

An Early Classification Approach for Improving Structural Rotor Fault Diagnosis

Aneesh G. Nath^{ID}, *Student Member, IEEE*, Anshul Sharma^{ID}, *Student Member, IEEE*, Sandeep S. Udmale^{ID}, *Member, IEEE*, and Sanjay Kumar Singh^{ID}, *Senior Member, IEEE*

Abstract—Artificial intelligence (AI)-based rotating machinery fault diagnosis has extreme importance in the industrial automation and control systems since rotating machinery constitutes approximately 40% of the overall machinery in the industry. However, the majority of AI-based solutions in rotor faults diagnosis are in an experimental stage due to: 1) inadequate and unrealistic faulty data; 2) lack of opportunities to utilize domain-specific fault features; and 3) limitations in employing deep learning and advanced learning strategies. Moreover, structural rotor fault (SRF) is one of the critical but least addressed faults in rotor faults diagnosis even though it is the root cause of the majority of rotating machinery issues. Hence, we develop an SRF diagnosis framework, which addresses the issues of industrial data acquisition by creating a subsampled data set incorporating distinctive frequency components (DFC). The data scarcity and imbalance problems are handled through an augmentation method using soft-dynamic time warping (soft-DTW), enhanced by fault information content (FIC)-based weighing scheme. At the fault classification phase, we proposed an early classification (EC) approach for SRF that predicts the faults with an acceptable tradeoff between earliness and accuracy. In this line, first, a sequential deep learning classifier is developed by considering accuracy only as an objective. Then, early decision policy is defined by taking accuracy and earliness into account. The model demonstrated exceptional performance over state-of-the-art methods in SRF diagnosis and achieved 99.5% accuracy with 14% earliness on the Meggitt testbed data set and 98.32% accuracy with 55.68% earliness on the MaFaulDa data set.

Index Terms—Augmentation, early classification (EC), faults diagnosis, machine health monitoring, recurrent neural networks (RNNs).

I. INTRODUCTION

THE emergence of Industry 4.0 [1] gives rise to the development of a more complicated industrial system with increased automation and high precision. Prognostics and health management [2] is an unavoidable part of Industry 4.0. It ensures optimum cost, safety, availability, and reliability by

Manuscript received September 27, 2020; revised November 12, 2020; accepted November 26, 2020. Date of publication December 21, 2020; date of current version January 5, 2021. The Associate Editor coordinating the review process was Zhigang Liu. (*Corresponding authors: Anshul Sharma; Sandeep S. Udmale*.)

Aneesh G. Nath is with the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi 221005, India, and also with the Department of Computer Engineering, TBM College of Engineering, Kollam 691005, India (e-mail: aneeshgnath.rs.cse18@itbhu.ac.in).

Anshul Sharma and Sanjay Kumar Singh are with the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi 221005, India (e-mail: anshul.rs.cse16@itbhu.ac.in; sks.cse@itbhu.ac.in).

Sandeep S. Udmale is with the Department of Computer Engineering and IT, VJTI, Mumbai 243122, India (e-mail: ssudmale@it.vjti.ac.in).

Digital Object Identifier 10.1109/TIM.2020.3043959

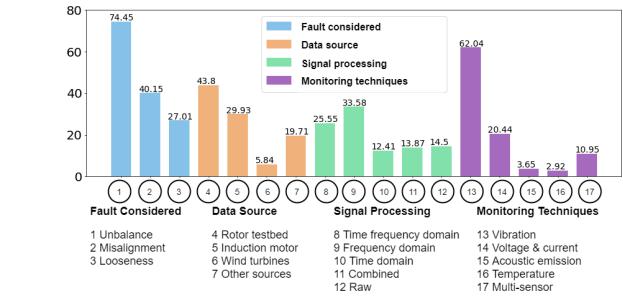


Fig. 1. Summary of the SRF literature.

avoiding chances of catastrophic failures, disastrous accidents, and unexpected shutdowns of the whole system. Rotating machinery constitutes approximately 40% of all machinery operated in the industry. Over a period of time, they are more prone to deterioration and failures in mechanical and electromechanical systems [3]. However, the majority of the studies that appeared in the literature of artificial intelligence (AI)-based fault diagnosis of rotating machinery were dealing only with bearing or gear faults, leaving rotor faults diagnosis least addressed [4]. Due to this, the rotor faults diagnosis literature looks fragmentary, lacking the opportunities for exploiting the fault-specific characteristics of the rotor faults and utilizing them in the feature engineering phase of AI to produce significant research improvements in the domain. Among the faults affecting the rotor, structural rotor fault (SRF) is a prevalent and straightforward fault, which includes unbalance (UB), misalignment, and looseness faults. Analyzing SRF is extremely important because it not only creates an unmediated baneful impact on the structural attributes and performance of the affected equipment, but it may cause secondary faults to the surrounding components, such as bearings and gears [5]. In addition, compared with bearing and gear faults, the SRF diagnosis is more challenging. It is characterized by spectral changes in rotating frequency and its harmonic frequencies in the vibration spectrum, known as distinctive frequency components (DFC). The significance of DFC in SRF diagnosis is compiled in Table I.

It is observed from the literature that there have been very few attempts made toward utilizing DFC in SRF, and the major contributions in this direction are given in Table II. A detailed review of SRF literature has been performed over 140 articles in [4], including machine learning and deep learning strategies, and the summary is shown in Fig. 1. The rotor testbed and

TABLE I
DFC AND SRF CORRELATIONS

Fault Type	A.P.	Symptomatic Frequency	Phase
Misalignment (Parallel)	Rad.	High 2x peak than 1x. Stronger 1x axial and torsional responses.	A phase shift of 180° in the radial position
UB (Static)	Rad.	1x is larger. Harmonics with less than 15% of the 1x	A phase shift of 0° in Radial direction
UB (Couple)	Rad.	1x is larger. Harmonics with less than 15% of the 1x	A phase shift of 180° in Radial direction
UB (Dynamic)	Rad.	1x is larger. Harmonics with less than 15% of the 1x	0° to 180° Phase shift.
Looseness	Axl. & Rad.	2x dominating. Then 1x to 10x harmonics, subharmonics, multiples of subharmonics	Unstable phase reading

A.P: Affected Plane, Rad.: Radial, Axl.: Axial

TABLE II
FAULT SPECIFIC COMPONENT USAGE

Category	Model	Significant contributions
machine learning	ANN	[6], [7], [8], [9] , [10], [11]
	SVM	[12], [13], [14], [15], [16]
	CNN	[17], [18]
deep learning	DBN:	[19], DNN: [20]

the induction motor are more commonly used as a data source. Besides, wind turbines and other sources, such as open-source data sets, have also been utilized. Among the signal processing methods, frequency-domain (FD) processing is widespread, which is followed by the time-FD (TFD) and time-domain (TD). Moreover, it is observed that the vibration signal-based health monitoring technique is the most popular in SRF. Along with that voltage and current, acoustic emission, temperature, and multisensor monitoring have also been employed.

The literature informs that the conventional TD feature extraction fails in identifying symptom parameters due to the irregularity of the raw data [5]. Thus, the FD or TFD feature extraction is most widely adopted, but the features extracted per sample basis turned out to be insufficient to use for deep learning models. Besides, the automatic feature extraction through deep learning models fails to utilize DFC efficiently. As a result, a significant amount of rotor faults diagnosis literature falls into the shallow learning category. In addition to that, SRF diagnosis often uses a vibration signal in the data monitoring method. However, these signal demands for the sequential data processing due to univariate or multivariate time-series (TS) characteristics. However, the TD or TFD feature processing lacks the temporal information in the data, which restricts the scope of using sequential learning models. The most commonly available data source for SRF is the rotor testbed, but the majority of rotor testbeds fail in reflecting various real-time industrial situations, such as varying speed and load conditions, unevenly sampled data (due to irregular triggering of sensors), and abnormal or noisy readings [21].

In order to mitigate the data collection and feature generation related issues, a data subsampling is introduced in this

TABLE III
AUGMENTATION IN ROTOR FAULTS' DIAGNOSIS

Work	Major Technique	TS
[36],[37],[38]	Slicing with overlap	No
[39],[17], [40]	Slicing without overlap	No
[41]	Combining consecutive samples	No
[42]	Label Dilation and semi-supervised data augmentation (SSDA)	No
[43]	Auxiliary classifier GAN (ACGAN)	Partial
[44]	Shrinking, randomly stretching, adding/ removing	Partial
[21]	Data simulation by resampling (DSR)	Partial

article. It creates summary information at regular intervals from the continuous stream of data and combines both TD values and DFC into a time-ordered sequential feature space. Since the subsampling reduces the number of data points, the lesser training samples tend to overfit. To alleviate this problem, this work proposed a weighted soft-dynamic time warping (soft-DTW)-based augmentation scheme that utilizes the fault information content (FIC). The data in SRF are most often periodic vibration signals that possess temporal behavior. Thus, it is enticing to have an augmentation scheme that follows TS property. However, the available augmentation literature hardly deals with TS data. Some of the important augmentation strategies in fault diagnosis and with their TS consideration are given in Table III. The process of creating new samples in TS with shrinking, randomly stretching, removing, or adding signal parts [22], perturbing signal [23], and so on partially mitigates the problem of preserving the signal properties. Also, certain works found in biosignal recordings [24]–[27] of medical signal processing that augments TS data. The most advanced and successful algorithm available in this literature is DTW barycenter averaging (DBA) [28]. It utilizes multiple aligned TS with weighted averaging to generate new samples [29] that uses DTW [30] as the distance metric. The suitability of DBA for the training of deep learning models was discussed by Fawaz *et al.* [31]. At the same time, in rotating machinery fault diagnosis, specific DL methods proved their efficiency in dealing with the sensor data properly without needing augmentation. The CNNs that extract features automatically from the 1-D raw vibration were described in [32] and [33]. The DBN model used in [34] employed independent DBNs on each sensor data with information fusion. Among the sequential DL models, LSTM was used by Lei *et al.* [35] to deal with multivariate TS data.

