IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 6, 2021, accepted February 4, 2021, date of publication March 2, 2021, date of current version March 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3063676

An Automated Sensor Fusion Approach for the RUL Prediction of Electromagnetic Pumps

UGOCHUKWU EJIKE AKPUDO, (Graduate Student Member, IEEE), AND HUR JANG-WOOK
Department of Mechanical Engineering (Department of Aeronautics, Mechanical and Electronic Convergence Engineering), Kumoh National Institute of Technology, Gyeongbuk 39177, South Korea

Corresponding author: Jang-Wook Hur (hhjw88@kumoh.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MIST) under Grant 2019R1/1A3A01063935.

ABSTRACT The remaining useful life (RUL) prediction of industrial cyber-physical system components demands the use of reliable prognostics parameters and frameworks. Against the traditional use of a single measure of degradation, data from multiple sensors provide abundant characteristic information for modeling, assessing, and extracting useful parameters via appropriate signal processing and sensor fusion methods. This study introduces a multi-sensor prognostics approach which merges highly prognosible statistical features from vibrational and pressure sensor measurements after a multi-level wavelet decomposition of the signals. The prognostic algorithm presented in this work for solenoid pump RUL prediction is a multi-objective genetic algorithm-optimized long short-term memory (MOGA-LSTM) which accepts the fused sensor features as input and returns the RUL of the pump as output. The framework was tested on a run-to-failure experiment on a VSC63A5 Solenoid pump following a significant pump malfunction caused by a clogged suction filter after the test. Using standard prognostic performance evaluation metrics, the performance of the prognostics framework was compared with other reliable state-of-the-art methods with a remarkable comparative advantage in addition to better automation potentials for real-time condition monitoring and RUL prediction.

INDEX TERMS Sensor fusion, remaining useful life prediction, electromagnetic pumps, MOGA-LSTM, wavelet decomposition.

I. INTRODUCTION

Reliability studies on industrial cyber-physical systems (ICPSs) have become one of the central research focus in academia owing to the fast-growing industry 4.0 revolution. Being a multidisciplinary topic, studies on ICPSs are currently receiving vast patronage and research interests covering vast modules ranging from industrial design technologies, manufacturing, smart control, robotics, engineering, etc.; nevertheless, regardless of the domain of interest, system reliability, safety, and maintainability are some of the key concerns which have motivated the on-going shift from traditional corrective (and routine-based maintenance) to the more effective condition-based maintenance(CBM) with prognostics and health management (PHM) at its core [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Edith C.-H. Ngai.

The economic, ethical and environmental impacts of unexpected equipment failure can be overwhelming and often result to unplanned maintenance (or repair) while production activities are delayed, or sometimes brought to a halt. Although deliberate attempts are being made to prevent occurrence of these undesired outcomes, absolute failure prevention and control seems unachievable with the major causes emanating from uncertainties, environmental factors, human errors, and equipment fatigue. These uncertainties and many past recorded accidents only validate the need for reliable process monitoring, fault diagnosis and prognostics system design schemes that can be applied to large-scale processes [2], [3].

PHM as a predictive maintenance approach enables real-time health assessment and future state estimation based on available information from sensing technologies, physics-of-failure(POF), modern statistics and reliability

