\lambda^K\right)^2} \left(x_q^m - w_q^J\right)^2 - \frac{d^J}{\left(d_\lambda^J + d_\lambda^K\right)^2} \left(x_q^m - w_q^K\right)^2 \right). \tag{9}$$

4. Distance Extensions for GRLVQ and GRLVQI Classifiers

Despite the various extensions, GRLVQI, as well as many LVQ variations, relies on the linear squared Euclidean distance measure. As noted above, in the introduction,

Euclidean distances present limitations because they are (1) adversely affected by high levels of dimensionality and (2) scale variant while being translational invariant. Of interest is thus extending GRLVQI to use non-Euclidean distances. Focusing on only one distance measure variation is not efficient, and thus, developing a framework to incorporate any desired distance measure is of interest. While distance extensions of GRLVQ were introduced by [49, 50], these were not easily generalizable. Of interest is thus developing a straightforward approach to incorporating alternative distances.

4.1. GRLVQI-D: Distance Extension Framework for GRLVQI. Because LVQ algorithms are trained using gradient descents, changing the distance measure necessarily requires computing the first derivative of the distance measure for appropriate inclusion into the cost function. Herein, distance-based extensions of LVQ, specifically GRLVQ and GRVLQI, are considered as a gradient-descent process. We can develop GRLVQI-D, a straightforward approach to changing GRLVQI distance measures by considering the various derivational pieces that are represented in equation (7). Using the developed derivative framework, GRVLQ and GRLVQI could be further extended with any differentiable distance measure. To accomplish this, we can represent equation (7) as

$$w^{J,K}(t+1) = w^{J,K}(t) \pm c\varepsilon(t) F_{\mu} R_{\mathrm{d}} D_m, \tag{10}$$

and equation (9) as

$$\psi_q = \psi_q - \xi(t) F_{\mu}\left(R_{\mathrm{d}} d^{J,K} - R_{\mathrm{d}} d^{J,K}\right), \tag{11}$$

where c is a constant, $c = 4$ for nominal GRLVQI, and

(1) $F_{\mu} = \partial f / \partial \mu(x^m)$; the component related to the derivative of the cost function and sigmoid equations (3) and (4)

(2) $R_{\mathrm{d}} = d^{K,J} / (d^J + d^K)^2$; the component related to the derivative of the relative distance difference metric equation (5)

(3) $D_m = \Psi(x^m - w^{J,K})$; the component related to the derivative of the squared Euclidean distance with relevance

Thus, for example, if one changes the distance measure, only D_m must be changed in equation (10), whereas the remainder of the expression is unchanged.

One final extension must be considered. Squared Euclidean distances are always positive and ensures that in equation (4), $\mu(x^m) \in [-1, +1]$, which is desirable to avoid $\mu(x^m)$ from creating unstable results. Nonsquared distances do not necessarily ensure a positive distance. Thus, the authors extend equation (5) to

$$\mu(x^m) = \tau \frac{(d^J)^2 - (d^K)^2}{(d^J)^2 + (d^K)^2}. \tag{12}$$

This creates a squared measure to ensure that $\mu(x^m) \in [-1, +1]$.

4.2. Cosine GRLVQI. As an example of using the straightforward GRLVQI-D process presented in Section 4.1, the authors will use this process to derive a Cosine GRLVQI algorithm. As mentioned previously, a cosine distance could be useful for discriminating exemplars that are similar in operational characteristic but differ based on minute characteristics, e.g., biometrics. A cosine distance measure is a similarity measure that computes the cosine angle of two vectors [51], i.e. a measure of orientation and not magnitude of the distance (translation variance but scale invariance). The cosine distance measure can be formulated as

$$d_{\cos} = \sum_{i=1}^{N_F} \frac{x_i w_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} w_i^2}}. \tag{13}$$

Following the discussion in Sections 3.1 and 4.1, a relevance learning can be formulated as

$$d_{\cos} = \sum_{i=1}^{N_F} \frac{\psi_i x_i w_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} w_i^2}}, \tag{14}$$

with its derivative via the quotient rule being

$$\frac{\partial d_{\cos}}{\partial w} = \sum_{i=1}^{N_F} \psi_i \frac{\left(x_i \sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} w_i^2} - x_i w_i^2 \sqrt{\sum_{i=1}^{n} x_i^2} / \sqrt{\sum_{i=1}^{n} w_i^2}\right)}{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} w_i^2}. \tag{15}$$