Moreover, from the machine health monitoring perspective, making informed decisions as early as possible without waiting for full-length TS data is highly desirable. In the literature, early class prediction of TS, based on partially observed data points, is called an early classification (EC) [45], [46]. On the other hand, the traditional classification (TC) approach performs class prediction when full-length TS becomes available, which delays the decision and also increases the response time. In recent times, EC became an exciting research topic among researchers of various fields [47]. In some early

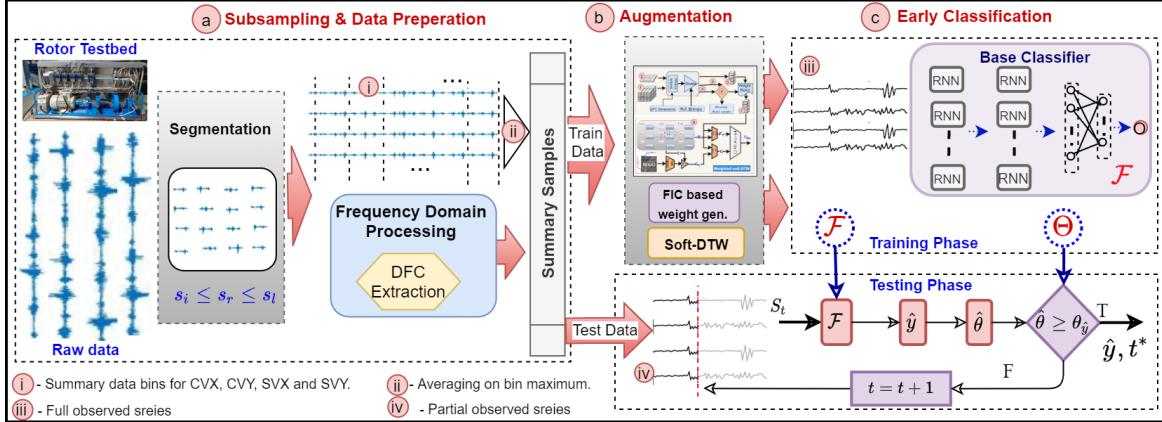


Fig. 2. Overall framework.

attempts, Bregón *et al.* [48] presented an idea of classifying incomplete TS for fault classification in dynamic system [48]. Xing *et al.* [45] proposed an instance-based EC approach by introducing the concept of the minimum required length and also addressed the tradeoff between the earliness and the accuracy. Furthermore, they introduced a feature-based early distinctive shapelet classification method whose results are highly interpretable to the end-users [49]. Mori *et al.* [46] proposed an EC method by learning the reliability threshold and also discriminating the classes over time. Similarly, the confidence threshold was defined by fusing the classifier's true prediction capability at successive time steps in [50]. An optimization-based EC approach on multivariate TS has been introduced by learning optimal decision rule in [51]. Besides, EC approach has been adopted in many useful applications in various domains, such as early drought prediction [52], early disease prediction [53], early malware detection [54], and early transportation mode detection [55], [56]. The EC approach is highly beneficial for the prediction of machinery faults, but it is still unexplored in the SRF literature. Thus, an EC-based deep learning approach is introduced for SRF diagnosis in this article. The main contributions of the article can be summarized as follows.

- 1) The proposed framework presents a sequential data representational form to bridge the gap between the practical industrial data and the experimental laboratory data. Also, it deals with unevenly sampled or missed data, varying operating speed conditions, and suppresses other sensor issues related to raw data, incorporating domain-specific SRF information.
- 2) A weighted soft-DTW-based data augmentation scheme is introduced and enhanced with FIC, ensuring the TS property. It augments the data even with a few reference samples of different lengths, which is essential in solving practical data collection issues.
- 3) This framework encourages the use of DFC and provides results over sequential deep learning models along with EC strategy, emphasizing the effect of DFC.
- 4) An EC-based deep learning approach is proposed for early SRF prediction that learns the classwise confidence

threshold for reliable class prediction by balancing the tradeoff between reliability in prediction and earliness.

- 5) Finally, the proposed method is tested using two data sets. A novel SRF data set has been created by a test-bed, simulating industrial scenarios, and is used as the primary data set. Along with that, a publicly available data set with varying speed and load conditions has also been utilized, and the model's performance is compared with state-of-the-art methods.

The rest of this article is organized as follows. Section II describes the three important processes in the proposed framework. Experimental settings, data collection processes, results, and discussions are given in Section III. Section IV draws the concluding remarks.

II. PROPOSED METHOD

In this section, we present the overview of the proposed framework for SRF diagnosis, which is depicted in Fig. 2. The proposed paradigm consists of three components, and they are described in the following subsections. Section II-A describes the subsampled data preparation that transforms the raw data into a feature space without losing the sequential properties of data. Section II-B illustrates the data augmentation process that helps to overcome the overfitting problem of the model and improve testing performance. Finally, Section II-C demonstrates the EC model (ECM) that consists of two phases: training and testing. In the training phase, ECM learns the base classifier (\mathcal{F}) and classwise reliability threshold $\Theta \in \mathcal{R}^k$. In the testing phase, the ECM processes the incoming signal in transformed feature space and predicts the class label (fault type) when the corresponding reliability threshold is satisfied.

A. Subsampling and Data Preparation

The key idea is to utilize the summary information at regular intervals by subsampling data points from the continuous stream of data instead of learning from raw data directly. This helps in mitigating the data acquisition issues in real plant scenario. Each of the summary data points combines both TD signals and FD DFCs into a time-ordered sequential feature space to exploit the advantages of both the domains.

Then, a sufficient number of sequential summary data points concerning the monitoring period of actual industrial scenarios are assigned to each sample. Such samples, created in a subsampled feature space, have the advantages of the temporal property of data points and sufficient discriminative information content. The raw data summarization interval (S_t) is selected based on two constraints. First, it ensures a minimum number of raw data points to confirm proper DFC extraction. Second, it provides a proper subsampling ratio, which is much closer to the unit observation period of the industrial data-based solutions. Let r_s be the speed of revolution in rpm, f_r the rotational frequency ($r_s/60$), and f_s the sensing frequency, and then, the number of sampling points per rotation (s_r) is $s_r = 60f_s/r_s$ or f_s/f_r . The minimum number of sampling points per segment (s_l) is decided in such a way that it satisfies the condition: $s_l \leq s_r \leq s_l$, where s_l is the number of interval points between the samples (it is set to zero in nonoverlapping segmentation). Thus, the condition $s_r \leq s_l$ ensures every segment contains one or more rotation data. Similarly, the maximum points are decided based on the data sampling rate of industry-based vibration monitoring solutions. Hence, the subsampling duration of one second is agreed to result in a segment length of f_s points, in our experiments.

In order to extract the DFCs, first, the FFT spectrum is evaluated on each segment. Since the fluctuating speed and changing operating conditions of the industry are considered in this experiment, normalization [5] is needed to reduce the rate difference under varying operating conditions. This process is given by

$$\mu_{\text{ampl}} = \frac{1}{S_f} \sum_{i=1}^{S_f} f_{iX\text{ampl}} \quad (1)$$

$$N(f_{iX}) = \frac{f_{iX\text{ampl}}}{\mu_{\text{ampl}}} \quad (i = 1, 2, \dots, S_f) \quad (2)$$

where S_f is the number of the significant frequency band, $f_{iX\text{ampl}}$ denotes the amplitude analogous to f_{iX} frequency, μ_{ampl} is the mean of amplitudes corresponding to the significant frequency bands, and $N(f_{iX})$ is the normalized amplitude values. Then, the DFCs are given by $N_i = N(f_{iX})$ for $i = 1, 2, \dots, h_f$, where h_f is the number of rotating frequency components. These are extracted by a multipass filter in range $f_{iX} - \Delta f, f_{iX} + \Delta f$, where Δf is the fluctuating frequency range. Moreover, it is observed that the rpm variation is less than 10%, and the impact of Δf in amplitude is comparatively less. Similarly, the phase information corresponding to these frequencies is also extracted. From Table I, it is evident that the first three harmonics of the rotating frequency (1X, 2X, 3X) and its combinations with phase in both radial and axial directions are sufficient to distinguish between different SRFs. The summary data values in the TD are generated by finding the maximum values in every bin of size s_r data points and averaging such f_s/s_r number of values. It is shown as numbers ‘i’ and ‘ii’ in pink circles in Fig. 2. This new format of data meets the data representation used by vibrosight industrial solution with the Meggitt data set [57]. This phase creates a new uniformly subsampled, ordered sequential set of values represented as $X = [X_1, X_2, \dots, X_N]$, where each

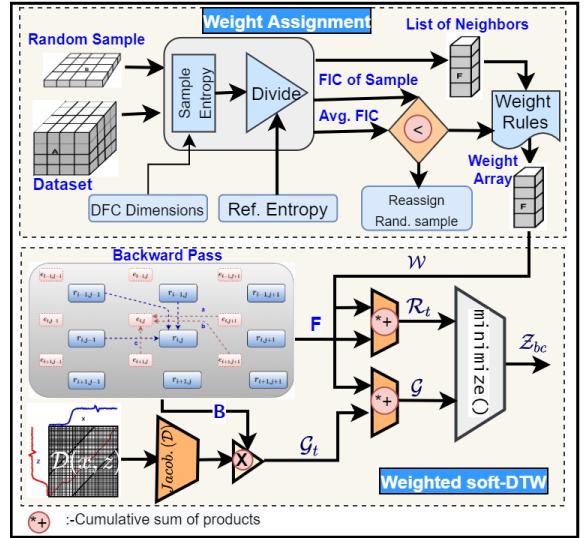


Fig. 3. Soft-DTW barycenter approach.

$X_i \in R^{(L \times M)}$ and L is the length of TS, each with dimension M in the subsampled space.

B. Soft-DTW-Based Augmentation

To avoid overfitting and increase the quality and the diversity of training samples, a soft-DTW-based data augmentation scheme is adopted in this work. This scheme is based on a DBA algorithm that uses an approximation approach to find the consensus sequence from a set of sequences. Forestier *et al.* [29] changed the objective function of DBA, incorporating a weight factor with each sequence to generate the average sequence. We have modified this scheme for augmenting TS signals in the subsampled space in the following ways.

