Optimal Window-Symbolic Time Series Analysis for Pattern Classification and Anomaly Detection

Ibrahim F. Ghalyan , *Senior Member, IEEE*, Najah F. Ghalyan , and Asok Ray , *Life Fellow, IEEE*

Abstract—This article proposes the optimal window-symbolic time series analysis (OW-STSA) methodology to optimize parameters of feature extraction and pattern classification in industrial processes. The underlying theory is built upon minimization of an empirical risk function to discriminate between nominal and anomalous operations of the physical process under consideration. In particular, the proposed methodology produces: optimized windows of the time series used for pattern classification and anomaly detection, and optimized identification of feature extractors and classifier parameters. The algorithm is realized by segmenting a given time series into windows of equal size. Then, the stationary state probability vector is computed for each window in the sense of OW-STSA for anomaly prediction with locally optimal accuracy of detection performance. The proposed methodology has been experimentally validated in laboratory environment with different classifiers for two distinct industrial processes. The first experiment addresses detection of fatigue failure in polycrystalline alloy structures using time series of ultrasonic signals. The second experiment investigates detection of thermoacoustic instability in an emulated combustion system using time series of pressure-wave signals. In both experiments, the proposed OW-STSA methodology yielded excellent detection performance of anomalous behavior with multiple classification techniques.

Index Terms—Anomaly detection, empirical risk function (ERF), symbolic dynamics, time series signals.

I. INTRODUCTION

DETECTION of anomalies by time series analysis has attracted tangible attention in multiple disciplines, such as engineering [1], medicine [2], and finance [3]. Such a wide spectrum of application domains brought in a significant impact on the progress of the science and art of anomaly detection by time series signals as described below.

Wang *et al.* [4] proposed an anomaly detection approach for monitoring nonlinear and multimode industrial processes. Their approach used a Dirichlet process Gaussian mixed model to classify the modes of multimode processes from the monitored data. Benkabou *et al.* [5] used a combination of entropy and dynamic time warping to identify anomalies in signals. Guo *et al.* [6] leveraged the gated recurrent unit to discriminate between nominal and anomalous behaviors of data sequences. Zhang *et al.* [1] used temporal dependence and cross-variable association for anomaly detection in power plants, using sensed time series signals. Convolutional neural network (CNN) was used by Piekarski *et al.* [7] to detect anomalies of sensed signals, such as instability in synchrotron radiation. Yin *et al.* [8] used a fixed sliding window concatenated with hybrid CNN and recurrent autoencoder to classify nominal and anomalous regions of data samples. Long short-term memory based techniques were used to identify anomalies in time series [9]–[11]. Various other techniques were proposed to detect anomalies in time series signals, such as classical recurrent neural networks, generative adversarial networks, among other techniques (e.g., Siegel *et al.* [12] and the references therein). Anomalous behavior in data of medical Internet of Things networks was detected by using a rough set theory based fuzzy core vector machine technique [13]. Granular computing was used for detection and scoring anomalous samples of a dataset, where fuzzy semantics produced a degree of anomaly of data samples [14]. Despite their excellent performance, most of the aforementioned techniques and the references therein rely on magnitudes of signals for detection of anomalous data samples.

Symbolic time series analysis (STSA) [15], [16] addresses anomaly detection, based on both magnitude and texture (i.e., dynamical behavior) of sensed signals, where a time series is converted into a string of symbols. Then, the state-transition probability [system state probability (SSP)] vector is computed from windows of symbol strings. The SSP is the feature of each window, which is used to determine if the corresponding part of the time series is nominal or anomalous [17]. Nevertheless, in [17], the window length was selected to be long enough to compute a realistic SSP vector, which enables reasonable classification of nominal and anomalous regions. Indeed, selection of an insufficient window length may lead to lack of stationarity in the SSP, which degrades the detection performance. A measure preserving transformation (MPT) based STSA technique

Ibrahim F. Ghalyan is with the Bank of New York Mellon Corporation, New York, NY 10286 USA (e-mail: ibrahim.ghalyan@bnymellon.com).

Najah F. Ghalyan is with the Department of Mechanical Engineering, The Pennsylvania State University, University Park, PA 16802 USA, and also with the Department of Mechanical Engineering, University of Kerbala, Kerbala 56001, Iraq (e-mail: nfjasim76@gmail.com).

Asok Ray is with the Department of Mechanical Engineering and the Department of Mathematics, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: axr2@psu.edu).

addressed selection of window length for low-delay anomaly detection [18], [19]; however, the satisfaction of required properties, such as measure preservation, may complicate its implementation. Furthermore, there is no direct relation to the recognition performance in the MPT-based STSA, potentially leading to nonoptimal window length and detection accuracy.

This article proposes an optimal window-STSA (OW-STSA) methodology to jointly optimize the window length and the empirical risk function (ERF) in the classification stage. The joint optimization is realized by plugging the window length into the ERF of the classification of nominal and anomalous parts; then, the window length is estimated such that the ERF is minimized. The OW-STSA technique is experimentally validated on two laboratory-scale apparatuses: fatigue damage detection in polycrystalline alloy structures and detection of thermoacoustic instability (TAI) in an electrically heated Rijke tube [20], which emulates an industrial-scale combustion system.

The main contributions of this article are as follows.

1) *Development of a low time-latency methodology for anomaly detection:* The proposed OW-STSA methodology simultaneously optimizes the window length of symbol strings and the ERF for accuracy of anomaly classification.

2) *Experimental validation of the methodology:* The proposed OW-STSA methodology has been successfully validated, using multiple classification techniques, in the laboratory environment for two industrial applications, namely, fatigue crack detection in mechanical structures and monitoring of TAI in combustion systems.

The rest of this article is organized as follows. Section II explains the notion of STSA. Section III details the proposed OW-STSA methodology. Section IV presents the results of experimental validation. Finally, Section V concludes this article.