engineering concepts [4]. Traditionally, the model-based approaches rely on a system model with known structures and parameters, and are well developed in the framework of state observers; despite that, their performance is based on costly assumptions and take no recognition of uncertainties and unexpected nonlinearities. On the other hand, data-driven methods are structured on transforming monitoring/historical data into befitting models for better health assessments and RUL prediction without knowledge of the system's failure mechanism [4], [5]. Current research trends aim at hybrid methods (a combination of model-based and data-driven methods); regardless, whether model-based, data-driven or hybrid, prognostics algorithms could be of three types: reliability-based (where historical failure time data is utilized), stressor-based (where environmental conditions are utilized) or degradation-based (where measured/inferred system condition are utilized) [6]. On a different note, the use of multiple degradation measurements for prognostics is usually more reliable and with appropriate fusion techniques, these measurements– extracted features and/or sensor data can be combined to develop a single health indicator(HI) using dimensionality reduction techniques like principal component analysis(PCA), independent component analysis(ICA), self organising maps(SOM), etc. [7]–[9]; however, not all measurements are prognosible measurements. The choice of prognosible measurements has been one of the open areas of on-going research studies; nevertheless, characteristics such as trendability, monotonicity and prognosibility provide reliable paradigm for selection of prognosible measurements [6], [10]. By computing and assessing these metrics on features, the high-ranking features can be selected to reduce computational load on the optimization scheme while ensuring overall computational advantage and accurate prognostics performance by a bio-inspired predictor like the LSTM– a popular recurrent neural network (RNN) with Bayesian architecture for time-series forecasting.

In compensation for the growing complexity of recent times technological advancement, state-of-the-art prognostics schemes demand the use of highly reliable predictive models of which, deep learning (DL) methods provide strong paradigms. Particularly, RNNs have strong predictive capabilities beyond PHM to stock forecasting [31], weather forecasting [32], text generation [33], etc. due to its feedback–enabled architecture and its use of gates for long–range memorization for accurate posterior estimation; nonetheless, deliberate efforts need to be made for automatically identifying its parameters based on available data characteristics, failure mechanisms and operating conditions. Although the LSTM has these remarkable efficiencies, there are still some drawbacks associated with its optimal use including a highly complex computational process, hyper-parameter dependence, and over-fitting tendencies. Consequently, this study aims at developing a reliable automatic parameter selection framework for solving this problem by capturing prognosible characteristics of a population of measured data (sensor data and/or features). The Genetic algorithm (GA)— a global meta-heuristic global search algorithm is efficient for this purpose (against the limitations of other evolutionary and /or meta-heuristic search algorithms like the particle swarm optimization, ant colony optimization, firefly algorithm, etc.) [14]–[16].

The rest of the paper is structured thus: Section II presents the motivation and related works to solenoid pump prognostics while Section III describes the theoretical background of multi-sensor data fusion, ICA- based HI construction, and GA. Section IV presents the proposed multi-objective GA-LSTM (MOGA-LSTM) prognostics model while the experimental validation is presented in section V. Section VI concludes the paper.

II. MOTIVATION AND RELATED WORKS

Electromagnetic pumps, like other types of pumps are usually operated for long periods which expose them to diverse failure modes from sources ranging from unfavorable environmental conditions, fluid contamination, electrical and/or mechanical stress, fatigue, uncertainty, filter clogging, overuse/underuse, etc. [26], [34]. It is on these grounds that the need for real-time condition monitoring (and assessment) becomes crucial for accurate predictive maintenance of these pumps. Interestingly, solenoid pump and valves are identical in their mode of operation— solenoid magnetization by electrical current through the coil which causes the electromagnetic core to move against a spring to slide a diaphragm into the discharge position. Although solenoid valves have witnessed a reasonable number of research studies on **FDI** and prognostics, unfortunately, only a few exist for solenoid pumps out of which, a majority cover only the design and fault diagnosis [5], [26], [34].

A. SIGNAL DE-NOISING

Raw vibrational and pressure sensor data are non-stationary in nature, contain background noise, and usually provide little information for direct use for most prognostics problems, including the one presented in this work [17]. This has essentially prompted the need for several signal processing techniques for reliable feature extraction. Particularly, choosing a safe threshold for signal de-noising depends on the targeted system's dynamics, engineer/analyst's level of expertise, and/or familiarity in the domain; nevertheless, by separating useful signals from background noise, a more reliable health assessment for accurate prognostics can be achieved [17], [18].