- 1) The proposed augmentation scheme incorporates the DFC of SRF in weight distribution, thus making it more domain-specific. It utilizes FIC to select the contributing samples for synthesizing more prudent samples, increasing the classification accuracy.
- 2) The soft-DTW barycenter approach is adopted to achieve smoother barycenters for effective learning [58].

This augmentation scheme preserves the sequential property and class discriminative ability of synthesized samples, and it deals with fewer reference samples and is able to work with a different length series. Algorithm 1 along with Fig. 3 demonstrates soft-DTW barycenter generation. It has two main processes: the first one is the weight assignment process, and the second is the soft-DTW barycenter generation using these weights.

The weight assignment process is based on the sample entropy [59]. The entropies of the DFC columns in a sample are averaged to find the representative entropy of a sample. The ratio of the representative entropies of faulty and healthy samples is then calculated to find the FIC of a particular sample. This process is given in Algorithm 1 from steps 1 to 9. Next, in order to assign weights, initially, a TS X_i has been randomly selected from the data set of a particular fault. Then, its FIC is compared with the average FIC of the whole

data set. If it is less, then a new random sample is selected, and the process is repeated. This process has been given in steps 10–13 in Algorithm 1 and in the first part of Fig. 3.

The weight distribution process is shown in step 15 of Algorithm 1. In the process, the random sample that meets the minimum FIC criteria is assigned a weight of 0.3. After that, its $N \times 0.1$ (10% of total samples) nearest indexed neighbors are selected and ranked based on the sample FIC. The first two neighbors with the highest FIC are assigned a weight of 0.15 each, and then, the next two neighbors are given the weight 0.1 each. To ensure the normalized sum of weights, the rest of the samples in the neighboring subset shares the remaining 0.2 weight equally to create the weight array. These weight values are identified to keep a balance between two objectives, i.e., diversity and the discriminative properties of the synthesized samples. Moreover, the approach of Fawaz *et al.* [31] is also considered with some modification to give the first four neighbors more weightage. The described scheme assists in selecting the most significant participating samples, and thereby, a proportional reduction in computational cost has been achieved.

The weight assignment process follows the barycenter calculation by DBA with the help of DTW, which uses sequential computation. However, since the min operator is not continuous in DTW, it limits gradient or subgradient calculation. Mit Shah *et al.* [60] used a soft-min operation for solving this issue. Thus, the differentiability of this DTW function is achieved by a generalized min operator with a smoothening parameter β , and it is expressed as

$$\min^\beta \{p_1, \dots, p_n\} = \begin{cases} \min_{i \leq n} p_i, & \beta = 0 \\ -\beta \log \sum_{i=1}^n e^{-p_i/\beta}, & \beta > 0. \end{cases} \quad (3)$$

Then, soft-DTW with β can be defined as

$$dtw_\beta(x, y) = \min^\beta \{\langle \mathcal{P}, \mathcal{D}(x, y) \rangle, \mathcal{P} \in P_{m \times n}\} \quad (4)$$

where $\mathcal{D}(x, y)$ is the pairwise distance matrix and $\mathcal{P} \in P_{m \times n}$ is the path alignment matrix in the basic DTW calculation. Similarly, a global alignment kernel (\mathcal{K}), defined by Cuturi *et al.* [61], is also able to compare two TSs, which is given by

$$\mathcal{K}^\beta(x, y) = \sum_{\mathcal{P} \in P_{m \times n}} e^{-\langle \mathcal{P}, \mathcal{D}(x, y) \rangle / \beta}. \quad (5)$$

Later, the same soft-min idea was used to describe the gradient of the soft-DTW with the chain rule [58], and it is shown as

$$\nabla_x dtw_\beta(x, y) = \left(\frac{\partial \mathcal{D}(x, y)}{\partial x} \right)^T \frac{\sum_{\mathcal{P} \in P_{m \times n}} e^{-\langle \mathcal{P}, \mathcal{D}(x, y) \rangle / \beta}}{\mathcal{K}^\beta(x, y)} \quad (6)$$

where $(\sum_{\mathcal{P} \in P_{m \times n}} e^{-\langle \mathcal{P}, \mathcal{D}(x, y) \rangle / \beta}) / (\mathcal{K}^\beta(x, y))$ is the average alignment matrix, and it is denoted by $E_\beta[\mathcal{P}]$.

To find the differentiation of $dtw_\beta(x, y)$ algorithmically, first, the forward pass by Bellman's equation is performed. This stores the intermediary computations, which results in

Algorithm 1 FIC-Based Weighted Soft-DTW

Input: X^f is a set of N^f TS of fault type f , and $z = X^f[k]$, where k is a random number $1 \leq k \leq N^f$, $z \in \mathbb{R}^{(L \times M)}$

Output: \mathcal{Z}_{bc} : optimal average sequence

Initialization :
Let $W^f \in \mathbb{R}^{N^f}$ is the weight vector initialized to 0,
 D_ϕ^{if} be the d^{if} DFC columns of i^{th} sample of X^f , ϵ^h be the entropy of reference sample, d be the distance fuction and set $S = 0$, $\mathcal{G} = 0$, $\mathcal{R}_t = 0$.

```

1: for  $i \leftarrow 1$  to  $N^f$  do
2:   for  $j \leftarrow 1$  to  $d^{if}$  do
3:      $S \leftarrow S + \text{Sample\_entropy}(D_\phi^{if}[j])$ 
4:   end for
5:    $\epsilon^{if} \leftarrow S/d^{if}$  /*Entropy of  $i^{th}$  sample of fault  $f$ 
6:    $\epsilon^f[i] \leftarrow \epsilon^{if}/\epsilon^h$ . /*FIC assigned per sample
7:    $E[i, 0] \leftarrow \epsilon^f[i]$ 
8:    $E[i, 1] \leftarrow i$ 
9: end for
10:   $\epsilon_{cls} \leftarrow \epsilon^f/N^f$  /* Average FIC of fault class
11:  if  $(E[k, 0] < \epsilon_{cls})$  then
12:    Find new random index  $k$  and reassign  $X^f[k]$  to  $z$ .
13:  end if
14:  /*Sort sample indices based on the key FIC
15:   $E \leftarrow \text{Sort}(E, \text{key} : E[:, 0])$ 
16:  Distribute weights in  $W^f$  w.r.t index position  $k$ 
17:  return  $\mathcal{Z}_{bc}$ 

```

Procedure $\text{Fun_sdtw}()$

```

18: for  $l \leftarrow 1$  to  $N^f$  and  $W^f[l] \neq 0$  do
19:    $x \leftarrow X_l^f$ 
20:   Compute the cost matrix  $\mathcal{D}(x, z) \in \mathbb{R}^{|x| \times |z|}$ 
/*Forward pass to compute  $F \leftarrow dtw_\beta(x, z)$ 
21:   for  $i \leftarrow 1$  to  $|x|$  do
22:     for  $j \leftarrow 1$  to  $|z|$  do
23:        $f_{i,j} \leftarrow d(x, z) + \min^\beta \{f_{i-1,j-1}, f_{i-1,j}, f_{i,j-1}\}$ 
24:     end for
25:   end for
/*Backward pass to compute error  $B$ , assuming
 $b_{i,|x|+1} = b_{|z|+1,j} = 0$  and  $b_{|x|+1,|z|+1} = 1$  for
 $i \in |x|, j \in |z|$ 
26:   for  $i \leftarrow |x|$  to 1 do
27:     for  $j \leftarrow |z|$  to 1 do
28:        $a_1 \leftarrow e^{\frac{1}{\beta}(f_{i+1,j} - f_{i,j} - d_{i+1,j})}$ ,
 $a_2 \leftarrow e^{\frac{1}{\beta}(f_{i,j+1} - f_{i,j} - d_{i,j+1})}$ , and
 $a_3 \leftarrow e^{\frac{1}{\beta}(f_{i+1,j+1} - f_{i,j} - d_{i+1,j+1})}$ 
29:        $b_{i,j} \leftarrow b_{i+1,j} \cdot a_1 + b_{i,j+1} \cdot a_2 + b_{i+1,j+1} \cdot a_3$ 
30:     end for
31:   end for
32:    $\mathcal{G}_t \leftarrow \text{Jacobian}(\mathcal{D}).B$ 
33:    $\mathcal{G} \leftarrow \mathcal{G} + W^f[l] * \mathcal{G}_t$ 
34:    $\mathcal{R}_t \leftarrow \mathcal{R}_t + W^f[l] * F$ 
35: end for
36: return  $\mathcal{R}_t, \mathcal{G}$ 

```

$F = [f_{(i,j)}]$. Thereafter, the chain rule is used to find the impact of change in $f_{(i,j)}$ affecting the end result $f_{(m,n)}$ of the forward pass. In the backward pass, the entire matrix $\mathcal{B} = [b_{(i,j)}]$ is calculated starting from $b_{(m,n)}$ down to $b_{1,1}$. The forward and backward passes of soft-DTW are specified in lines 21–31 in Algorithm 1. The heterogeneity of the synthesized samples is ensured by the randomness of the samples and weights used in the weight assignment process. The vectorized operation of applying these weights to generate diversity in barycenters is given in lines 32–34 in the algorithm. These operations are shown in the second part of Fig. 3. Finally, the minimization function returns the optimal soft-DTW barycenter.

C. Early Classification Model

The EC is the process of classifying the sequence data by observing partial information, with an acceptable level of accuracy. In this work, the proposed ECM for fault diagnosis follows a twofold process. In the first fold, the base classifier \mathcal{F} is developed by taking accuracy only as the objective similar to the TC approach in which class prediction is made once the complete TS becomes available. In the second fold, the classwise confidence threshold is defined by taking the reliability of prediction and earliness into account.