II. SYMBOLIC TIME SERIES ANALYSIS

Let $X = \{x(1), \ldots, x(T)\}$ to be a time series of the signal having length T, where $x(t)$'s are (possibly) noisy sensor measurements generated by a dynamical system. The main objective of STSA is to detect and identify the anomalous behavior, if any, in the time series X. A schematic description of STSA is depicted in Fig. 1, where the decision output $\hat{y}$ is a binary signal that identifies if a part of the time series represents the nominal (indicated by blue solid lines) or anomalous (indicated by dotted red lines) behavior. In STSA, a time series X is converted into a string of symbols $S = \{s_1, \ldots, s_T\}$ with $s_i \in \mathcal{A}$, where $\mathcal{A}$ is the alphabet of *finitely many* (i.e., $|\mathcal{A}|$) symbols. Such a conversion is performed by partitioning the state space of the dynamical system into $|\mathcal{A}|$ disjoint regions. Each region is assigned with a distinct symbol from $\mathcal{A}$. Several partitioning techniques can be used; examples are uniform partitioning (UP), maximum entropy partitioning (MEP), and K-means. In a UP process, the state space is partitioned into regions with equal size, whereas in MEP process, the time series is partitioned into segments with equal probability such that the total entropy is maximized. In
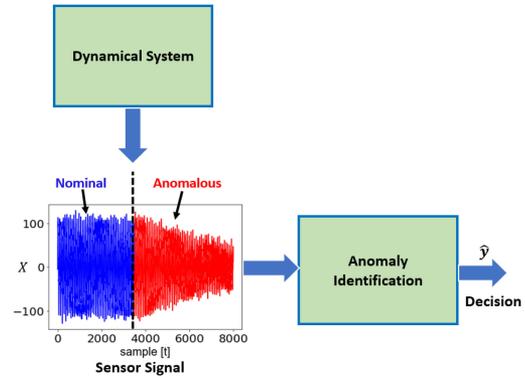

Fig. 1. Schematic diagram for detection of nominal (solid blue lines) and/or anomalous (dotted red lines) behavior of the dynamical system as indicated by the time series of signals.

K-means, the state space is partitioned by using the K-means clustering process.

For accurate discrimination between nominal and anomalous behaviors of the dynamical system, the time series is segmented into windows of measurements $\chi_t = \{x((t-1)W_\mathrm{d}), \ldots, x(tW_\mathrm{d} - 1)\}$, $t = 1, 2, \ldots, J$, where J is the total number of windows of length W_d, generated from the given time series. It is possible that each window χ_t of X can be converted into strings of symbols by different partitioning methods; however, in this article, the same partitioning method has been used for all windows. Now, one may use the notion of STSA to extract features from these windows of measurements to draw conclusions (i.e., nominal or anomalous) about the behavior of the dynamical system.

To model the behavior of the generated string of symbols, D-Markov machines have been employed to capture the sequential behavior of symbol strings [15], [16]. The notion of D-Markov machines uses an algebraic structure, called the probabilistic finite-state automaton (PFSA), defined by $\mathcal{K} = (\mathcal{A}, \mathcal{Q}, \delta, \mathcal{M})$, where $\mathcal{A}$ is a finite-cardinality alphabet, $\mathcal{Q}$ is a finite set of states, $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ is a state transition map, and $\mathcal{M} : \mathcal{Q} \times \mathcal{A} \to [0, 1]$ generates individual entries of the emission (also called morph) matrix. The maps δ and $\mathcal{M}$ determine the $(|\mathcal{Q}| \times |\mathcal{Q}|)$ state transition probability matrix (STPM), which is used to generate the SSP vector π as the (sum-normalized) left eigenvector corresponding to the unique eigenvalue 1, provided that the constructed STPM is ergodic [18], [19]. Indeed, a D-Markov machine is a PFSA for which the probability of the emitted symbols depends only on the previous D consecutive symbols [15], [16]. In this way, a time series X of sensor data is converted to low-dimensional feature vectors that can efficiently detect anomalies in the dynamical behavior of time series signals. Algorithm 1 summarizes the STSA procedure and further details are reported in [16] and references therein.

Step 1 of Algorithm 1 is the partitioning process, where a given time series is segmented into logical partitions by using one of the aforementioned partitioning techniques [21]. Then, each sample of the time series signal is converted into a symbol assigned to its partition and as given in Step 1 of Algorithm 1.
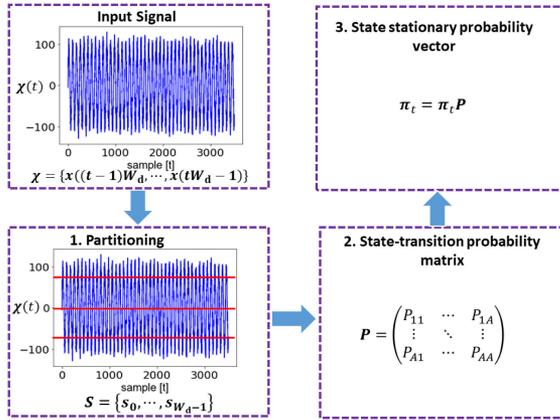
Fig. 2.   Schematic diagram for implementation of STSA in Algorithm 1.

Algorithm 1: STSA Feature Extraction.

    INPUT: An alphabet $\mathcal{A}$, a Markov depth D, and a data block $\{x((t-1)W_{\mathrm{d}}), \ldots, x(tW_{\mathrm{d}} - 1)\}$.
    OUTPUT: A feature vector π_t.
1:   Convert $\{x((t-1)W_{\mathrm{d}}), \ldots, x(tW_{\mathrm{d}} - 1)\}$ into a symbol string $\{s_0, \ldots, s_{W_{\mathrm{d}}-1}\}$, $s_i \in \mathcal{A}$, using one of the STSA partitioning methods.
2:   Using frequency counting, construct a D-Markov machine based on $\{s_0, \ldots, s_{W_{\mathrm{d}}-1}\}$ to obtain the state-transition probability matrix $\mathcal{P}(t)$.
3:   Find the feature vector as the state stationary probability (SSP) vector π_t computed as the left eigenvector of $\mathcal{P}(t)$ corresponding to the eigenvalue $\lambda = 1$, i.e., $\pi_t = \pi_t \mathcal{P}(t)$.