Considering equation (14) with a derivative for relevance yields the following:

$$\frac{\partial d_{\cos}}{\partial \psi} = \sum_{i=1}^{N_F} \frac{x_i w_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} w_i^2}}. \tag{16}$$

Since cosine distance measures do not ensure a positive distance, the formulation of equation (12) will be used. Applying the quotient rule to equation (16) with $v = (d^J)^2 + (d^K)^2$, $u = (d^J)^2 - (d^K)^2$, and $v^2 = ((d^J)^2 + (d^K)^2)^2$ yields

$$R_{\mathrm{d}} = \left(\frac{\partial (d^{J,K})^2}{\partial w^{J,K}}\right) \frac{\left(2 (d^{K,J})^2\right)}{\left((d^J)^2 + (d^L)^2\right)^2}, \tag{17}$$

where $\partial (d^{J,K})^2 / \partial w^{J,K}$ is the product of $2d^{J,K}$ and equation (11) for Cosine. Taking $D_m = \partial / \partial w^{J,K} [(d_{\cos})^2]$ and inserting equation (17) into equations (10) and (11) produces the Cosine GRLVQI update expressions.

5. Application and Example Results

To enable the industrial IoT, CPS devices are finding increasing use in critical infrastructure, smart metering, and home automation [2, 52]. CPS devices employ either open or proprietary protocols, with open protocols offering more aftermarket security options, but possibly more threats, while proprietary protocols offer "security through obscurity," but less additional aftermarket security options [53]. Of the proprietary wireless protocols, the most commonly used are Z-wave, which is based on the International Telecommunications Union—Telecommunications (ITU-T)

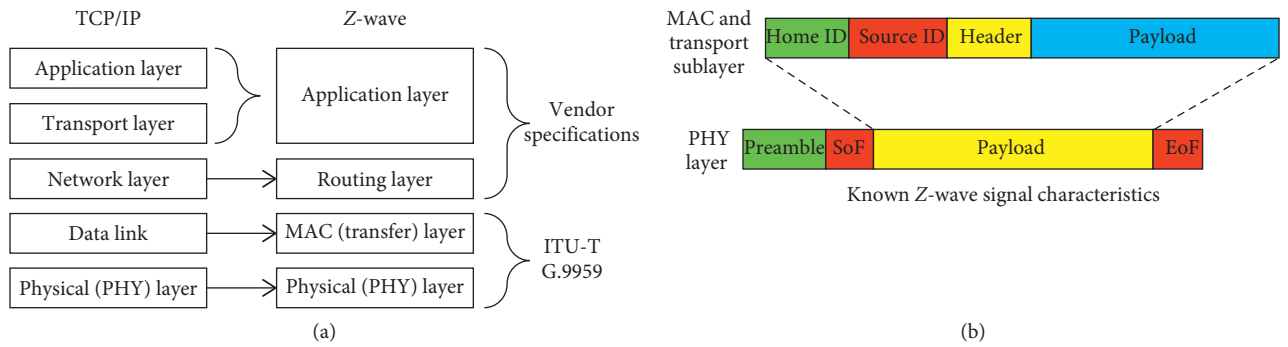

Figure 3: Z-wave characteristics: (a) protocol, and (b) signal. Extended from discussions in [56, 58, 60].

G.9959 recommendation [52]. Since Z-wave is known to have security vulnerabilities [54], vetting the claimed identity of Z-wave devices through RF fingerprinting is important.

5.1. Z-Wave Devices. Z-wave wireless communication devices are low-cost WPAN technologies used primarily for residential automation and similar in operations yet simpler to work with when compared with previously described devices [55–58]. However, Z-wave is generally considered as less secure than other WPAN technologies due to (1) an original lack of built in encryption [56] and (2) a proprietary standard that makes it difficult for third parties to provide enhancements [58].

Integration of Z-wave devices with IoT largely involves vendors. To produce a Z-wave device, and thus gain access to the proprietary Z-wave standard, a vendor must coordinate sign a Non-Disclosure Agreement (NDA) with Sigma Designs [59]. Vendors then gain access to hardware and software to develop Z-wave [59]. However, without a signed NDA, only general characteristics of the Z-wave protocol are known [59].

General Z-wave signal characteristics are known and presented in Figure 3 and Table 1. To facilitate incorporation of Z-wave with other communication devices, Z-wave follows the ITU-T G.9959 protocol at the physical layer (PHY) and medium access (MAC) layer [61]. However, the routing and application layer specifications are proprietary. Thus, third-party security at the network and routing levels is very difficult.