Several approaches to signal de-noising and decomposition have been presented including but not limited to the sparsity–based models, empirical mode decomposition (EMD) whose functionality is based on the Hilbert Huang transform (HHT), discrete wavelet transform (DWT), Bayesian-filter-based methods, etc. [19], [20]. Of these techniques, the EMD and DWT are the most popular due to their robustness for noise reduction/elimination; however, studies show that the wavelet decomposition has better de-noising capabilities and this has motivated its usage in this study [20].

B. MULTI-SOURCE DATA FUSION

Reliable PHM schemes usually demand the use of multiple sensors(data sources) or extracted features for providing abundant useful information for RUL estimation; however, some of these data sources may be insignificant and this creates a need– to choose from a multitude of sensors, which one(s) are useful and more importantly, to combine these useful sensor data/features in the most comprehensive and efficient manner [21]. Although the insignificance of a sensor(s) or extracted feature for prognostics does not imply absolute insignificance for other purposes, it is important to ensure that redundant sensors/features are eliminated for accurate and cost–efficient prognostics results. Furthermore, the abundant partial information contained in these sensor data/features can be fused using appropriate fusion techniques to create a comprehensive health indicator (HI) for easier system representation and degradation assessment [7], [11], [22].

Although every fusion technique have their respective pros and cons, the choice of a technique usually depend on the kind of problem at hand. For instance, PCA's performance is greatly limited on normally distributed data while LLE is affected by the choice of nearest neighbours. On the other hand, the ICA is especially robust in cases of non–Gaussian input distributions that are statistically independent; just as the case study presented is this study [8], [23]. To ensure dependable safety–critical systems using multiple on–chip embedded instruments (EIs), Bagheriye *et al.* [23] compared the ICA and the auto-encoder (AE) for synchronous data capture of different IJTAG compatible EIs and a ML-based system–level model for determining the end of life (EOL). Results show that the ICA–based EI fusion approach is comparatively more effective in capturing the latent variables as new representation of inputs for EOL prognostics. For optimal human brain diagnosis, the authors in [24] employed the ICA for merging the functional magnetic resonance images (fMRI), structural magnetic resonance images (sMRI) and electroencephalogram (EEG) signals with results validating the usefulness of joint brain imaging data for revealing unique information that cannot be evaluated from a single modality. For pump fault diagnostics, J. Weidong [25] proposed a novel ICA based compound neural network– FastICA whose functionality is based on feature extraction of multi-channel vibration measurements.

C. CHOICE OF PROGNOSTICS ALGORITHM AND PARAMETER OPTIMIZATION

The choice of prognostics algorithm(s) is one of the key factors worth considering for every PHM scheme [6], [12]. With the dominance and vast applicability of artificial intelligence (AI), machine learning (ML), and DL algorithms provide reliable learning and predictive capabilities against purely mathematical/probabilistic models whose effectiveness are limited on complex few data measurements; however, for optimal usage, deliberate attempts are required for assessing, choosing and validating the ML/DL algorithm for prognostics and RUL prediction [12]. Quite remarkably, the choice of these prognostics algorithms depend on their predictive robustness which rely hugely on their individual constituent parameters. The deeper the architecture of the predictor, the more effective it is for accurate modelling for posterior estimation; however, this often leads to an increase in the number of parameters required for optimum prognostics results. Consequently, the need for optimal parameter selection arises [10], [13]. Traditionally, such selection depend on the analyst's intuition. This is unreliable considering human errors. In contrast, meta-heuristic algorithms provide reliable (and automated) solutions to optimal parameter selection [27]–[30].