Step 2 of Algorithm 1 computes probabilities of transition between symbols to compute the STPM $\mathcal{P}$. Step 3 of Algorithm 1 computes the SSP π as the left eigenvector corresponding to the unity eigenvalue. The aforementioned explanation is reinforced with a schematic diagram given by Fig. 2 that helps visualizing the STSA algorithm.

A judicious choice of the window length W_d is essential for early prediction of anomalies with low time-delay and accurate performance. The next section presents an empirical approach to optimize the selection of W_d to achieve low-delay prediction of anomalies with good performance.

III. OPTIMAL WINDOW-STSA

This section develops the OW-STSA algorithm. Given a (time series) window segment χ of length W_{d}, let $L(y, f(\chi, W_{\mathrm{d}}))$ be a classification loss function described by the indicator function

$$L(y, f(\chi, W_{\mathrm{d}})) = \begin{cases} 0, & \text{for } y = f(\chi, W_{\mathrm{d}}) \\ 1, & \text{for } y \neq f(\chi, W_{\mathrm{d}}) \end{cases} \quad (1)$$

where y is the (discrete-valued) class label of χ, and $f(\chi, W_{\mathrm{d}})$ is the estimated class label, i.e.,
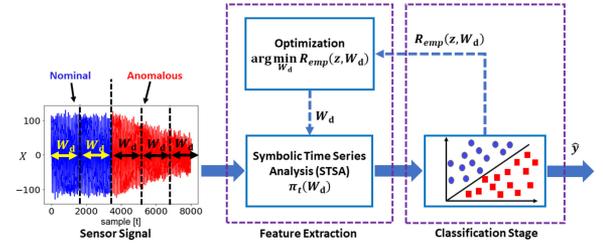
$$\hat{y} = f(\chi, W_{\mathrm{d}}). \quad (2)$$

Fig. 3.   Schematic diagram of the proposed OW-STSA for detection of nominal (solid blue lines) and/or anomalous (dotted red lines) behavior of the dynamical system, as indicated by the time series of signals.

The risk functional $R(z, W_{\mathrm{d}})$ is written as

$$R(z, W_{\mathrm{d}}) = \int L(y, f(\chi, W_{\mathrm{d}})) dP(z) \quad (3)$$

where $z \triangleq (\chi, y)$ and $P(z)$ is the joint probability distribution. Then, given a dataset $z_N = \{(\chi_1, y_1), \ldots, (\chi_N, y_N)\}$, (3) is approximated by the ERF

$$R_{\mathrm{emp}}(z_N, W_{\mathrm{d}}) = \frac{1}{N} \sum_{n=1}^{N} L(y_n, f(\chi_n, W_{\mathrm{d}})). \quad (4)$$

An optimal value W_{d}^* of the window length is obtained by minimizing $R_{\mathrm{emp}}(z, W_{\mathrm{d}})$ in (4) as

$$W_{\mathrm{d}}^* = \arg\min_{W_{\mathrm{d}}} R_{\mathrm{emp}}(z, W_{\mathrm{d}}) \quad (5)$$

for which it is necessary that the derivative of $R_{\mathrm{emp}}(z, W_{\mathrm{d}})$ with respect to W_{d} is zero at the optimal point W_{d}^*, i.e.,

$$\left. \frac{\partial R_{\mathrm{emp}}(z, W_{\mathrm{d}})}{\partial W_{\mathrm{d}}} \right|_{W_{\mathrm{d}} = W_{\mathrm{d}}^*} = 0 \quad (6)$$

which leads to the fact that

$$\lim_{h \to 0} \frac{R_{\mathrm{emp}}(z, W_{\mathrm{d}} + h) - R_{\mathrm{emp}}(z, W_{\mathrm{d}})}{h} = 0. \quad (7)$$

The limit in (7) implies that for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $0 < h < \delta$

$$\left| \frac{R_{\mathrm{emp}}(z, W_{\mathrm{d}} + h) - R_{\mathrm{emp}}(z, W_{\mathrm{d}})}{h} \right| < \varepsilon. \quad (8)$$

It follows from (8) that

$$|R_{\mathrm{emp}}(z, W_{\mathrm{d}} + h) - R_{\mathrm{emp}}(z, W_{\mathrm{d}})| < h\varepsilon \triangleq \acute{\varepsilon}. \quad (9)$$

Arbitrarily small choices of h and ε (and hence the choice of $\acute{\varepsilon}$) enable (9) to be a stopping condition to estimate W_{d}^*; the sufficiency for minimization is established by checking that

$$\left. \frac{\partial^2 R_{\mathrm{emp}}(z, W_{\mathrm{d}})}{\partial W_{\mathrm{d}}^2} \right|_{W_{\mathrm{d}} = W_{\mathrm{d}}^*} > 0. $$

The window-length optimality of OW-STSA is realized from the stopping condition of (9), as detailed in Algorithm 2.

The realization of Algorithm 2 simultaneously yields minimal window length W_{d}^* and local optimal accuracy of anomaly detection. The schematic diagram in Fig. 3 helps visualization of how OW-STSA is applied in Algorithm 2.

Algorithm 2: Optimal Window-Symbolic Time Series Analysis (OW-STSA).