To identify Z-wave devices, Z-wave communications follow a predefined preamble and Start of Frame (SoF) [60], which is conceptualized in the PHY packet structure seen in [56, 58, 60]. Since the preamble and SoF should not vary from device to device, of interest is collecting such regions of interest for comparison and fingerprinting of devices at a serial number level.

5.2. Signal Collection and Dataset Generation. Consistent with [54, 62], $N_D = 3$ Aeotec Z-Stick S2 WPAN devices were considered in this research. Experimentally, each device was placed 10 cm from a vertically oriented LP0410 log periodic antenna [54]. The antenna was connected via a Gigabit

Table 1: General Z-wave device characteristics.

Device	Z-wave
Standard	Proprietary
Frequency	906 MHz
Bit rate	40 kbits/s
Security	None (200 and 300 series models) AES 128 (400 series models)
Latency	~1000 ms
Range	30–100 m
Message size (bytes)	64 (max)
Topology	Star, cluster, mesh

Ethernet cable to an NI USRP-2921 software defined radio with in-phase and quadrature (I/Q) samples collected as 16-bit integers, sampled at 2 Msps [62]. Amplitude-based leading edge detection with a −6 dB threshold was used for transmission (burst) detection [54]. Using this setup, a total of $N_P = 230$ preamble signals (the first segment of the signal per Figure 3(b), and the first 8.3 ms of the signal) were collected per device [54].

The collected data had Signal-to-Noise Ratio (SNR) at $\text{SNR}_C = 24.0$ dB and was like filtered [54]. To provide multiple operating points to consider noisy environments, Additive White Gaussian Noise (AWGN) was added to collected signals to achieve SNR ∈ (0 24) dB operating points in 2 dB steps [54]. Since the data collected were for 3 devices, this research does not consider identity impersonation attacks by "rogue" devices, and all devices were considered as "authorized."

Following the RF-DNA fingerprinting process in Section 2.1, Z-wave fingerprint generation parameters included the $N_{TD} = 3$ (a, ϕ, f) time-domain responses. Within each response, $N_R = 20$ regions per signal were considered, and $N_S = 3$ $(\sigma^2, \gamma, \kappa)$ RF fingerprint statistics were computed per region. Thus, a full-dimensional feature set had $N_F = 189$ features, which included statistics computed for each entire region. The Z-wave fingerprint features were divided into classifier model development, Training (TNG), and sequestered model assessment, Testing (TST), datasets using a 50% split. Thus, with $N_D \times N_P/2$, a total of $N_{Tng} = 345$ TNG fingerprints and $N_{Tst} = 345$ sequestered TST fingerprints were available per AWGN realization, each with 189 features. The TNG sets were used for

```
Select subset of {amplitude, frequency, phase}
Select subset of {variance, skewness, kurtosis}
Select classifier model {Cosine GRLVQI, GRLVQI, GRLVQI-SD, MDA}
Select classifier hyperparameters
for SNR = 0 to 24 dB in 2 dB steps do
    Select training data
    Train selected classifier model
    Classify test set
    Record classification accuracy
end for
    Select appropriate SNR and evaluate all possible combinations for
    verification accuracy
    Record verification accuracy
```

Algorithm 1: Classifier training process for RF fingerprinting.

Table 2: Classifier hyperparameter settings.

Classifier	Algorithm settings				
	A	*B*	*C*	*D*	*E*
Cosine GRLVQI (proposed herein)	0.025	0.005	2.5	0.35	10
GRLVQI-SD optimized [28]	0.078	0.016	2.527	0.319	7
GRLVQI (baseline) [29]	0.025	0.005	2.5	0.35	10
MDA [29]	N/A	N/A	N/A	N/A	N/A

classifier model development with TST sets were used to confirm results.

5.3. Classifier Algorithm Performance. Classifier models were developed for the *Z*-wave dataset using four classifiers: (1) the proposed Cosine GRLVQI (Section 3.2), (2) MDA, (3) the baseline GRLVQI of Harmer et al. [6], and (4) the GRLVQI-SD of Bihl and Steeneck [28]. Consistent with [24, 62], for these classifiers, the following process of Algorithm 1 was employed for each SNR.