1) *Sequential Classifier*: In practice, most of the existing works gave the least importance to the temporal properties of TS data in SRF classification. They have been developed based on global features set, which seems to produce unrealistic solutions. The commonly used sequential deep learning models, such as a simple recurrent neural network (RNN), long short-term memory (LSTM) [62], and gated recurrent unit (GRU) [63], have the sequential connections between their nodes. It makes them capable of learning temporal dynamic behavior for the input time sequence. The simple RNN replaces the whole activation, while LSTM and GRU regulate the information in each cell, which helps the latter models to tackle vanishing/exploding gradient problems. With this motivation, we proposed a deep learning architecture that consists of L number of recurrent layers. These layers capture the temporal information and their long-term relationships from the signal. Thereafter, the fully connected layer is added, which provides a higher level representation of data and also helpful in discriminating the classes well. Finally, the last layer performs the fault classification. We developed two classifiers, \mathcal{M}_1 and \mathcal{M}_2 , respectively, based on LSTM and GRU architectures.

We determine the hyper parameters of the classification models, \mathcal{M}_1 and \mathcal{M}_2 , by considering the categorical cross-entropy as a loss function and Adam as a optimizer. Moreover, the proposed model uses tanh, relu, and softmax as activation functions in RNN, dense, and output layers, respectively. We have searched over recurrent layers $L \in \{1, 2, 3\}$, hidden nodes HN $\in \{16, 32, 64, 128\}$, and learning rate $\eta \in \{0.1, 0.01, 0.001\}$ for 300 epoch. Finally, the best parameters $L = 2$ and $\eta = 0.001$ are considered.

2) *Confidence Threshold*: The proposed ECM defines the classwise confidence threshold (θ) to take the early decision on incoming TS. The confidence threshold is used

to measure the reliability of class prediction that analyzes whether the observed sequence is sufficient enough for class prediction or not. Basically, ECM processes the incoming TS X at every time step t and computes the confidence of predicted class $\hat{y} = \mathcal{F}(X_t)$, denoted by $\hat{\theta}_t$. Moreover, ECM predicts the class label only if computed $\hat{\theta}_t$ is higher than the predefined confidence threshold θ .

3) *Training Phase*: In this phase, first, the base classifier $\mathcal{F} \in \{\mathcal{M}_1, \mathcal{M}_2\}$ is trained using full-length summary data. To learn the more accurate model, the training set is increased by augmentation with the proposed approach, as discussed in Section II-B. The classifier \mathcal{F} is trained with complete training set by considering accuracy only as objective. Next, ECM learns the confidence thresholds $\{\theta_1, \theta_2, \dots, \theta_k\}$ and follow a similar approach, as defined in [50]. However, the proposed ECM learns the classwise threshold, which is more adaptable for SRF diagnosis. The complete learning process of ECM is presented in Algorithm 2.

To learn the confidence threshold, we consider \mathcal{F} as a pretrained model and compute the performance of classifier \mathcal{F} at each timestep t , denoted by \mathcal{F}_t^P . Basically, \mathcal{F}_t^P measures the possibility of class prediction y , while the predicted class label is \hat{y} . It is formally defined as

$$\mathcal{F}_t^P(y, \hat{y}) = \frac{\|\{X^i | (y^i = y) \wedge (\mathcal{F}(X^i) = \hat{y})\}\|}{\|\{X^i | \mathcal{F}(X^i) = \hat{y}\}\|}. \quad (7)$$

Thus, the confidence in prediction (\hat{y}) at single time point t is computed as $\mathcal{F}_t^P(y = \hat{y}, \hat{y})$. To evaluate the \mathcal{F}_t^P , partial data \mathcal{D}_t are used in which each $X_t \in \mathcal{D}_t$ contains only t data points. As the classifier \mathcal{F} has been trained using full length (T), X_t is not acceptable by \mathcal{F} . To deal with this problem, each X_t is padded with current mean in prefix manner. It fulfills the two requirements. First, it makes the input acceptable to the pretrained model. Second, it captures the mean distribution of current signals for unobserved series. This step has been shown in lines 6–9 in Algorithm 2.

As the TS data are collected over time, we can get the class prediction at each time step t . Therefore, all the predictions up to t are utilized to compute the composite confidence for class prediction, which is defined as

$$\psi(\hat{y}_t) = 1 - \prod_{j=1}^t (1 - \mathcal{F}_j^P(\hat{y}_t | \hat{y}_j)). \quad (8)$$

Finally, we define the optimal θ_c for each class of SRF that supports reliable class prediction early in time. In this process, all possible threshold candidates ($\theta_t^i, i \in [1, N], t \in [1, T]$) are computed for the training set. Next, the best classwise threshold is selected by balancing the tradeoff between the reliability score of predictions and earliness. This step has been shown by lines 10–38 in Algorithm 2.

4) *Prediction Phase*: In the prediction process, the ECM processes the incoming X at each time point t and computes the class label $\hat{y}_t = \mathcal{F}(X_t)$. Furthermore, it computes the confidence of \hat{y}_t by (8). If incoming X at time t satisfies the reliability threshold condition, ECM will predict the class label; otherwise, ECM will wait to add more data points in the

Algorithm 2 ECM Training

Input: Data set $\mathcal{D} = \{(X^i, y^i)\}_{i \in [1, N]}\}$, The set of classes $\mathcal{C} = \{c_1, c_2, \dots, c_k\}$, and the balancing parameter α

Output: Trained classifier \mathcal{F} , Classifiers performance matrices $\{\mathcal{F}_t^P(y|\hat{y})|y, \hat{y} \in \mathcal{C}, 1 \leq t \leq T\}$ and Confidence threshold $\{\theta_1, \theta_2 \dots \theta_k\}$

```

1:  $\mathcal{D}' \leftarrow \text{dataAugmentation}(\mathcal{D})$ 
2:  $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}'$ 
3:  $\mathcal{F} \leftarrow \text{trainModel}(\mathcal{D})$ 
4: for  $t \leftarrow 1$  to  $T$  do
5:    $\mathcal{D}_t \leftarrow \text{truncateData}(\mathcal{D}, t)$ 
6:    $\mathcal{D}'_t \leftarrow \text{paddingDataWithMean}(\mathcal{D}_t)$ 
7:   predict class label  $\mathcal{F}(\mathcal{D}'_t)$ 
8:   compute  $\mathcal{F}_t^P(y|\hat{y})$  using Eq. (7)
9: end for
10: for  $i \leftarrow 1$  to  $N$  do
11:   for  $t \leftarrow 1$  to  $T$  do
12:      $\hat{y}_t^i \leftarrow \mathcal{F}(X_t^i)$ , store  $\hat{y}_t^i$ 
13:      $\theta_t^i \leftarrow \psi(\hat{y}_t^i)$  using Eq. (8), store  $\theta_t^i$ 
14:      $\Theta \leftarrow \Theta \cup \theta_t^i$ 
15:   end for
16: end for
17:  $\Theta \leftarrow \text{sort}(\Theta)$ 
18: for  $j \leftarrow 2$  to  $\text{len}(\Theta)$  do
19:    $\theta' \leftarrow \text{mid}(\Theta[j - 1], \Theta[j])$ 
20:   for  $i \leftarrow 1$  to  $N$  do
21:     for  $t \leftarrow 1$  to  $T$  do
22:       if  $\theta_t^i \geq \theta'$  then
23:          $R_i \leftarrow (\hat{y}_t^i, \frac{t}{T})$ , break
24:       end if
25:     end for
26:   end for
27:   for  $c$  in  $\mathcal{C}$  do
28:     Compute  $TP_c$ : True positive,  $FP_c$ : False positive and  $FN_c$ : False negative
29:      $P_c \leftarrow TP_c / (TP_c + FP_c)$ 
30:      $Q_c \leftarrow P_c / (TP_c + FN_c)$ 
31:      $FS_c \leftarrow 2 * (Q_c * P_c) / (Q_c + P_c)$ 
32:      $ER_c \leftarrow \sum_{i=1}^{N} \frac{R_i[2][R_i[1]=c]}{N}, i = 1, 2, \dots, N$ 
33:      $CF_c^\theta \leftarrow \alpha * (1 - FS_c) + (1 - \alpha) * ER_c$ 
34:   end for
35: end for
36: for  $c$  in  $\mathcal{C}$  do
37:    $\theta_c \leftarrow \arg \min_{\theta \in \Theta} \{CF_c(\theta)\}$ 
38: end for

```

TS and repeats the process. The complete prediction process of ECM is demonstrated in Algorithm 3.

III. EXPERIMENTS AND DISCUSSION

A. Experimental Settings and Data Collection

To evaluate the effectiveness of the proposed framework, the data have been collected from a rotor kit setup (Meggit-Mi 19003), which simulates the real plant varying speed working environment. In addition, a publically available data set called

Algorithm 3 ECM Prediction Process

Input: X : incoming TS, \mathcal{F} : trained classifier, $\{\theta_1, \theta_2 \dots \theta_k\}$: confidence thresholds

Output: \hat{y} : predicted class label, t^* : time point at which prediction is made

```

1: for  $t = 1$  to  $T$  do
2:    $X_t \leftarrow \text{paddingDataWithMean}(X_t)$ 
3:    $\hat{y} \leftarrow \mathcal{F}(X_t)$ 
4:   compute  $\hat{\theta} = \psi(\hat{y})$ , using Eq. 8
5:   if  $\hat{\theta} \geq \theta_{\hat{y}}$  then
6:     return  $\hat{y}$  and  $t^* \leftarrow t$ 
7:   end if
8: end for

```

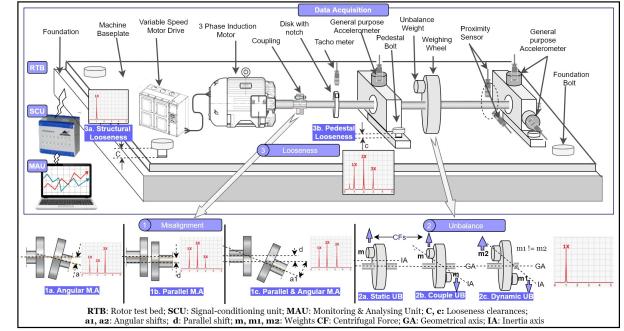


Fig. 4. Data collection and fault simulation setup.

Machinery Fault Database (MaFaulDa) [64] is operated on the proposed paradigm. MaFaulDa is an extensive set of multivariate TS of SRF and utilizes SpectraQuest's machinery fault simulator for data generation process.