INPUT: Nominal X_h and anomalous X_f components of a time series signal, step size ΔW, tolerance $\acute{\varepsilon}$, initial value of window length W_d and its lower and upper bounds $\{\underline{W}, \overline{W}\}$

OUTPUT: Optimal parameters – window length W_d^*, symbolic state probability (SSP) vector P_t^*, and classifier C^*

1:   Initialize: $W_d \leftarrow \underline{W}$
2:   Segment X_h and X_f into windows of size W_d
3:   Compute the corresponding SSP vector P_t of each window using Algorithm 1.
4:   Compose the training input X_{tr} to be matrix of vectors of SSP of X_h and X_f
5:   Compose the corresponding output targets Z_{tr} of X_{tr}
6:   Develop the classifier C given the data (X_{tr}, Z_{tr})
7:   Compute the empirical risk R_{emp}^* of C
8:   **repeat**
9:     $W_d \leftarrow W_d + \Delta W$
10:   Segment X_h and X_f into windows of size W_d
11:   Compute the SSP vector π_t of each window using Algorithm.1.
12:   Compose the training input X_{tr} to consist of SSP vectors X_h and X_f
13:   Compose the corresponding output targets Z_{tr} of X_{tr}
14:   Develop the classifier C given the data (X_{tr}, Z_{tr})
15:   Compute the empirical risk function R_{emp} of C
16:   **if** $R_{emp} < R_{emp}^*$ **then**
17:     $R_o \leftarrow R_{emp}^*, R_{emp}^* \leftarrow R_{emp}$
18:   **end if**
19:   **until** $|R_{emp}^* - R_o| < \varepsilon$ or $W_d \geq \overline{W}$
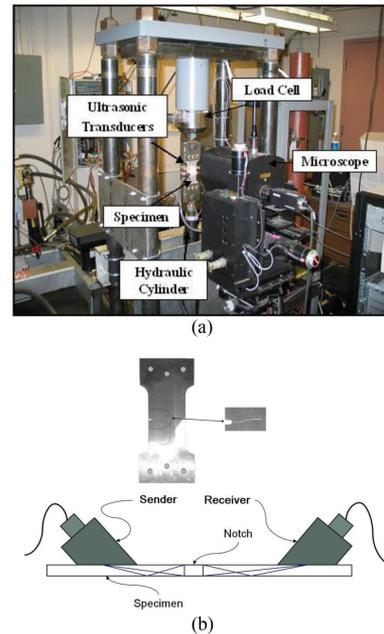20: $W_d^* \leftarrow W_d, \pi_t^* \leftarrow \pi_t$, and $C^* \leftarrow C$

Fig. 4. Test apparatus of Experiment #1. (a) Fatigue testing machine. (b) Details of a (aluminum alloy 7075-T6) test specimen with the ultrasonic sensor.
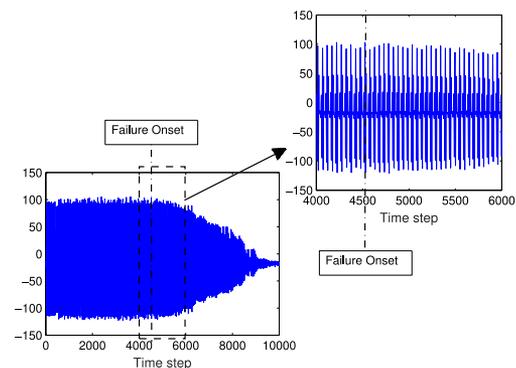

Fig. 5. Ultrasonic signal for a specimen with fatigue onset at approximately cycle number 4525.

IV. VALIDATION OF THE OW-STSA METHODOLOGY

This section validates the OW-STSA methodology with two laboratory experiments: fatigue damage in polycrystalline alloy structures and TAI in an emulated combustion system. Time series signals were acquired for both experiments, which have healthy (or stable) and damaged (or unstable) phases. The classification model uses ten-fold cross-validation [22]. In Experiments #1 and #2, the number of states was selected to be 8, i.e., the dimension of the STPM $\mathcal{P}$ is 8×8.

A. Experiment #1: Fatigue Damage Detection

The test apparatus for fatigue damage detection is depicted in Fig. 4(a). The apparatus is equipped with an ultrasonic sensor that has a sender and a receiver of ultrasonic signal waves to monitor the health of tested materials. Fig. 4(b) depicts the mechanism of capturing the ultrasonic signal of a specimen made of polycrystalline aluminum alloy AL7075-T6. The specimen has dimensions of 3-mm thickness and 50-mm width with a side-notch of 1.58×4.57 mm^2, to increase the stress concentration factor and thereby ensure local crack initiation at the

notch end. To gradually initiate and propagate fatigue damage, a cyclic sinusoidal load under tension–tension stress is applied at the specimen, with a frequency of 12.5 Hz using the MTS 831.1 fatigue testing machine in Fig. 4(a). When the specimen is healthy, the receiver captures the nominal signal generated by the sender. Once fatigue damage is initiated, the behavior of the ultrasonic signal, captured by the receiver, drifts toward the anomalous mode of operation. Twenty four specimens were tested and Fig. 5 shows the ultrasonic time series signal that is generated on a typical specimen. As the fatigue damage accumulates, the observed behavior of the time series changes, for which OW-STSA is used to distinguish between the nominal and anomalous behaviors of the signal time series, thereby providing a proxy to detect fatigue damage. In this article, each time series of signals under consideration has been downsampled to $\sim$ 10000 data samples, as shown in Fig. 5. The stochastic nature of fatigue damage in polycrystalline alloys randomizes the onset
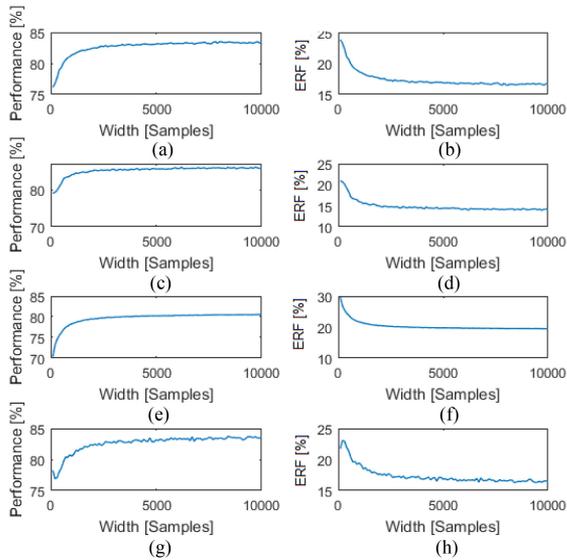
Fig. 6. Classification performance in Experiment #1. (a) Performance of AdapBoost. (b) ERF of AdapBoost. (c) Performance of BagEns. (d) ERF of BagEns. (e) Performance of NB. (f) ERF of NB. (g) Performance of DT. (h) ERF of DT.