Hyperparameter settings, e.g., learning rates, are critical to classifier performance and reproducibility. Developing and finding high-performing hyperparameter settings is research in itself [28, 29]. The hyperparameter settings used for classifiers in this paper are presented in Table 2. For both Cosine GRLVQI and the baseline GRLVQI, nominal algorithmic settings were used, consistent with [6], with (1) $\varepsilon = 0.025$, gradient descent learning rate; (2) $\xi = 0.005$, relevance learning rate; (3) $\gamma = 2.0$, conscience rate 1; (4) $\beta = 0.35$, conscience rate 2; and (E) $N_{PV} = 10$ PVs per class. GRLVQI-SD uses settings obtained from Stochastic Optimization via Sequential Design after 28 iterations. MDA has no distinct operational settings and was performed as described in [24].

Following the process of Algorithm 1 for each classifier presented in Table 2, we first evaluate classification performance at each SNR point. Figure 4 presents the classification results for each classifier on the sequestered TST set. Classification performance was evaluated in Figure 4 as Average Percent Correct (%C) on the *y*-axis and SNR (dB) on the *x*-axis, consistent with [6, 29]. Visible in Figure 4 is that both MDA and GRLVQI underperform Cosine

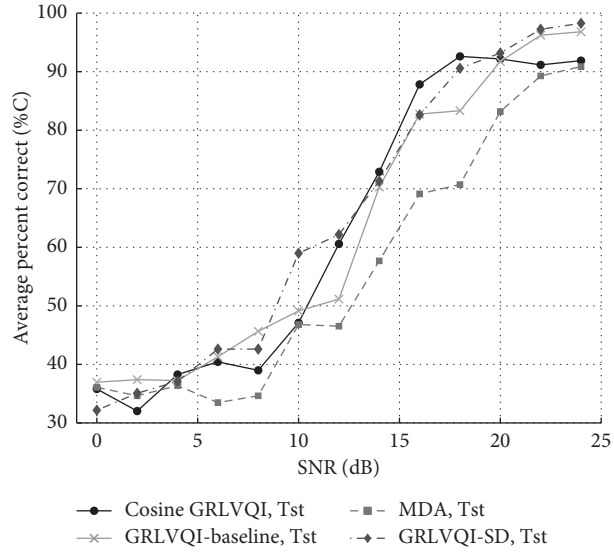

Figure 4: Classification performance with average percent correct versus SNR for testing (Tst) performance for Cosine GRLVQI, GRLVQI, MDA, and GRLVQI with settings optimized per [28].

GRLVQI and require higher SNR to achieve the same %C. GRLVQI-SD is seen to outperform Cosine GRLVQI at only SNR > 20 dB, where performance of all algorithms is largely over %C = 90%. Overall, at testing, Cosine GRLVQI is seen to provide over +6.00 dB gain at 90%C relative to MDA testing performance, whereas baseline GRLVQI offers only +3.32 dB gain at 90%C relative to MDA testing performance.

Verification performance is considered in Figure 5 as Receiver Operating Characteristic (ROC) curves. In Figure 5, models were evaluated at SNR = 20 dB where

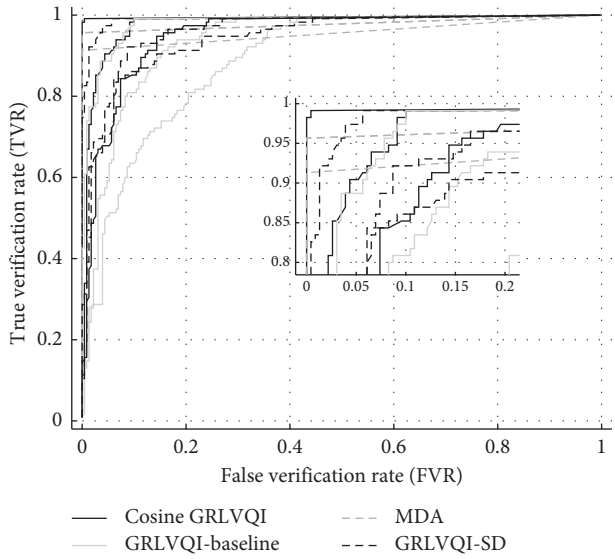

Figure 5: Verification performance at SNR = 20 dB for Cosine GRLVQI, GRLVQI, MDA, and GRLVQI with settings optimized per [28]. Inset shows performance in the [0%, 20%] FVR and [80%, 100%] TVR range.

Table 3: Classification and verification performance for each algorithm.