1) *Meggitt Data Set (DS-1)*: The data acquisition setup contains the element like rotor testbed, signal-conditioning unit (SCU), and monitoring and analyzing unit (MAU), as shown in Fig. 4. It is comprised of an electric motor controlled by variable frequency drive (VFD, which is responsible for varying the frequency and voltage supply) and a rotating shaft with a flexible coupling, supported by bearing housings. The contact-type general-purpose accelerometers (CA202 piezoelectric accelerometer) are placed in the radial direction in bearing housings to measure casing vibrations. The noncontact-type eddy current-based proximity sensors (TQ 402/EA 402) are situated close to the shaft for sensing shaft vibration and the speed for reference. The data set consists of six different fault classes: healthy, static UB, couple UB, dynamic UB, misalignment, and looseness. The testbed setup and data collection conditions are given in Table IV.

2) *MaFaulDa Data Set (DS-2)*: The fault simulator of the MaFaulDa data set is mainly equipped with two sets of accelerometers arranged in three orthogonal (axial, radial, and tangential) directions. It also contains a tachometer and a microphone to measure the system rotation frequency and operating sound, respectively. The machine is capable of capturing data under normal and UB conditions with seven different weights. The horizontal and vertical misalignments with different shift distances were considered for the data collection process. It comprises a total of 1951 different

TABLE IV
TESTBED SETUP

Features	Specification / values
Sensors	Piezo-electric accelerometer, Eddy current based proximity sensors
RPM / Frequency	RPM Range: 0 - 3600 Subsampled Frequency: 1Hz
Sample Length	5 minutes/sample after subsampling
Motor Power	1 Kwh
Rotor Speeds (rpm)	750, 1500, 2250, 3000, 3250 with Runup and Rundown
Unbalance weight	13g, 26g (Dynamic Unbalance)
Diameter	Shaft:12.66mm, Weighing Disc: 100.18mm

TABLE V
TEST SETUP CONDITION: MAFAULDA DATA SET

Features	Specification/ values
Sensors	Accelerometers (radial, axial tangential directions). Analog tachometer and Microphone.
Frequency RPM	Sampled Frequency: 51.2 KHz RPM Range: 700-3600 rpm
Sampling Time	5 seconds/sample
Motor Power	$\frac{1}{4}$ CV DC
Rotor Speeds	10Hz to 60Hz
Unbalance weight	6g, 10g, 15g, 20g, 25g, 30g, 35g
Misalignment (mm)	Vertical: 0.51, 0.63, 1.40, 1.90, 1.27, 1.78 Horizontal: 0.50, 1.0, 1.50, 2.0
Diameter	Shaft: 16 mm, Rotor: 152.4 mm

scenarios, each one described by eight signals acquired at 50 kHz over a time interval of 5 s. A detailed description of the test setup is given in Table V.

The original data collected from the rotor testbed are summarized in the subsampled space at 1-s duration. The data set contains six classes with five different speeds, and each sample is organized to have 5-min duration. Then, it is augmented to improve the size of the training data so that the original and augmented data together form the overall training samples. The same subsampling, but with overlapping, is performed on the MaFaulDa data set selecting the closest matching speeds of DS-1 with all load conditions.

The effectiveness of the proposed framework is evaluated to attain the earliness of fault classification by keeping a decent tradeoff with accuracy. To mine the useful fault pattern information of SRF, the individual and combined effects of TD and DFC features in the subsampled space with the sequential models are analyzed. Besides, the capacity of the framework in dealing with real plant data is examined with a separate pipeline denoted as PL-2. The performance of the model on PL-2 is compared with the standard data specified by PL-1. The PL-2 data set is designed to simulate unevenly sampled or missed data situation by randomly removing a few data points from subsampled space.

TABLE VI
TRAINING PERFORMANCE

	\mathcal{M}_1			\mathcal{M}_2		
	train acc	val acc	test acc	train acc	val acc	test acc
DS-1	99.29	97.73	98.73	100.00	99.15	99.49
DS-2	99.82	99.76	97.96	99.93	99.88	98.78

B. Performance Evaluation

Each data set is initially partitioned in the ratio of 70%–30% for training and testing, respectively. Besides, to evaluate the performance of ECM, we use the following metrics.

1) *Accuracy*: It measures the percentage of correctly classified samples over total number of testing samples, formally defined as

$$\text{Accuracy} = \frac{\sum_{i=1}^N (\hat{y}^i = y^i)}{N} \times 100. \quad (9)$$

2) *Earliness*: It used to measure performance of ECM on TS. ECM predicts the class label by observing partial TS. Hence, earliness of testing TS refers to the number of data points ($t^* \leq T$) utilized for class prediction. Therefore, earliness of ECM is the average percentage of predicted length (t^*) to full length of TS (T). It is defined as

$$\text{Earliness} = \frac{1}{N} \sum_{i=1}^N \frac{t_i^*}{T} \times 100. \quad (10)$$

C. Impact of Subsampling and Augmentation

Table VI demonstrates the impact of representing the raw data in the sequentially subsampled feature space and the effect of augmentation on overall performance. Also, Fig. 5 notifies the learning of the models on both the data sets. The accuracy results are provided for the combined features as a performance benchmark. Table VI indicates that \mathcal{M}_1 and \mathcal{M}_2 achieve satisfactory performance on both the data sets, and it specifies the ability of the subsampled feature space to boost the fault diagnosis results. Comparing the training and validation accuracies of both the models in both the data sets asserts the fact that the models are free from data overfitting, indicating the effectiveness of augmentation. Besides, model \mathcal{M}_2 achieves slightly superior performance in comparison to \mathcal{M}_1 on both the data sets. It is also observed that both the models that able to achieve decent performance with few epochs for each data set and the learning curve of \mathcal{M}_2 are comparatively smoother than \mathcal{M}_1 . It is worth noting from the results that adding highly discriminative augmented data increases data diversity, which helps improve the parameter training of the model and shows an increase in accuracy with basic sequential learning models.

D. Performance Analysis of Sequential Models

Fig. 6 shows the performance comparison of classifiers with TD, DFC, and combined features for both the data sets. As the model's structural complexity increases, a corresponding increase in accuracy is observed for both classifiers in both data sets with TD, DFC, and the combined features. It is observed that the combined features provide decent

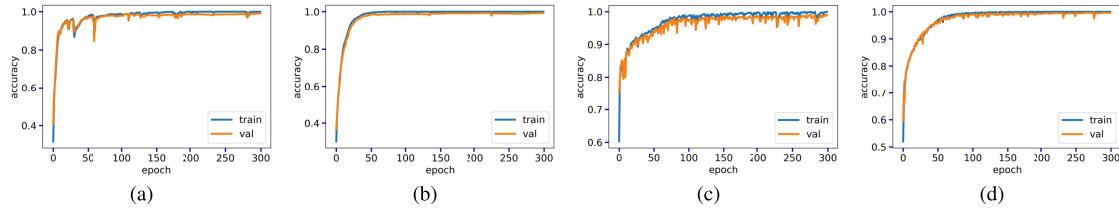


Fig. 5. Accuracy versus epoch graph for the models \mathcal{M}_1 and \mathcal{M}_2 . (a) DS-1 (\mathcal{M}_1). (b) DS-1 (\mathcal{M}_2). (c) DS-2 (\mathcal{M}_1). (d) DS-2 (\mathcal{M}_2).

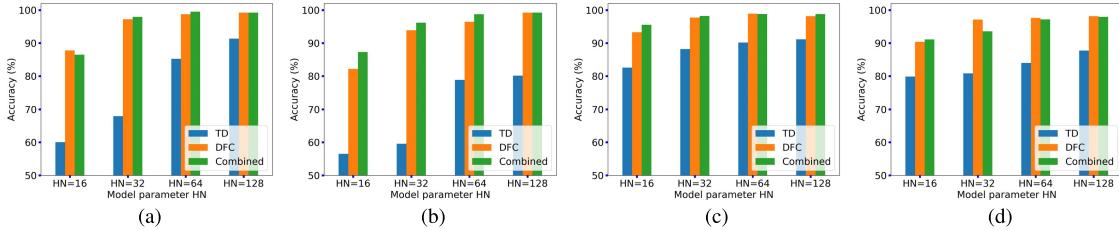


Fig. 6. Model comparison with different nodes. (a) DS-1 (\mathcal{M}_1). (b) DS-1 (\mathcal{M}_2). (c) DS-2 (\mathcal{M}_1). (d) DS-2 (\mathcal{M}_2).

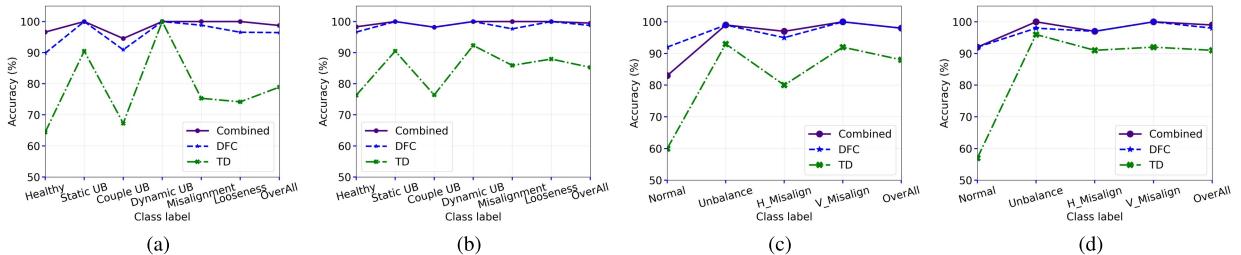


Fig. 7. Comparative analysis of TD, DFC, and combined features. (a) DS-1 (\mathcal{M}_1). (b) DS-1 (\mathcal{M}_2). (c) DS-2 (\mathcal{M}_1). (d) DS-2 (\mathcal{M}_2).

performance, and DFC itself is able to mark its importance in getting significant accuracy. In contrast, TD features are proven to be inferior in performance for DS-1, while, for DS-2, TD features are managed to give a minimum accuracy of 80%. The most important observation from the results of DS-1 is that both the sequential models with 64 HN are sufficient to give a decent accuracy closer to 100%. Interestingly, it shows that the \mathcal{M}_2 model with 64 HN outperformed \mathcal{M}_2 with 128 HN. Thus, it is concluded that the classifiers with 64 HN provide acceptable performance without compromising to higher complexity for negligible performance enhancement for DS-1.