Fig. 7. Classification performance in Experiment #1. (a) Performance of k-NN. (b) ERF of k-NN. (c) Performance of LDA. (d) ERF of LDA. (e) Performance of LR. (f) ERF of LR. (g) Performance of SVM. (h) ERF of SVM.

TABLE I
EXPERIMENT 1: OPTIMAL WINDOW LENGTH AND ACCURACY OF OW-STSA AND CONVENTIONAL STSA AT $W_d = \{1000, 6000\}$ WITH MULTIPLE CLASSIFICATION TECHNIQUES

Classification	W_d^*	OW-STSA $[W_d^*]$	STSA [1000]	STSA [6000]
AdaBoost	3100	83.12%	81.43%	83.57%
BagEnsemble	3000	85.56%	83.61%	85.83%
NB	2900	80.00%	78.15%	80.21%
DT	2800	82.81%	80.82%	82.95%
k-NN	2900	83.21%	81.04%	83.97%
LDA	2900	80.51%	78.74%	80.59%
LR	2700	83.75%	80.05%	84.03%
SVM	3300	78.58%	77.14%	78.70%

Fig. 8. Test apparatus of Experiment #2: Electrically heated Rijke tube.

points of failure in individual specimens, even under identical loading conditions. Furthermore, the change in the signal pattern at the failure onset is relatively small, as shown in Fig. 5, which makes identification of the onset points a challenging task [18].

To detect and identify the anomalous behavior in polycrystalline alloys due to accumulation of fatigue damage, the proposed OW-STSA methodology was implemented with multiple classification techniques, namely, adaptive boosting (AdapBoost), bagged ensemble (BagEns), naive Bayes (NB), decision tree (DT), k-nearest neighbor (k-NN), linear discriminant analysis (LDA), logistic regression (LR), and support vector machine (SVM) (see machine learning and statistical modeling literature, such as the works in [22] and [23], for further details about the aforementioned classification techniques). Figs. 6 and 7 show the performance and ERF of the aforementioned eight classification techniques, with respective optimal widow lengths, obtained using OW-STSA. In each case, the convergence parameter $\hat{\varepsilon}$ [see (9)] was selected to be 0.01. Using the OW-STSA, pairs of optimal window length (in # of samples) and accuracy [in percent (%)] (W_d^*, Acc) are listed in Table I. For comparison of OW-STSA with conventional STSA, where window length is manually selected, both relatively small (e.g.,
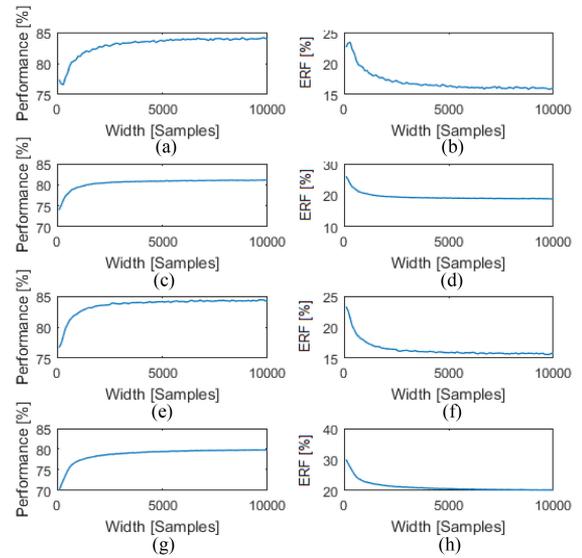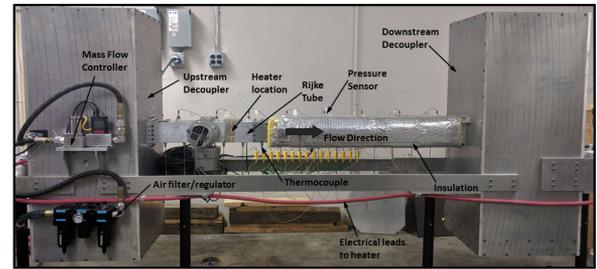
$W_d = 1000$ samples) and large (e.g., $W_d = 6000$ samples) were investigated; these results are also listed in Table I.

It is worth mentioning that image analysis techniques can be efficiently used to identify abrupt behavioral changes in many industrial systems (e.g., thermal imaging to detect faults in machines [24]). However, for fatigue damage detection, online capturing of microscope images of the materials under consideration are needed, which requires more sophisticated, expensive, and fragile devices to capture such kind of images that limit its practical applications [25]. Furthermore, image-based techniques can detect fatigue damage on specific spots of the material under the coverage of microscopes, which adds more constraints in their usage to detect fatigue damage. Thus, behavioral changes of sensed ultrasonic signals can provide a more efficient and realistic means to detect fatigue damage.

B. Experiment #2: Detection of TAI

This experiment serves to detect TAI, by using sensed time series signals from a laboratory-scale electrically heated Rijke tube apparatus, as depicted in Fig. 8, which emulates commercial-scale combustion systems [20]. The apparatus consists of a 1.5-m-long horizontal tube with an external cross-section of
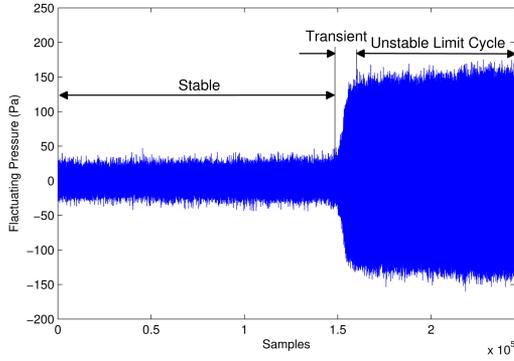
Fig. 9. Unsteady pressure signals showing the transience from stable (nominal) to unstable limit cycle (anomalous) behavior.