Method	Performance results					
	Classification				Verification at SNR = 20 dB	
	G_{SNR} (dB) at % C = 90%		RAP		*%Aut*	AUC_M
	TNG	TST	TNG	TST		
Cosine GRLVQI (proposed herein)	**+5.23**	**+6.00**	1.14	1.14	66	**0.973**
GRLVQI-SD optimized [28]	+5.16	+5.05	1.16	**1.17**	66	**0.965**
GRLVQI (baseline) [29]	+3.72	+3.32	1.14	1.13	33	0.936
MDA [29]	+1.68	0.00	**1.23**	1.0	**100**	**0.971**

Performance results bolded are the best values and those within 5% of the best values.

baseline GRLVQI %C = 90%, consistent with the results and process of Bihl et al. [54]. Verification performance shows MDA (dashed grey lines) and Cosine GRLVQI (solid black lines) both outperforming the baseline GRLVQI (solid grey lines) and GRLVQI-SD (dashed black lines). Notably, baseline GRLVQI only correctly authorizes 2 *Z*-wave devices, missing the third device by a considerable margin, e.g., the solid grey line that intersects TVR = 0.80 and FVR = 0.20.

When comparing MDA, Cosine GRLVQI, and GRLVQI-SD, the results are less clear. The insert in Figure 5 enlarges the 0–20% FVR and 80–100% TVR range and shows that the performance of these three classifiers for verification is very close, and that while only MDA provides 100% verification accuracy, both Cosine GRLVQI and GRLVQI-SD only barely miss the dichotomous *%Aut* threshold for this experiment. Since dichotomous results, e.g., for $ND = 3$ devices $\%Aut \in [0, 33, 66, 100]$, are not always reliable, as seen inspecting Figure 5, the authors also investigate AUC_M to enable relative performance differences to be evaluated between competing classifiers.

Table 3 presents performance for training (TNG) and testing (TST) data and shows an advantage of Cosine GRLVQI over GRLVQI-SD, GRLVQI, and MDA for classification. Verification performance in Table 3 was evaluated with binary grant/deny network access decisions based on a verification criteria, e.g., TVR ≥ 90% at FVR ≤ 10% with $\%Auth \in [0, 33, 66, 100]$ for $N_D = 3$. When evaluating AUC_M, the results show that these three methods (MDA, Cosine GRLVQI, and GRLVQI-SD) perform very similarly in verification performance with Cosine GRLVQI slightly outperforming all algorithms. As noted in discussing Figure 5, Cosine GRLVQI and GRLVQI-SD barely miss $\%Auth = 100\%$. Considering Table 3 overall, both MDA and Cosine GRLVQI outperform the baseline GRLVQI and GRLVQI-SD, while Cosine GRLVQI provides the best classification performance for this problem, thus illustrating the benefit of changing the cost function in GRVLQI.

6. Conclusions

Herein, the authors addressed problems in identifying cyber-physical systems (CPS) using radio frequency (RF) emissions. The authors considered the Generalized Relevance Learning Vector Quantization-Improved (GRLVQI) classifier algorithm, which is known to accurately classify RF fingerprint features but underperforms other methods

in the literature when protecting against masquerade attacks. Prior work optimized GRLVQI classifier settings to find improved operating points, but this is computationally costly and not generalizable. Since GRLVQI, and the LVQ family of algorithms, revolves around a distance measure to train their architecture, logically changing this distance measure can change results. However, LVQ algorithms are gradient descent based with the distance measure being related to the cost function of the gradient descent itself, thus changing the distance measure requires changing the derivatives. Since this can be complex, it has not been pursued previously.

The authors thus developed GRLVQI-D, a straightforward modularization of the GRLVQI algorithm and developed an update methodology which facilitates changing the distance measure of GRLVQI and GRLVQ, as well as other LVQ algorithms. To illustrate the application of the update methodology, the authors further developed a Cosine GRVLQI algorithm. Example results were then then presented for experimentally collected *Z*-wave RF data. RF fingerprints were developed for these devices to create a biometric library. Then, following a general biometric process, both classifier model development and identify verification were considered. Results show that the proposed Cosine GRLVQI algorithm outperforms both MDA and baseline squared Euclidean GRLVQI in classification and verification accuracy. Additionally, the results presented for Cosine GRLVQI were better than the iteratively obtained optimal results of GRLVQI-SD, which required 28 iterations to obtain improved algorithmic settings. Naturally, future extensions of this work would be to find optimal Cosine GRLVQI, per the approach of [28].

Data Availability

Publication is cleared for public release under case: 88ABW-2019-4252. Data results are publishable, but data and code are not cleared for public release.