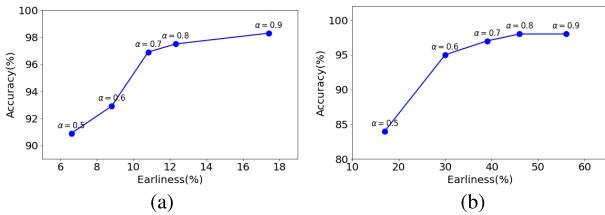
Compared with DS-1, DS-2 shows consistent improvement in accuracy as the model complexity increases. It is worth noticing that even the least complex model with 16 hidden layers with TD features could produce an accuracy of around 80% with both the classifiers. As in DS-1, the supremacy of \mathcal{M}_2 over \mathcal{M}_1 and the dominance combined features over TD or FD features persist with DS-2 also. Regarding the model selection, there is a slight performance improvement for the 64-HN model compared 32 HN model for both the classifiers, where the 128-HN model could give negligible performance enhancement for \mathcal{M}_1 only. Thus, we have chosen the model with 128 HN for comparative analysis.

Fig. 7 shows the importance of DFC in SRF diagnosis for both the classifiers on the data sets. The faultwise analysis

assists in understanding the significance of DFC in fault diagnosis. The noteworthy fluctuations have been observed for TD features with different types of faults. Both classifiers have displayed unacceptable performance for the healthy and couple UB class labels for DS-1. Also, similar observation has been noted for misalignment and looseness class using \mathcal{M}_1 model. However, DFC and the combined features provide acceptable performance irrespective of the fault type for both the models. Also, \mathcal{M}_2 performs slightly better than \mathcal{M}_1 in terms of accuracy. It is observed that the classwise behavior of DFC and combined features in DS-2 is almost similar to that of DS-1, producing no significant fluctuations between different fault types. However, TD features show unacceptable performance with the class normal in both the classifiers; at the same time, \mathcal{M}_2 is able to make up the performance degradation of the horizontal misalignment class. However, as with DS-1, \mathcal{M}_2 performs better than \mathcal{M}_1 for all the classes.

E. Performance Analysis of ECM

1) *Impact of α Parameters in ECM:* In the proposed model, parameter α decides the tradeoff between accuracy and earliness and holds a value between 0 and 1. Fig. 8 demonstrates the trend of $\alpha \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$, and it is observed that the accuracy and earliness substantially increase with the increase in the value of α . Moreover, it is also realized that the trend of α depends on the characteristic of the data set. For

Fig. 8. Effect of α . (a) DS-1. (b) DS-2.TABLE VII
PERFORMANCE OF ECM ON THE DS-1 DATA SET

		TD		DFC		Combined	
		Acc	Ear	Acc	Ear	Acc	Ear
PL-1	ECM- \mathcal{M}_1	78.40	56.80	96.40	12.70	98.20	20.20
	ECM- \mathcal{M}_2	86.30	80.70	98.50	12.80	99.50	14.90
PL-2	ECM- \mathcal{M}_1	77.30	60.02	93.30	15.60	96.23	24.34
	ECM- \mathcal{M}_2	85.40	82.80	96.01	16.08	97.91	18.19

TABLE VIII
PERFORMANCE OF ECM ON THE DS-2 DATA SET

		TD		DFC		Combined	
		Acc	Ear	Acc	Ear	Acc	Ear
PL-1	ECM- \mathcal{M}_1	88.53	86.24	97.01	63.07	97.55	62.66
	ECM- \mathcal{M}_2	90.29	90.83	97.51	54.21	98.32	55.68
PL-2	ECM- \mathcal{M}_1	88.07	86.17	94.65	63.87	95.60	63.17
	ECM- \mathcal{M}_2	88.93	90.98	93.06	54.79	96.96	56.66

example, change in α from 0.6 to 0.7 accuracy is improved by 4% and 2% for data sets DS-1 and DS-2, respectively, whereas the earliness is increased by 2% and 9%, respectively, for corresponding data sets. In this experimental work, the value of α is considered to be 0.9, as it exhibits adequate performance in terms of accuracy and earliness for both the data sets. However, any value of α can be selected as per the requirement of the system.

2) *Performance of ECM on Two Data Sets:* The performance analysis of the proposed ECM on the DS-1 and DS-2 data sets is provided in Tables VII and VIII, respectively. Along with that, the performance of the PL-2 data set has also been compared. In Table VII, it is observed that the ECM- \mathcal{M}_2 achieves the highest accuracy of 99.50% for combined features, whereas ECM- \mathcal{M}_1 with DFC features achieves the best earliness of 12.70%. As the TD signals have the least discriminative ability, it requires longer series to get an acceptable accuracy, while DFC and combined signals can make quick conclusions with less than 20% earliness without compromising accuracy. Similarly, model \mathcal{M}_2 is capable of capturing information in a short length of the series compared with model \mathcal{M}_1 . As a result, ECM- \mathcal{M}_2 outperformed the ECM- \mathcal{M}_1 in both the objectives, i.e., accuracy and earliness. A similar conclusion has been drawn for DS-2 also, as shown in Table VIII. However, due to more diversity in speed and load conditions, the DS-2 data set demands more length sequences to make the confident classification. Hence, the earliness value of DS-2 is around 60%. However, still, the overall accuracy of DS-2 is slightly better than the DS-1 in this framework.

Considering the performance of the model with the PL-2 data, it is evident from the first observation that there is a trend

in decreased accuracy and increased earliness value compared with PL-1, in both the data sets. Since data acquisition irregularity is created by randomly removing the samples, and the data missed timestamps are filled with mean values, in the TD data, this trend is not significant. That is, the TD data points in the subsampled space are almost similar to the PL-1 data because of the averaging on maximum bin operation. However, the DFC extracted values in the subsampled space of PL-2 are a bit different from the DFC data sequence generated in PL-1. This justifies the reduction in accuracy and needs a lengthier pattern (increased value of earliness) in both DFC and combined feature scenarios. However, it is worth noticing that the overall model accuracy and earliness of PL-2 can be maintained almost similar to PL-1 because of the training with the proposed augmentation method.

Furthermore, the confusion matrix has been shown in Fig. 9 to analyze the performance of ECM for classifying the healthy and faulty components using combined features. The proposed ECM demonstrates the good performance for healthy and faulty components except for Couple UB, as shown in Fig. 9(a) and (b). This effect has been observed because the properties of static UB and dynamic UB are almost similar and it creates adversity in classification. For the DS-2 data set also, ECM demonstrated good performance and is able to classify UB and vertical misalignment accurately, as shown in Fig. 9(c) and (d). ECM- \mathcal{M}_1 misclassified 17% normal samples as horizontal misalignment. However, ECM- \mathcal{M}_2 outperformed with 97% accuracy and misclassified only 3% normal sample as horizontal misalignment. Overall, the proposed ECM is capable of providing exceptional results for SRF classification.

3) *Comparative Analysis of ECM With Traditional Approach:* We have compared the proposed ECM with the TC approach for SRF, which is depicted in Fig. 10. The TC approach considers the full-length sequence of data for classification, whereas the proposed ECM is able to predict the fault with the partial sequence of data. ECM- \mathcal{M}_1 classified dynamic UB fault with 100% accuracy by utilizing only 5.19% data points, as shown in Fig. 10(a) and (b). Moreover, ECM- \mathcal{M}_1 classified couple UB using 25.19% of full-length data only compared with the TC approach with approximately 1.8% deficiency in the accuracy. It is also observed that the proposed ECM achieved better accuracy than the TC approach for a healthy condition, as shown in Fig. 10(a) and (c). Moreover, ECM- \mathcal{M}_2 achieved better earliness compared with ECM- \mathcal{M}_1 except for the dynamic UB class. For the DS-2 data set, ECM- \mathcal{M}_2 demonstrated similar or even higher accuracy for normal class, compared with DS-1, by utilizing 85.9% data points, as shown in Fig. 10(g) and (h). Both models utilized more lengthy sequences with DS-2 compared with DS-1 due to the diverse nature of the data set. It is also observed that the accuracy and earliness patterns vary for both the classifiers, \mathcal{M}_1 and \mathcal{M}_2 , with respect to the faults under consideration. However, the earliness and the accuracy of \mathcal{M}_2 are always found superior to that of \mathcal{M}_1 . Thus, based on the above observation, it is concluded that the proposed EC approach for SRF is highly efficient compared with the traditional one.

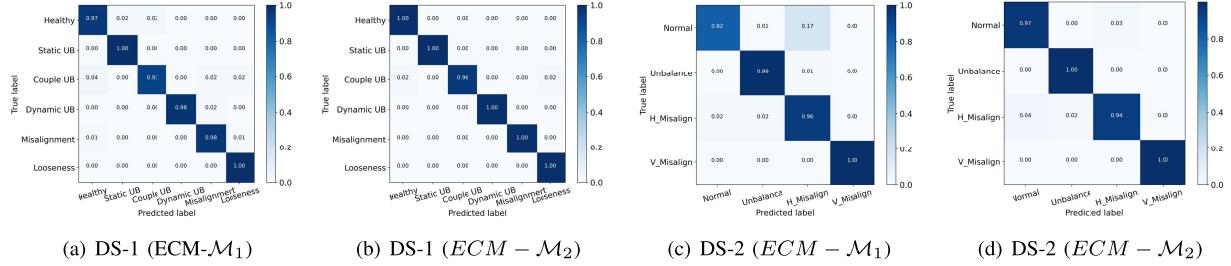


Fig. 9. Confusion matrix for models (ECM- \mathcal{M}_1 & ECM- \mathcal{M}_2) on data sets (DS-1 and DS-2). (a) DS-1 (ECM- \mathcal{M}_1). (b) DS-1 (ECM- \mathcal{M}_2). (c) DS-2 (ECM- \mathcal{M}_1). (d) DS-2 (ECM- \mathcal{M}_2).