10×10 cm and a wall thickness of 6 mm. The apparatus has a controller that regulates the flow of air (Q) at atmospheric pressure through the tube, a heating element, and a programmable dc power supply that controls the power input (E_{in}) to the heater. Multiple experiments were conducted with flow rates ranging from 130 to 250 L/min with increment of 20 L/min. First, the system is heated to a steady state, using the primary heater power input E_{in} with a power of approximately 200 W. Then, the power input is abruptly increased to a higher value to have a limit-cycle behavior (see Mondal *et al.* [20] for further details). For multiple air-flow rates and heat inputs, bifurcating transition from stable to unstable behavior results in the acoustic response of the chamber. Pressure sensors are installed in the tube to check TAI of the underlying system.

Fifteen experiments were conducted on the Rijke tube apparatus. In each experiment, the process started with nominal (i.e., stable) behavior and gradually drifted to anomalous (i.e., unstable) state, and a time series of pressure oscillations was captured over 30 s, sampled at 8192 Hz, filtered to attenuate the effects of low-frequency environmental acoustics. A typical profile of the pressure time series is shown in Fig. 9, where the process started with a stable behavior, then entered into a transient mode, and finally transitioned into the unstable state.

Similar to what was done in Experiment #1, the proposed OW-STSA methodology was tested with the data collected from Experiment #2. Setting the convergence parameter $\check{\varepsilon} = 0.01$ [see (9)], Figs. 10 and 11 show the performance and ERF of the aforementioned eight classification techniques, with respective optimal window length W_{d}^*, obtained by OW-STSA. Pairs of optimal window length (in # of samples) and accuracy [in percent (%)] (W_{d}^*, Acc) are listed in Table II. For comparison of OW-STSA with conventional STSA, where window length is manually selected, both relatively small (e.g., $W_{\mathrm{d}} = 200$ samples) and relatively large (e.g., $W_{\mathrm{d}} = 3000$ samples) were investigated; these results are also listed in Table II.

C. Discussion of Results of Experiments #1 and #2

Referring to Figs. 6 and 7 with respect to Experiment #1, and from Figs. 10 and 11 with respect to Experiment #2, it is seen that an increase of the window length W_{d} in the construction
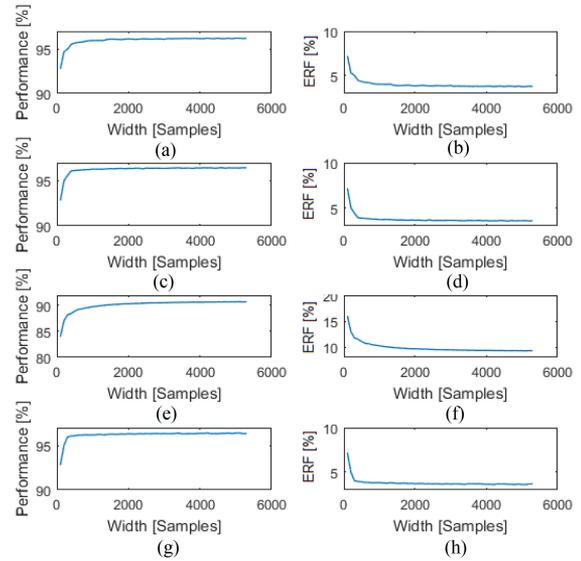

Fig. 10. Classification performance in Experiment #2. (a) Performance of AdapBoost. (b) ERF of AdapBoost. (c) Performance of BagEns. (d) ERF of BagEns. (e) Performance of NB. (f) ERF of NB. (g) Performance of DT. (h) ERF of DT.
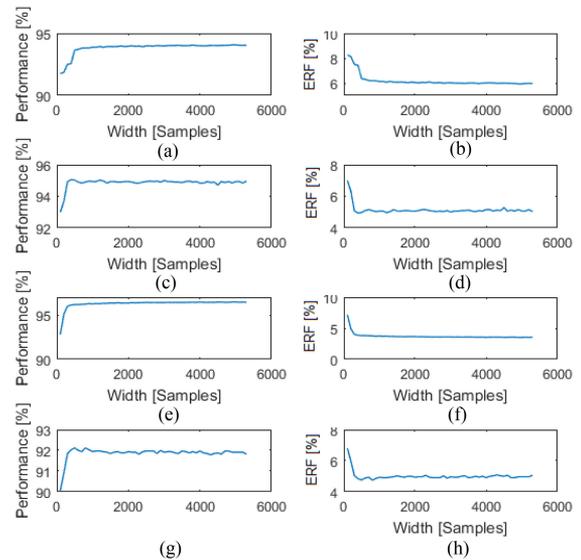

Fig. 11. Classification performance in Experiment #2. (a) Performance of k-NN. (b) ERF of k-NN. (c) Performance of LDA. (d) ERF of LDA. (e) Performance of LR. (f) ERF of LR. (g) Performance of SVM. (h) ERF of SVM.

TABLE II
EXPERIMENT 2: OPTIMAL WINDOW LENGTH AND ACCURACY OF OW-STSA
AND CONVENTIONAL STSA AT $W_{\mathrm{d}} = \{200, 3000\}$ WITH MULTIPLE
CLASSIFICATION TECHNIQUES

Classification	W_{d}^*	OW-STSA [W_{d}^*]	STSA [200]	STSA [3000]
AdaBoost	1100	95.10%	93.98%	95.70%
BagEnsemble	900	96.36%	94.34%	96.42%
NB	2000	90.05%	87.75%	90.71%
DT	700	96.27%	95.14%	96.40%
k-NN	900	93.87%	92.17%	93.96%
LDA	1100	94.90%	93.91%	94.98%
LR	600	96.37%	94.87%	96.50%
SVM	800	92.10%	90.96%	92.19%

of STSA leads to enhanced accuracy with reduced ERF. As $W_{\rm d}$ increases, a higher discrimination capability of the D-Markov machine is obtained due to the following.