Disclosure

This material is declared a work of the US Government and is not subject to copyright protection in the United States.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the U.S. Air Force Research Laboratory, Sensors Directorate. The views expressed in this article are those of the authors and do not reflect the official policy of the United States Air Force, Department of Defense, or the US Government.

References

[1] D. Shih, H. Chiang, B. Lin, and S. Lin, "An embedded mobile ECG reasoning system for elderly patients," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 854–865, 2010.

[2] T. Morris, A. Srivastava, B. Reaves et al., "Engineering future cyber-physical energy systems: challenges, research needs, and roadmap," in *Proceedings of the 41st North American Power Symposium*, pp. 1–6, IEEE, Starkville, MS, USA, October 2009.

[3] P. Padilla, J. F. Valenzuela-Valdés, and J. L. Padilla, "Radiofrequency identification of wireless devices based on RF fingerprinting," *Electronics Letters*, vol. 49, no. 22, pp. 1409-1410, 2013.

[4] W. E. Cobb, E. D. Laspe, R. O. Baldwin, M. A. Temple, and Y. C. Kim, "Intrinsic physical-layer authentication of integrated circuits," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 14–24, 2012.

[5] H. Kholidy and F. Baiardi, "CIDD: a cloud intrusion detection dataset for cloud computing and masquerade attacks," in *Proceedings of the 2012 Ninth International Conference on Information Technology-New Generations*, pp. 397–402, IEEE, Las Vegas, NV, USA, April 2012.

[6] P. K. Harmer, D. R. Reising, and M. A. Temple, "Classifier selection for physical layer security augmentation in cognitive radio networks," in *Proceedings of the 2013 IEEE International Conference on Communications (ICC)*, pp. 2846–2851, IEEE, Budapest, Hungary, June 2013.

[7] H. Patel, M. Temple, and R. Baldwin, "Improving ZigBee device network authentication using ensemble decision tree classifiers with radio frequency distinct native attribute fingerprinting," *IEEE Transactions on Reliability*, vol. 64, no. 1, pp. 221–233, 2014.

[8] S. Rogers and M. Kabrisky, *An Introduction to Biological and Artificial Neural Networks for Pattern Recognition*, SPIE Press, Bellingham, WA, USA, 1991.

[9] M. J. Mendenhall, *A neural relevance model for feature extraction from hyperspectral images, and its application in the wavelet domain*, Ph.D. dissertation, Rice University, Houston, TX, USA, 2006.

[10] J. V. Lee, *Nonlinear Dimensionality Reduction*, Springer, Heidelberg, Germany, 2007.

[11] A. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on web-page clustering," in *Proceedings of the Workshop on Artificial Intelligence for Web Search (AAAI)*, pp. 58–64, Austin, TX, USA, July 2000.

[12] Z. Chuan, L. Xianliang, and X. Qian, "A novel anti-spam email approach based on LVQ," in *Parallel and Distributed Computing: Applications and Technologies*, vol. 3320, pp. 180–183, Springer, Heidelberg, Germany, 2005.

[13] H. S. Behera, D. K. Acharya, and S. S. Panda, "Modified linear vector quantization technique for classification of heart disease data," *International Journal of Advanced Research in Computer Science*, vol. 3, no. 4, pp. 315–319, 2012.

[14] Z. Chuan, L. Xianliang, H. Mengshu, and Z. Xu, "A LVQ-based neural network anti-spam email approach," *ACM SIGOPS Operating Systems Review*, vol. 39, no. 1, pp. 34–39, 2005.

[15] Z. Lu and T. Peng, "The VoIP intrusion detection through a LVQ-based neural network," in *Proceedings of the IEEE International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 1–6, IEEE, London, UK, November 2009.

[16] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.

[17] Q. Chen and S. Abdelwahed, "Towards realizing self-protecting SCADA systems," in *Proceedings of the 9th Annual Cyber and Information Security Research Conference*, pp. 105–108, ACM, Oak Ridge, TN, USA, April 2014.

[18] B. Ramsey, M. Temple, and B. E. Mullins, "PHY foundation for multi-factor ZigBee node authentication," in *Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 795–800, IEEE, Anaheim, CA, USA, December 2012.

[19] M. Majzoobi, F. Koushanfar, and M. Potkonjak, "Techniques for design and implementation of secure reconfigurable PUFs," *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, vol. 2, no. 1, pp. 1–33, 2009.