Meta-heuristic algorithms like the GA, particle swarm optimization, Ant Colony Optimization, fruit fly optimization, and a host of hybrid methods have been reported (and compared) in many studies for several optimization problems including but not limited to feature selection [26], fault detection and identification [15], structural optimization of predictive algorithms [27], [28], and functional process/system control [29]. Particularly for multi-objective probelms, the MOGA has proven its superiority over the multi-objective PSO (MOPSO) as verified in [30] and has also shown reliable efficiencies for global parameterization of time-series predictors [27], [28]. Consequently, this paper makes the following contributions:

- Proposal of an ICA–based HI construction of highly prognosible features from the de-noised signals for accurate RUL prediction. This was achieved by evaluating the statistical features extracted from each level of DWT decomposition stage based on monotonicity, prognosibility, and trendability scores.
- Design of a computationally efficient (and automated) DL-based prognostics scheme which combines vibration and pressure measurements for capturing diversified solenoid pump dynamics for accurate solenoid pump RUL prediction
- Proposal and validation of the MOGA-LSTM predictor which automatically selects the appropriate LSTM parameters.

III. BACKGROUND OF STUDY

This section provides the theoretical background of the key components which the proposed study relies on. This include DWT de-noising/decomposition, ICA–based HI construction, and the proposed MOGA-LSTM RUL predictor

A. DWT DE-NOISING/DECOMPOSITION

The wavelet transform represents a signal in the form of wavelet series– a representation of a square–integrable (real or complex–valued) function by a certain orthogonal series generated by a wavelet [17], [18]. Similar to discrtete Fourier transform (DFT) and short-time Fourier transform (STFT), wavelet transform can be viewed as the projection of a signal into a set of basis functions called wavelets. Such basis functions offer localization in the frequency domain. The major difference is: Fourier transform decomposes the signal

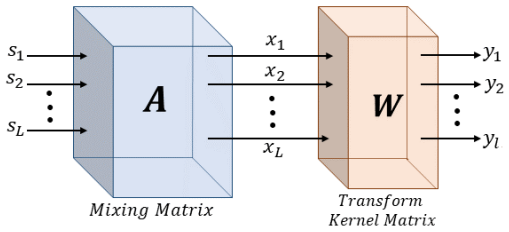


FIGURE 1. Linear model of FastICA.

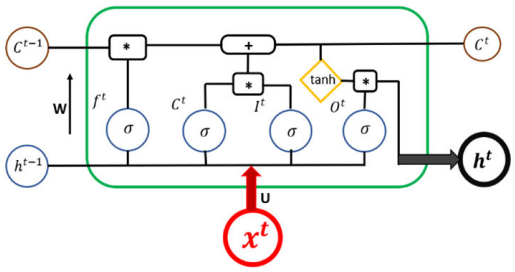


FIGURE 2. An LSTM block.

into sines and cosines whereas the wavelet transform uses functions (wavelets) that are localized in both the real and Fourier space, thereby providing more intuitive information in a signal.

Generally, a wavelet transform is defined as:

$$WT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi(t)^* \left(\frac{t-b}{a}\right) dt \tag{1}$$

where $\psi(t)^*$ is the complex conjugate of the single function $\psi_{a,b}(t)$ obtained by translation and dilation of the mother wavelet $\psi(t)$ as shown in (2).

$$\psi_{a,b}(\mathrm{t}) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a}\right) \tag{2}$$

where a is the scaling parameter ($a > 0$) is the localization parameter ($b \in R$) and $\frac{1}{\sqrt{a}}$ is a normalization factor for energy preservation.

Realistically, the Wavelet transform is an infinite set of various transforms, depending on the merit function used for its computation. Based on this, one can use *orthogonal* wavelets for DWT development and *non − orthogonal* wavelets for continuous wavelet transform (CWT) development. For decomposition/de-noising purposes (as intended in this work), the DWT is very effective and has been employed for many diagnostics and prognostics purposes.

B. ICA–BASED HI CONSTRUCTION

Originally, ICA was a tool designed for solving blind source separation problems in image and audio processing with a primary goal– to extract a set of statistically independent components from the observed signal; however, its use for dimensionality reduction (feature fusion) purposes has motivated the development of the FastICA [35]. The mutual information between multiple features is usually hidden in the high–order statistic characteristics, and FastICA is effective for reducing the high–order correlation while maintaining the mutual independence between them.