1) Richer information to capture the dynamical behavior.
2) Reduced dominance of the effect of transient on the onset of anomalies.
3) Better convergence of the probability transition matrix to a constant matrix.

Experiments #1 and #2 validated identification of optimal window length by using OW-STSA, as summarized in Tables I and II. Nevertheless, the improvement in accuracy beyond the estimated optimal window lengths $W_{\rm d}^*$ of OW-STSA is not significant but it considerably increases the delay in detection of anomalies. Indeed, Tables I and II show that by comparison of OW-STSA results with those of conventional STSA, very slight increase in the accuracy is obtained for all classification techniques with a tangible delay of detection. On the other hand, selection of a relatively small window length significantly degrades the classification accuracy, which compromises the reliability of anomaly detection. Tables I and II also verify this fact when comparing the conventional STSA with relatively small manual selection of window length, where tangible performance degradation is caused with all classification techniques. As a result, the use of conventional STSA with nonoptimal window length may either lead to delayed detection of anomalies or compromised accuracy of detection. Therefore, the use of optimal window length (for a specific classification method) in the proposed OW-STSA methodology is reasonable, because it provides early prediction of anomalies with locally optimal accuracy of detection.

Table I shows that $W_{\rm d}^*$ depends on the classification technique used in the detection process. Thus, $W_{\rm d}$ might vary from one classification technique to another. Such a variation is caused by the difference in reaching the steady-state performance of a classification technique as $W_{\rm d}$ increases. Therefore, usage of OW-STSA to estimate optimal window length is influenced by the selection of a classification technique, which justifies the notion of the ERF in the optimization process. The minimal window length obtained in Experiment #1 was attained with the LR technique, where the optimal window length was computed to be 2700 with corresponding accuracy of 83.75%. However, the highest accuracy was found to be 85.56% when using the bagged ensemble classification technique with optimal window length of 3000. Therefore, for Experiment #1, bagged ensemble apparently produces the most reliable detection, yet the earliest detection of anomalies can be attained with LR.

Similarly for Experiment #2, Table II demonstrates that both optimal window length and classification accuracy might change for one classification technique to another, emphasizing the impact of the classification stage to the optimal window length and accuracy of detection. Here also, the minimal optimal window length $W_{\rm d}^*$ is attained with the use of LR, which is 600 samples with the corresponding accuracy of 96.37%. Indeed, the highest accuracy is obtained with the use of LR, which achieves both early prediction of anomalies and highest accuracy compared to other classification techniques. It is obvious from Tables I and II that values of $W_{\rm d}^*$ in Experiment #2 are less than those in Experiment #1. Likewise for detection accuracy, values obtained in Experiment #2 are more than those in Experiment #1. Such minimal window lengths and enhanced values of accuracy in Experiment #2 give a clear evidence of the challenge in discriminating between the measurements of healthy and anomalous phases of Experiment #1. Indeed, the nature of the datasets itself plays a key role to determine both optimal window length and accuracy of detection.

Optimal window length was shown to be different from one experiment to another, even if the same classification technique is used. Indeed, each process has its own dynamical behavior, which generates its own time series signal, rendering the nature of optimal description of features, such as window length, to vary from an application to another. However, the algorithm remains the same and it is apparently capable of estimating a local value of $W_{\rm d}^*$ for optimal detection of anomalous behavior. Therefore, it is conjectured that OW-STSA would be applicable for similar problems, provided that a sufficient number of time series samples are available, and the system generating the signal exhibits variations in its dynamical behavior after an anomalous behavior is encountered.

V. CONCLUSION

This article proposed the OW-STSA technique for detection of anomalous behavior in dynamical systems, using sensed time series signals. To detect an anomalous behavior, the time series was first converted into a string of symbols, by one of the available partitioning processes. A specific window of resulting string of symbols was used to compute the STPM, by computing the frequency of transition from a symbol to another over the given window of symbol strings. Then, the state stationary probability vector was determined, which was the (sum-normalized) left eigenvector corresponding to unity eigenvalue of the STPM. The resulting state stationary probability vector was considered as input features for the classification stage to detect if the corresponding window of symbol strings was anomalous. The window length of the symbol string was optimized by minimizing the ERF of the classification process that discriminates between the nominal and anomalous phases of time series signals of the physical process.

The OW-STSA methodology was experimentally validated for two industrial processes, namely, detection of fatigue damage in polycrystalline alloy structures and detection and prediction of TAI in combustion systems. In both experiments, OW-STSA demonstrated not only less time delay to detect anomalous behavior in typical industrial processes, but also enhanced accuracy in the anomaly detection from the respective time series.

Despite the excellent performance of the proposed OW-STSA methodology, future research is recommended on the following topics.

1) *Problem formulation for global optimization:* This research could be realized by selecting convex ERFs, where derivative-free optimization would prevent being trapped into local optima.
2) *Inclusion of window length in the models of STPM:* This research could be realized by rigorously constructing an appropriate risk functional, instead of using the ERF, and then by minimizing the risk functional.

3) *Integration of the OW-STSA and classification stages:* This research will make the optimization problem non-convex, which would require reformulation of the OW-STSA and classification processes as a more challenging multiobjective optimization problem.

REFERENCES

[1] Y. Zhang, Z. Y. Dong, W. Kong, and K. Meng, "A composite anomaly detection system for data-driven power plant condition monitoring," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4390–4402, Jul. 2020.

[2] F. C. Bauer, D. R. Muir, and G. Indiveri, "Real-time ultra-low power ECG anomaly detection using an event-driven neuromorphic processor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1575–1582, Dec. 2019.

[3] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," *Comput. Sci. Rev.*, vol. 40, 2021, Art. no. 100402.

[4] B. Wang, Z. Li, Z. Dai, N. Lawrence, and X. Yan, "Data-driven mode identification and unsupervised fault detection for nonlinear multimode processes," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3651–3661, Jun. 2020.