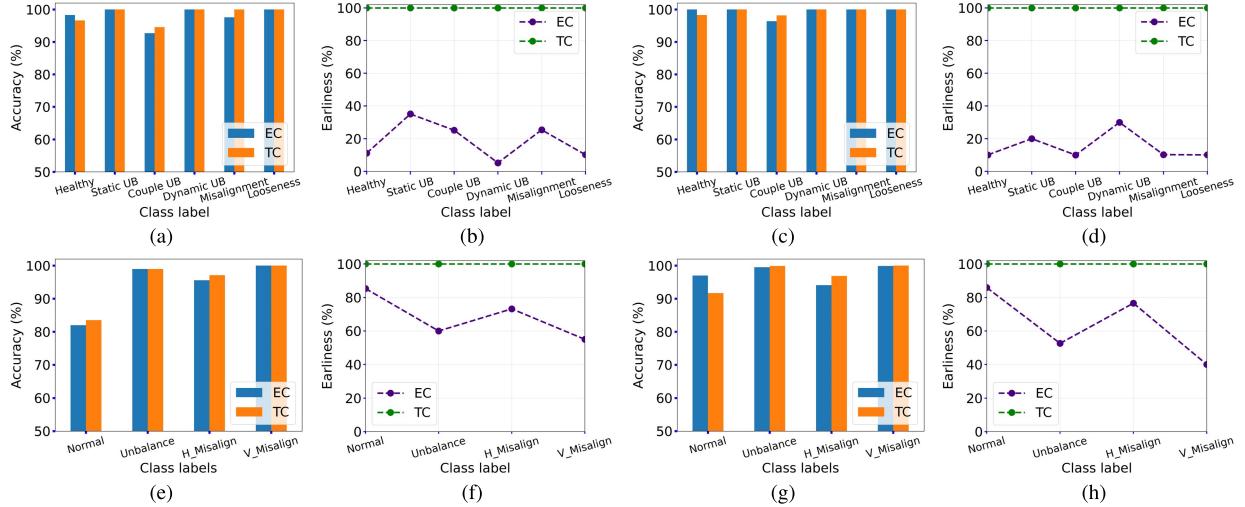


Fig. 10. Comparative analysis of the proposed model with the TC approach. (a) DS-1 (ECM- \mathcal{M}_1). (b) DS-1 (ECM- \mathcal{M}_1). (c) DS-1 (ECM- \mathcal{M}_2). (d) DS-1 (ECM- \mathcal{M}_2). (e) DS-2 (ECM- \mathcal{M}_1). (f) DS-2 (ECM- \mathcal{M}_1). (g) DS-2 (ECM- \mathcal{M}_2). (h) DS-2 (ECM- \mathcal{M}_2).

TABLE IX

COMPARISON WITH OTHER DEEP LEARNING-BASED APPROACHES

Work	Model	Data input	Accuracy (%) DS-1	Accuracy (%) DS-2	Early Classification
[32]	CNN	Raw data	90.02	95.60	No
[33]	CNN	Raw data	95.86	96.23	No
[34]	DBN	Raw data	92.76	93.50	No
[35]	LSTM	Raw data	67.98	80.15	No
[35]	LSTM	Features	89.50	92.00	No
ECM- \mathcal{M}_1	LSTM	Features	98.20	97.55	Yes
ECM- \mathcal{M}_2	GRU	Features	99.50	98.32	Yes

4) *Comparison With State-of-the-Art Deep Learning Approaches:* Table IX demonstrates the comparison of the proposed model with the excellently performing state-of-the-art deep learning models in rotating machinery fault diagnosis. Since most deep learning studies in rotating machinery fault analysis deal with CNNs [4], we have selected two recent CNN models [32], [33], which extracts features automatically from the 1-D raw vibration and gives special attention in dealing with sensor signals. It is observed that the method of [33] performs better and achieved an accuracy around 95% or slightly more compared with the method of [32] because of its feature extraction (by deeper CNN architecture) and feature fusion strategy. The performance of CNN models on DS-2 is always better than DS-1 because of the high acquisition frequency of DS-2 raw vibration signals.

The sequential model's performance, such as LSTM [35] with raw vibration input, is not in an acceptable range. We have tested the model [35] on raw data and extracted frequency features to analyze the performance of the LSTM for comparison. LSTM with frequency features shows an improved performance compared with the raw data input as expected; also, it could not meet the other compared models' accuracy range. The deep belief network (DBN) model [34], which used independent DBNs on each sensor data with information fusion, worked well with our data sets and achieved an accuracy around 93%, showing the performance between CNNs and LSTM. Our proposed method reached accuracy around 99% or more on full-length data with the carefully selected DFC features and augmentation compared with these methods. Moreover, it is worth noticing that, with EC strategy, 96% or more accuracy was achieved with only around 25% of data length for DS-1 and that with 60% of data length for DS-2 by the proposed model.

IV. CONCLUSION

This article presents an SRF diagnosis framework with an EC strategy. A novel SRF data set has been created and tested, simulating the real plant industrial environment. The proposed data subsampling method, which incorporated SRF specific DFC and TD features, facilitated the model

to perform well irrespective of the industrial data acquisition issues. The soft-DTW-based augmentation enriched the subsampled input training data set and eliminated the class imbalance issue. The FIC-based weighing scheme used in augmentation provided more heterogeneity and discriminative features to the synthesized samples. Finally, the proposed EC model made the framework capable of detecting various faults based on a partial observed sequence in a real-time environment. The paradigm's advantage is that it provides decent performance at an early stage with a minimum input sequence. As a result, it provides a significant amount of time for maintenance activity than the traditional fault diagnosis method. The decision policy is the heart of the EC paradigm and plays a crucial role in providing reliable performance by controlling the earliness parameter. Hence, the experimental results demonstrated that the proposed framework achieved decent outcomes in both objectives: accuracy and earliness. Our future work will focus on exploiting TS characteristics and utilizing DFC and optimization techniques in SRF to achieve more reliable fault prediction.

ACKNOWLEDGMENT

The authors would like to thank IMAGENOUS Engineering Pvt. Ltd., Vadodara, India, and Meggitt India Pvt. Ltd., Bengaluru, India, for providing the experimental environment.

REFERENCES

- [1] J. Qin, Y. Liu, and R. Grosvenor, "A categorical framework of manufacturing for industry 4.0 and beyond," *Procedia CIRP*, vol. 52, pp. 173–178, Sep. 2016.
- [2] G.-Y. Lee *et al.*, "Machine health management in smart factory: A review," *J. Mech. Sci. Technol.*, vol. 32, no. 3, pp. 987–1009, 2018.
- [3] P. Chen, *Foundation and Application of Condition Diagnosis Technology for Rotating Machinery*. Nagoya, Japan: Sankeisha Press, 2009, pp. 49–51.
- [4] A. G. Nath, S. S. Udmale, and S. K. Singh, "Role of artificial intelligence in rotor fault diagnosis: A comprehensive review," *Artif. Intell. Rev.*, pp. 1–60, Sep. 2020. [Online]. Available: <https://link.springer.com/article/10.1007%2Fs10462-020-09910-w>
- [5] H. Xue, H. Wang, P. Chen, K. Li, and L. Song, "Automatic diagnosis method for structural fault of rotating machinery based on distinctive frequency components and support vector machines under varied operating conditions," *Neurocomputing*, vol. 116, pp. 326–335, Sep. 2013.
- [6] H. Srinivas, K. Srinivasan, and K. Umesh, "Application of artificial neural network and wavelet transform for vibration analysis of combined faults of unbalances and shaft bow," *Adv. Theor. Appl. Mech.*, vol. 3, no. 4, pp. 159–176, 2010.
- [7] A. El-Shafei, T. A. Hassan, A. K. Soliman, Y. Zeyada, and N. Rieger, "Neural network and fuzzy logic diagnostics of 1X faults in rotating machinery," *J. Eng. Gas Turbines Power*, vol. 129, no. 3, pp. 703–710, 2007.
- [8] C.-S. Chen and J.-S. Chen, "Rotor fault diagnosis system based on sGA-based individual neural networks," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 10822–10830, Sep. 2011.
- [9] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network," *Mech. Syst. Signal Process.*, vol. 27, pp. 696–711, Feb. 2012.
- [10] H. K. Srinivas, K. S. Srinivasan, and K. N. Umesh, "Role of an artificial neural network and a wavelet transform for condition monitoring of the combined faults of unbalance and cracked rotors," *Int. J. Acoust. Vibrat.*, vol. 15, no. 3, p. 121, 2010.
- [11] S. Cacciola, I. M. Agud, and C. L. Bottasso, "Detection of rotor imbalance, including root cause, severity and location," *J. Phys., Conf. Ser.*, vol. 753, no. 7, Sep. 2016, Art. no. 072003.
- [12] W. Fengqi and G. Meng, "Compound rub malfunctions feature extraction based on full-spectrum cascade analysis and SVM," *Mech. Syst. Signal Process.*, vol. 20, no. 8, pp. 2007–2021, Nov. 2006.
- [13] X. Tang, L. Zhuang, J. Cai, and C. Li, "Multi-fault classification based on support vector machine trained by chaos particle swarm optimization," *Knowl.-Based Syst.*, vol. 23, no. 5, pp. 486–490, Jul. 2010.
- [14] L. M. R. Baccarini, V. V. Rocha e Silva, B. R. de Menezes, and W. M. Caminhas, "SVM practical industrial application for mechanical faults diagnostic," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6980–6984, Jun. 2011.
- [15] H. Xue, H. Wang, L. Song, and P. Chen, "Structural fault diagnosis of rotating machinery based on distinctive frequency components and support vector machines," in *Proc. Int. Conf. Intell. Comput.* Berlin, Germany: Springer, 2011, pp. 341–348.
- [16] B. Pang, G. Tang, C. Zhou, and T. Tian, "Rotor fault diagnosis based on characteristic frequency band energy entropy and support vector machine," *Entropy*, vol. 20, no. 12, p. 932, Dec. 2018.
- [17] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
- [18] Y. Yao, Y. Li, P. Zhang, B. Xie, and L. Xia, "Data fusion methods for convolutional neural network based on self-sensing motor drive system," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, vol. 1, Oct. 2018, pp. 5371–5376.
- [19] Z. Li, Y. Wang, and K. Wang, "A deep learning driven method for fault classification and degradation assessment in mechanical equipment," *Comput. Ind.*, vol. 104, pp. 1–10, Jan. 2019.
- [20] L. Bo, J. Ying, C. Y. Ping, and F. Xiaolong, "The study on rotating machinery fault diagnosis based on deep neural networks," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City (ICITBS)*, Dec. 2016, pp. 125–129.
- [21] T. Hu, T. Tang, and M. Chen, "Data simulation by Resampling—A practical data augmentation algorithm for periodical signal analysis-based fault diagnosis," *IEEE Access*, vol. 7, pp. 125133–125145, 2019.
- [22] A. L. Guennec, S. Malinowski, and R. Tavenard, *Data Augmentation for Time Series Classification Using Convolutional Neural Networks*. Accessed: 2016. [Online]. Available: <https://halshs.archives-ouvertes.fr/halshs-01357973/>
- [23] T. Taewoong Um *et al.*, "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," 2017, *arXiv:1706.00527*. [Online]. Available: <http://arxiv.org/abs/1706.00527>
- [24] P. Cao *et al.*, "A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101675.
- [25] A. Marie Delaney, E. Brophy, and T. E. Ward, "Synthesis of realistic ECG using generative adversarial networks," 2019, *arXiv:1909.09150*. [Online]. Available: <http://arxiv.org/abs/1909.09150>
- [26] M. M. Krell, A. Seeland, and S. K. Kim, "Data augmentation for brain-computer interfaces: Analysis on event-related potentials data," 2018, *arXiv:1801.02730*. [Online]. Available: <http://arxiv.org/abs/1801.02730>
- [27] J. Dahmen and D. Cook, "SynSys: A synthetic data generation system for healthcare applications," *Sensors*, vol. 19, no. 5, p. 1181, Mar. 2019.
- [28] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, Mar. 2011.
- [29] G. Forestier, F. Petitjean, H. A. Dau, G. I. Webb, and E. Keogh, "Generating synthetic time series to augment sparse datasets," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 865–870.
- [30] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.
- [31] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Müller, "Data augmentation using synthetic data for time series classification with deep residual networks," 2018, *arXiv:1808.02455*. [Online]. Available: <http://arxiv.org/abs/1808.02455>
- [32] D. Kolar, D. Lisjak, M. Paják, and D. Pavković, "Fault diagnosis of rotary machines using deep convolutional neural network with wide three axis vibration signal input," *Sensors*, vol. 20, no. 14, p. 4017, Jul. 2020.
- [33] W. Zhao, C. Hua, D. Wang, and D. Dong, "Fault diagnosis of shaft misalignment and crack in rotor system based on MI-CNN," in *Damage Assessment of Structures (Lecture Notes in Mechanical Engineering)*. Singapore: Springer, Jul. 2019, pp. 529–540.