As illustrated in Fig. 4, from an observed vector $\boldsymbol{x} = [x_1, x_2, \cdots, x_L)]^T$ composed of linear combinations of sources s_i, the matrix A- a *transform kernel* calculated by the FastICA represents the mixed mode of independent non-Gaussian sources $\boldsymbol{s} = [s_1, s_2, \cdots, s_L]^T$ with the components s_i's mutually independent such that:

$$\boldsymbol{x} = A\boldsymbol{s} \tag{3}$$

The independent signals extracted from $\boldsymbol{x}$ by the FastICA are $\boldsymbol{y} = [y_1, y_2, \cdots, y_l]^T$ where $l \leq L$. Basically, the goal of the FastICA technology is to discover a linear mapping W such that the output signal $\boldsymbol{y}$ can approximate the independent source signals $s_1, s_2, \cdots, s_L$ with the minimum reconstruction error.

For HI construction, the goal is to generate a single comprehensive feature index y (where $l = 1$) by fusing the set of highly prognosible features.

C. LSTM NEURAL NETWORK

As shown in Fig. 2, a conventional LSTM block contains memory cell and three multiplicative gates– input gate, forget gate, and output gate whose functions are for regulating sequential information transfer for acquiring more accurate long–range dependencies. The recurrent connections between the cells and each gate provide steady operations for the cells. While each cell does the job of transporting state values over time steps, each gate conducts the write, read, and reset operations for the cells [40].

At each time step, the first layer considers a previous output and current input for determining what information to transfer from the previous state where the input value can only be preserved in the state of the cell if the input gate permits it. Eqs. (4), (5) and (6) present the definitions for the forget, input and output gates respectively.

$$f^t = \sigma\left(u^f x^t + w^f h^{t-1} + b^f\right) \tag{4}$$

$$I^t = \sigma\left(u^i x^t + w^i h^{t-1} + b^i\right) \tag{5}$$

$$O^t = \sigma\left(u^o x^t + w^o h^{t-1}\right) + b^o\right) \tag{6}$$

where u, w, and b are the weight matrices and bias for the respective gates while σ is a sigmoid activation function.

The current cell memory C^t is generated from C^{t*} and an element-wise multiplication of the previous memory C^{t-1} and f^t where C^{t*} is is the weighted sum of the prior output h^{t-t} and bias. This is expressed in (7) while C^t is expressed in (8)

$$C^{t*} = \tanh\left(w^c h^{t-1} + u^c x^t + b^c\right) \tag{7}$$

$$C^t = f^t \odot C^{t-1} + I^t \odot C^{t*}\right) \tag{8}$$

Finally, the LSTM output h^t is a dot product of the output layer value and the *tanh*-activated function of C^t and is

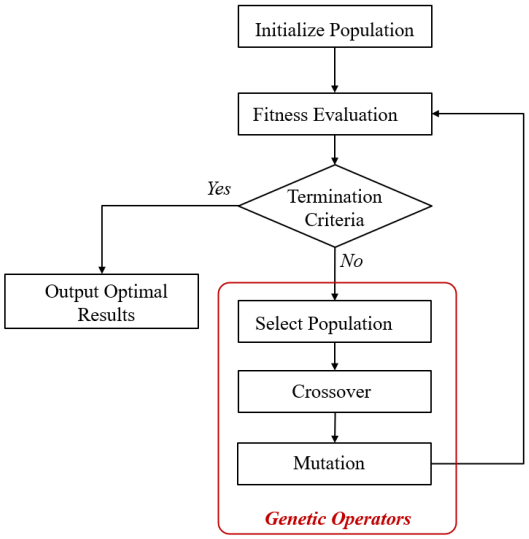


FIGURE 3. Genetic algorithm optimization flowchart.

expressed in (9).