[20] B. Danev, D. Zanetti, and S. Capkun, "On physical-layer identification of wireless devices," *ACM Computing Surveys (CSUR)*, vol. 45, no. 1, pp. 1–29, 2012.

[21] W. Suski II, M. Temple, M. Mendenhall, and R. Mills, "Using spectral fingerprints to improve wireless network security," in *Proceedings of the 2008 IEEE Global Telecommunications Conference*, pp. 1–5, IEEE, New Orleans, LO, USA, December 2008.

[22] A. King and P. Wahjudi, "Dynamic free text keystroke biometrics system for simultaneous authentication and adaptation to user's typing pattern," *Journal of Management & Engineering Integration*, vol. 6, no. 2, pp. 86–93, 2013.

[23] J. Sammons, *The Basics of Digital Forensics: The Primer for Getting Started in Digital Forensics*, Elsevier, Amsterdam, The Netherlands, 2012.

[24] T. J. Bihl, K. W. Bauer, and M. A. Temple, "Feature selection for RF fingerprinting with multiple discriminant analysis and using zigbee device emissions," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1862–1874, 2016.

[25] J. Hall, M. Barbeau, and E. Kranakis, "Detection of transient in radio frequency fingerprinting using signal phase," *Wireless and Optical Communications*, pp. 13–18, 2003.

[26] M. Lukacs, M. Temple, and P. Collins, "Classification performance using "RF-DNA" fingerprinting of ultra-wideband noise waveforms," *Electronics Letters*, vol. 51, no. 10, pp. 787–789, 2015.

[27] A. K. Jain, P. W. Duin, and J. Jianchang Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[28] T. Bihl and D. Steeneck, "Multivariate stochastic approximation to tune neural network hyperparameters for criticial infrastructure communication device identification," in *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*, pp. 2225–2234, Waikoloa Village, HI, USA, January 2018.

[29] T. Bihl, M. Temple, and K. Bauer, "An optimization framework for generalized relevance learning vector quantization with application to *Z*-wave device fingerprinting," in *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS)*, pp. 2379–2387, Waikoloa Village, HI, USA, January 2017.

[30] C. Dubendorfer, B. Ramsey, and M. Temple, "ZigBee device verification for securing industrial control and building automation systems," in *Critical Infrastructure Protection VII*, vol. 417, pp. 47–62, Springer, Berlin, Germany, 2013.

[31] C. G. Looney, *Pattern Recognition Using Neural Networks*, Oxford University Press, Oxford, UK, 1997.

[32] T. Bihl, W. Young II, and G. Weckman, "Artificial neural networks and their applications in business," in *Encyclopedia of Information Science and Technology*, IGI Global, Hershey, PA, USA, 4th edition, 2018.

[33] J. Huysmans, B. Baesens, J. Vanthienen, and T. Van Gestel, "Failure prediction with self organizing maps," *Expert Systems with Applications*, vol. 30, no. 3, pp. 479–487, 2006.

[34] P. D. Wasserman, "Appendix C: training algorithms," in *Neural Computing: Theory and Practice*, P. D. Wasserman, Ed., Van Nostrand Reinhold, New York, NY, USA, 1989.

[35] P. D. Wasserman, "Counterpropagation networks," in *Neural Computing: Theory and Practice*, P. D. Wasserman, Ed., Van Nostrand Reinhold, New York, NY, USA, 1989.

[36] P. D. Wasserman, *Neural Computing, Theory and Practics*, Van Nostrand Reinhold, New York, NY, USA, 1989.

[37] R. Suurmond and E. Bergkvist, *Artificial Neural Networks and Statistical Approaches to Classifying Remotely Sensed Data*, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1996.

[38] J. Liu, B. Zuo, X. Zeng, P. Vroman, and B. Rabenasolo, "Nonwoven uniformity identification using wavelet texture analysis and LVQ neural network," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2241–2246, 2010.

[39] B. Hammer and T. Villmann, "Generalized relevance learning vector quantization," *Neural Networks*, vol. 15, no. 8-9, pp. 1059–1068, 2002.

[40] M. Kaden, M. Lange, D. Nebel, M. Riedel, T. Geweniger, and T. Villmann, "Aspects in classification learning—review of recent developments in learning vector quantization," *Foundations of Computing and Decision Sciences*, vol. 39, no. 2, pp. 79–105, 2014.

[41] D. Nova and P. A. Estévez, "A review of learning vector quantization classifiers," *Neural Computing and Applications*, vol. 25, no. 3-4, pp. 511–524, 2013.