[5] S.-E. Benkabou, K. Benabdeslem, and B. Canitia, "Unsupervised outlier detection for time series by entropy and dynamic time warping," *Knowl. Inf. Syst.*, vol. 54, pp. 463–486, 2018.

[6] Y. Guo, W. Liao, Q. Wang, L. Yu, T. Ji, and P. Li, "Multidimensional time series anomaly detection: A GRU-based Gaussian mixture variational autoencoder approach," in *Proc. 10th Asian Conf. Mach. Learn.*, 2018, pp. 97–112.

[7] M. Piekarski, J. Jaworek-Korjakowska, A. I. Wawrzyniak, and M. Gorgon, "Convolutional neural network architecture for beam instabilities identification in synchrotron radiation systems as an anomaly detection problem," *Measurement*, vol. 165, 2020, Art. no. 108116.

[8] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, "Anomaly detection based on convolutional recurrent autoencoder for IoT time series," *IEEE Trans. Syst., Man, Cybern. Syst.*, to be published, doi: 10.1109/TSMC.2020.2968516.

[9] T. Ergen and S. S. Kozat, "Unsupervised anomaly detection with LSTM neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 8, pp. 3127–3141, Aug. 2020.

[10] D. Wu, Z. Jiang, X. Xie, X. Wei, W. Yu, and R. Li, "LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5244–5253, Aug. 2020.

[11] X. Zhou, Y. Hu, W. Liang, J. Ma, and Q. Jin, "Variational LSTM enhanced anomaly detection for industrial big data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3469–3477, May 2021.

[12] B. Siegel, "Industrial anomaly detection: A comparison of unsupervised neural network architectures," *IEEE Sensors Lett.*, vol. 4, no. 8, Aug. 2020, Art. no. 7501104.

[13] L. Fang, Y. Li, Z. Liu, C. Yin, M. Li, and Z. J. Cao, "A practical model based on anomaly detection for protecting medical IoT control services against external attacks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4260–4269, Jun. 2021.

[14] A. Kiersztyn, P. Karczmarek, K. Kiersztyn, and W. Pedrycz, "Detection and classification of anomalies in large data sets on the basis of information granules," *IEEE Trans. Fuzzy Syst.*, to be published, doi: 10.1109/TFUZZ.2021.3076265.

[15] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Process.*, vol. 84, no. 7, pp. 1115–1130, Jul. 2004.

[16] K. Mukherjee and A. Ray, "State splitting and merging in probabilistic finite state automata for signal representation and analysis," *Signal Process.*, vol. 104, pp. 105–119, 2014.

[17] S. Gupta and A. Ray, "Symbolic dynamic filtering for data-driven pattern recognition," in *Pattern Recognition: Theory and Applications.* Commack, NY, USA: Nova, 2007, pp. 17–71.

[18] N. F. Ghalyan and A. Ray, "Symbolic time series analysis for anomaly detection in measure-invariant ergodic systems," *J. Dyn. Syst., Meas., Control*, vol. 142, no. 6, 2020, Art. no. 061003.

[19] N. F. Ghalyan and A. Ray, "Measure invariance of ergodic symbolic systems for low-delay detection of anomalous events," *Mech. Syst. Signal Process.*, vol. 159, 2021, Art. no. 107746.

[20] S. Mondal, N. F. Ghalyan, A. Ray, and A. Mukhopadhyay, "Early detection of thermoacoustic instabilities using hidden Markov models," *Combustion Sci. Technol.*, vol. 191, no. 8, pp. 1309–1336, Aug. 2019.

[21] V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," *Signal Process.*, vol. 86, no. 11, pp. 3309–3320, 2006.

[22] C. Bishop, *Pattern Recognition and Machine Learning.* New York, NY, USA: Springer, 2007.

[23] K. Murphy, *Machine Learning: A Probabilistic Perspective*, 1st ed. Cambridge, MA, USA: MIT Press, 2012.

[24] A. Glowacz, "Fault diagnosis of electric impact drills using thermal imaging," *Measurement*, vol. 171, 2021, Art. no. 108815.

[25] N. F. Ghalyan, I. F. Ghalyan, and A. Ray, "Modeling of microscope images for early detection of fatigue cracks in structural materials," *Int. J. Adv. Manuf. Technol.*, vol. 104, pp. 3899–3913, 2019.

Ibrahim F. Ghalyan (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the University of Baghdad, Baghdad, Iraq, in 2001 and 2003, respectively, and the Ph.D. degree in engineering science from the University of Luxembourg, Esch-sur-Alzette, Luxembourg, in 2016.

He was a Postdoctoral Researcher with New York University, New York, NY, USA. In 2020, he joined the Bank of New York Mellon, New York, where he is currently a Director of data science. His current research interests include machine learning, statistical signal processing, time series analysis, computer vision, and automation.

Najah F. Ghalyan received the B.S. and M.S. degrees in mechanical engineering from Al Nahrain University, Baghdad, Iraq, and the M.S. degree in electrical engineering and the Ph.D. degree in mechanical engineering from Pennsylvania State University, University Park, PA, USA, in 2000, 2003, 2016, and 2019, respectively.

He is currently a Lecturer of mechanical engineering with the University of Kerbala, Kerbala, Iraq. His current research interests include machine learning and AI risk management.
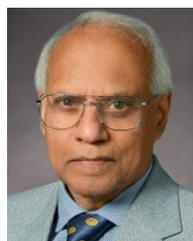
Asok Ray (Life Fellow, IEEE) received the graduate degrees in each discipline of electrical engineering, mathematics, and computer science, and the Ph.D. degree in mechanical engineering from Northeastern University, Boston, MA, USA, in 1976.

In 1985, he joined Pennsylvania State University, University Park, PA, USA, where he is currently a University Distinguished Professor of mechanical engineering and mathematics, a Graduate Faculty of electrical engineering, and a Graduate Faculty of nuclear engineering. He has authored or coauthored more than 650 research publications, including 320 scholarly articles in refereed journals.

Dr. Ray is a Fellow of ASME, AAAS, and WIF.