- [34] J. Yan, Y. Hu, and C. Guo, "Rotor unbalance fault diagnosis using DBN based on multi-source heterogeneous information fusion," *Procedia Manuf.*, vol. 35, pp. 1184–1189, Jan. 2019.
- [35] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on long short-term memory networks," *Renew. Energy*, vol. 133, pp. 422–432, Apr. 2019.
- [36] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.
- [37] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.
- [38] S. Ma, W. Cai, W. Liu, Z. Shang, and G. Liu, "A lighted deep convolutional neural network based fault diagnosis of rotating machinery," *Sensors*, vol. 19, no. 10, p. 2381, May 2019.
- [39] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.
- [40] Z. Meng, X. Guo, Z. Pan, D. Sun, and S. Liu, "Data segmentation and augmentation methods based on raw data using deep neural networks approach for rotating machinery fault diagnosis," *IEEE Access*, vol. 7, pp. 79510–79522, 2019.
- [41] S. Heo and J. H. Lee, "Fault detection and classification using artificial neural networks," *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 470–475, 2018.
- [42] X. Xu, H. Zheng, Z. Guo, X. Wu, and Z. Zheng, "SDD-CNN: Small data-driven convolution neural networks for subtle roller defect inspection," *Appl. Sci.*, vol. 9, no. 7, p. 1364, Mar. 2019.
- [43] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," *Comput. Ind.*, vol. 106, pp. 85–93, Apr. 2019.
- [44] A. Le Guennec, S. Malinowski, and R. Tavenard, "Data augmentation for time series classification using convolutional neural networks," in *Proc. ECML/PKDD Workshop Adv. Anal. Learn. Temporal Data*, Riva Del Garda, Italy, Sep. 2016.
- [45] Z. Xing, J. Pei, and P. S. Yu, "Early classification on time series," *Knowl. Inf. Syst.*, vol. 31, no. 1, pp. 105–127, Apr. 2011.
- [46] U. Mori, A. Mendiburu, E. Keogh, and J. A. Lozano, "Reliable early classification of time series based on discriminating the classes over time," *Data Mining Knowl. Discovery*, vol. 31, no. 1, pp. 233–263, Jan. 2017.
- [47] A. Gupta, H. P. Gupta, B. Biswas, and T. Dutta, "Approaches and applications of early classification of time series: A review," *IEEE Trans. Artif. Intell.*, to be published.
- [48] A. Bregón, M. A. Simón, J. J. Rodríguez, C. Alonso, B. Pulido, and I. Moro, "Early fault classification in dynamic systems using case-based reasoning," in *Proc. Conf. Spanish Assoc. Artif. Intell.* Berlin, Germany: Springer, 2005, pp. 211–220.
- [49] Z. Xing, J. Pei, P. S. Yu, and K. Wang, "Extracting interpretable features for early classification on time series," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2011, pp. 247–258.
- [50] J. Lv, X. Hu, L. Li, and P. Li, "An effective confidence-based early classification of time series," *IEEE Access*, vol. 7, pp. 96113–96124, 2019.
- [51] A. Sharma and S. K. Singh, "Early classification of multivariate data by learning optimal decision rules," in *Multimedia Tools and Applications*. Springer, Aug. 2020, doi: [10.1007/s11042-020-09366-8](https://doi.org/10.1007/s11042-020-09366-8).
- [52] M. Rußwurm, R. Tavenard, S. Lefèvre, and M. Körner, "Early classification for agricultural monitoring from satellite time series," 2019, *arXiv:1908.10283*. [Online]. Available: <http://arxiv.org/abs/1908.10283>
- [53] M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic, "Extraction of interpretable multivariate patterns for early diagnostics," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 201–210.
- [54] A. Sharma and S. Kumar Singh, "A novel approach for early malware detection," *Trans. Emerg. Telecommun. Technol.*, Apr. 2020, Art. no. e3968. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.3968>
- [55] A. Sharma, S. K. Singh, S. S. Udmale, A. K. Singh, and R. Singh, "Early transportation mode detection using smartphone sensing data," *IEEE Sensors J.*, early access, Jul. 14, 2020, doi: [10.1109/JSEN.2020.3009312](https://doi.org/10.1109/JSEN.2020.3009312).
- [56] A. Gupta, H. P. Gupta, B. Biswas, and T. Dutta, "An early classification approach for multivariate time series of on-vehicle sensors in transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5316–5327, Dec. 2020.
- [57] *VibroSight Vibration Monitoring Software*. Accessed: May 9, 2020. [Online]. Available: https://meggittsensing.com/wp-content/uploads/VibroSight_software_releas%20se_notes_4.1.0-en.pdf
- [58] M. Cuturi and M. Blondel, "Soft-DTW: A differentiable loss function for time-series," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 894–903.
- [59] Z. Guan, Z. Liao, K. Li, and P. Chen, "A precise diagnosis method of structural faults of rotating machinery based on combination of empirical mode decomposition, sample entropy, and deep belief network," *Sensors*, vol. 19, no. 3, p. 591, Jan. 2019.
- [60] M. Shah, J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, "Learning DTW-shapelets for time-series classification," in *Proc. 3rd IKDD Conf. Data Sci. (CODS)*, 2016, pp. 1–8.
- [61] M. Cuturi, J.-P. Vert, O. Birkenes, and T. Matsui, "A kernel for time series based on global alignments," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2007, p. 413.
- [62] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [63] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [64] *MaFaulDa Machinery Fault Database*. Accessed: May 11, 2020. [Online]. Available: http://www02.smt.ufrj.br/~offshore/mfs/page_01.html



Aneesh G. Nath (Student Member, IEEE) received the M.Tech. degree in computer engineering from the National Institute of Technology Karnataka (NITK), in 2014. He is currently pursuing the Ph.D. degree in computer science and engineering with IIT (BHU) Varanasi, Varanasi, India.

He is currently an Associate Professor with the Department of Computer Science and Engineering, TKM College of Engineering, Kollam, Kerala. His current research interests include machine learning, data science, signal processing, and optimization.



Anshul Sharma (Student Member, IEEE) received the M.Tech. degree in computer science from Aligarh Muslim University (AMU), Uttar Pradesh, India, in 2011. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi, India.

His research interests include machine learning, data analysis, and pattern recognition.

Mr. Sharma is a Student Member of ACM.



Sandeep S. Udmale (Member, IEEE) received the Ph.D. degree in computer science and engineering from IIT (BHU) Varanasi, Varanasi, India, in 2019.

He is currently an Assistant Professor with the Department of Computer Engineering and Information Technology, Veermata Jijabai Technological Institute (VJTI), Mumbai, India. His current research interests include machine learning, data science, and optimization, and pattern analysis.

Dr. Udmale is a member of the ACM and the Computer Society of India.



Sanjay Kumar Singh (Senior Member, IEEE) is currently a Professor with the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi, India. He has authored or coauthored more than 150 national and international journal publications, book chapters, and conference articles. He has four patents filed to his credit. His current research interests include machine learning, deep learning, computer vision, medical image analysis, pattern recognition, and biometrics.

Prof. Singh is a Senior Member of ACM and the Computer Society of India. He is also a guest editorial board member and a reviewer for many international journals of repute.