[42] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[43] R.-P. Li, M. Mukaidono, and I. B. Turksen, "A fuzzy neural network for pattern classification and feature selection," *Fuzzy Sets and Systems*, vol. 130, no. 1, pp. 101–108, 2002.

[44] A. S. Sato and K. Yamada, "Generalized learning vector quantization," in *Advances in Neural Information Processing Systems*, G. Tesauro, D. Touretzky, and T. LEen, Eds., MIT Press, Cambridge, MA, USA, 1995.

[45] A. I. Gonzalez, M. Grana, and A. D'Anjou, "An analysis of the GLVQ algorithm," *IEEE Transactions on Neural Networks*, vol. 6, no. 4, pp. 1012–1016, 1995.

[46] T. Bojer, B. Hammer, D. Schunk, and K. Tluk von Toschanowitz, "Relevance determination in learning vector quantization," in *Proceedings of European Symposium on Artificial Neural Networks (ESANN)*, pp. 271–276, Bruges, Belgium, April 2001.

[47] M. J. Mendenhall and E. Merenyi, "Relevance-based feature extraction for hyperspectral images," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 658–672, 2008.

[48] D. DeSieno, "Adding a conscience to competitive learning," in *Proceedings of the 1998 IEEE International Conference on Neural Networks*, pp. 117–124, IEEE, San Diego, CA, USA, July 1988.

[49] M. Strickert, U. Seiffert, N. Sreenivasulu, W. Weschke, T. Villmann, and B. Hammer, "Generalized relevance LVQ (GRLVQ) with correlation measures for gene expression analysis," *Neurocomputing*, vol. 69, no. 7-9, pp. 651–659, 2006.

[50] M. Biehl, B. Hammer, and T. Villmann, "Distance measures for prototype based classification," in *Proceedings of the International Workshop on Brain-Inspired Computing*, Cetraro, Italy, July 2013.

[51] S. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 4, no. 1, pp. 300–307, 2007.

[52] J. Fuller and B. Ramsey, "Rogue *Z*-wave controllers: a persistent attack channel," in *Proceedings of the 2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*, pp. 734–741, IEEE, Clearwater Beach, FL, USA, October 2015.

[53] A. Zobaa and T. Bihl, "Security methods for critical infrastructure communications," in *Big Data Analytics in Future Power Systems*, CRC Press, Boca Raton, FL, USA, 2018.

[54] T. J. Bihl, M. A. Temple, K. Bauer, and B. Ramsey, "Dimensional reduction analysis for physical layer device fingerprints with application to ZigBee and *Z*-wave devices," in *Proceedings of the 2015 IEEE Military Communications Conference (MILCOM)*, pp. 360–365, IEEE, Tampa, FL, USA, October 2015.

[55] I. Yaqoob, I. A. T. Hashem, Y. Mehmood, A. Gani, S. Mokhtar, and S. Guizani, "Enabling communication technologies for smart cities," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 112–120, 2017.

[56] M. Knight, "How safe is *Z*-wave? (Wireless standards)," *Computing and Control Engineering*, vol. 17, no. 6, pp. 18–23, 2006.

[57] S. Omatseye, "Z battle for Z wirelessly controlled home," *RCR News*, vol. 24, no. 5, p. 7, 2005.

[58] C. Gomez and J. Paradells, "Wireless home automation networks: a survey of architectures and technologies," *IEEE Communications Magazine*, vol. 48, no. 6, pp. 92–101, 2010.

[59] C. W. Badenhop, B. W. Ramsey, B. E. Mullins, and L. O. Mailloux, "Extraction and analysis of non-volatile memory of the ZW0301 module, a *Z*-wave transceiver," *Digital Investigation*, vol. 17, pp. 14–27, 2016.

[60] M. Galeev, *Catching the Z-Wave*, pp. 1–5, Electronic Engineering Times India, 2006.

[61] ITU, *ITU-T G.9959: Short Range Narrow-Band Digital Radio Communication Transceiver—PHY and MAC Layer Specifications*, International Telecommunication Union, Geneva, Switzerland, 2012.

[62] H. Patel and B. Ramsey, "Comparison of parametric and nonparametric statistical features for *Z*-wave fingerprinting," in *Proceedings of the 2015 IEEE Military Communications Conference (MILCOM)*, pp. 378–382, IEEE, Tampa, Florida, USA, October 2015.