$$h^t = O^t \odot \tanh\left(C^t\right) \tag{9}$$

To further improve the effectiveness of the LSTM, several LSTM cells could be stacked to form a series of deep architecture network whereby the output of one LSTM cell- h^t and C^t is propagated through time and becomes the input for the succeeding LSTM cell and so forth. Subsequently, for l number of LSTM cells (layers), the averaged sum over time as shown in (10) further ensures prior information for posterior estimations are well captured and/or modelled.

$$H = \sum_{i=1}^{l} \frac{h_l^t}{l} \tag{10}$$

D. GENETIC ALGORITHM (GA)

Like other optimizers, the goal of a GA is to search for the best solution from a set of initial parameter values following some update rule. GA optimizations mimics the evolution process by assessing a random population of initial parameter values with the goal of optimizing them. More often, the GA may skip a local minima while searching for global minimum and this is one of its unique attributes.

As shown in Fig. 3, the GA optimization steps consist of the six core stages– initialization, fitness evaluation, termination criteria assessment, selection, and mutation/mating. The *initialization* stage is the process of global search by the GA by defining a chromosome– an N–dimensional vector containing possible solutions to the problem. For each chromosome is an associated *fitness value* and is calculated with the predefined *fitness function*– a mathematical function used to numerically encode a chromosome's performance. Once the fitness function has been defined, the GA employs three genetic operators— *reproduction/selection, crossover/mating, and mutation* to generate new optimal solution chromosomes (see Fig. 3). The selection process reflects the Darwin's theory of survival of the fittest which ensures that the best fitness costs are retained (or improved) between successive generations. The *crossover/mating* process entails creating new offsprings of chromosomes from selected ones. This process creates a random binary vector commonly referred to as *the mask*, from which an offspring is produced by accepting the genes of the first parent at a mask value of 1. Also, at a mask value of 0, genes of the second parent are selected. After *crossover*, comes the *mutation* process wherein a random alteration of a set of variables in the list of chromosomes takes place in each generation. This mimics the typical evolutionary process in biology and serves the purpose of introducing diversity and novelty into the solution pool by arbitrarily interchanging/alternating solutions.

IV. PROPOSED PROGNOSTICS SCHEME

In this study, we propose a hybrid approach to the RUL prediction of electromagnetic pumps. The scheme consists of a FastICA-based HI construction from highly prognosible vibrational and pressure signals (following a multi-level DWT decomposition, feature extraction and selection) and a MOGA-optimized LSTM predictor. Fig. 4 shows the proposed prognostics scheme.

A. DE-NOISING AND FEATURE EXTRACTION

As shown in Fig. 4, from the raw vibrational and pressure signals, a multi-level decomposition of the signals is implemented for de-noising. From these levels, the statistical features defined in Table 1 are extracted. The DWT returns a data vector of the same length as the input signal by decomposing it into a set of wavelets (functions) that are orthogonal to its translations and scaling. By so doing, same or lower number of the wavelet coefficient spectrum decomposition outputs can be obtained; however, the choice of wavelet has been a major challenge for DWT decomposition. In practical applications, wavelet Shannon entropy provides a reliable paradigm for solving this problem.

Because the wavelet transform measures the similarity between a signal and scaled version of a base wavelet, the level of similarity between them is directly proportional to the wavelet decomposition coefficient and the corresponding energy concentration. Therefore, a befitting wavelet is expected to extract the maximum amount of energy while minimizing the Shannon entropy of the corresponding wavelet coefficients [18]. Eq. 11 computes the energy-Shannon entropy ratio (E-SER).

$$\text{E-SER}(j) = \frac{E(j)}{S_{\text{entropy}}(j)} \tag{11}$$

where $E(j)$ are the energy wavelet coefficients at j^{th} scale and is obtained using (12) below:

$$E(j) = \sum_{k=1}^{K} |wt(j,k)|^2 \tag{12}$$

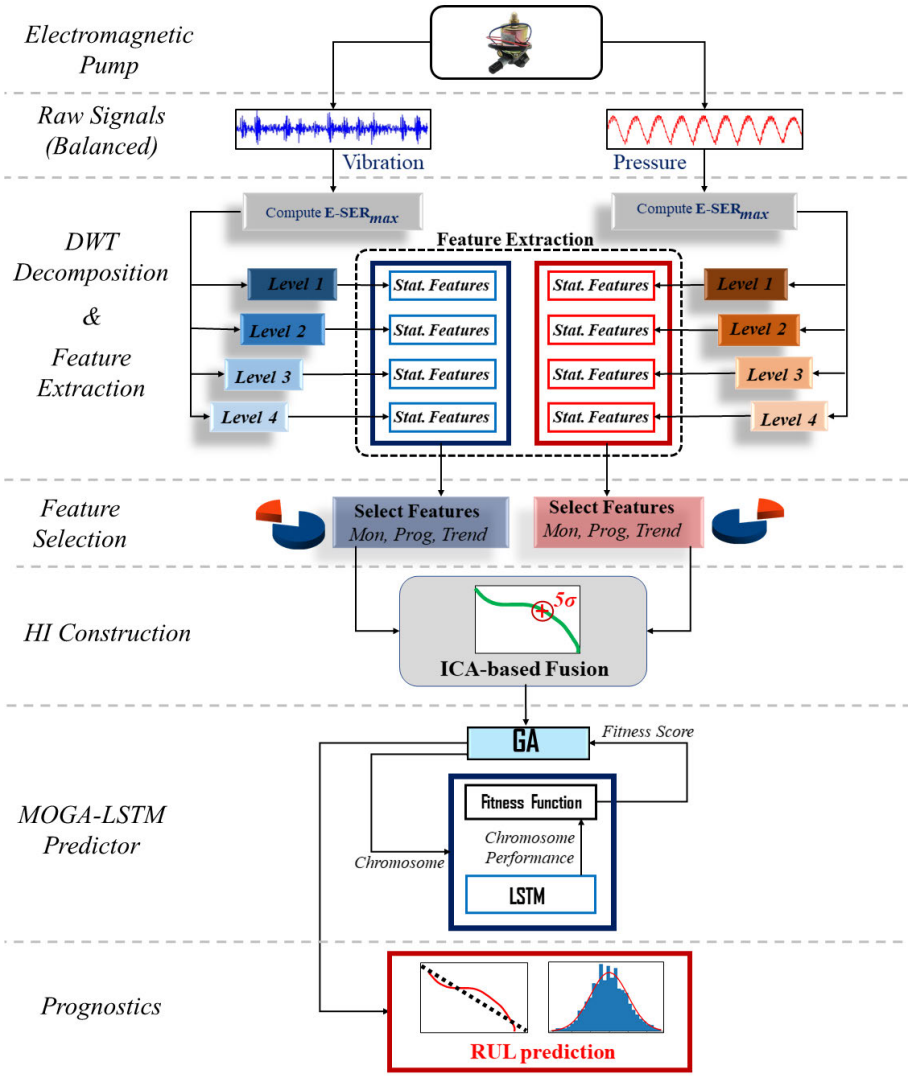


FIGURE 4. system model.

where K is the total number of wavelet coefficients, and $wt(j, k)$ is the k-th wavelet coefficient of j scale.

The $S_{\text{entropy}}(j)$ represents the Shannon entropy at j^{th} scale and is obtained using (13) below:

$$S_{\text{entropy}}(j) = -\sum_{k=1}^{K} p_k \log_a p_k \tag{13}$$

where p_k is the energy probability distribution of wavelet coefficients.

B. FEATURE SELECTION AND HI CONSTRUCTION

Practically, not all extracted features are prognosible parameters Although visual inspection and engineer/analyst's experience may be useful, such a biased approach is unreliable. As proposed in [36], the choice of a prognostics parameter depends on its characteristics— monotonicity (*Mon*), prognosability (*Prog*), and trendability (*Trend*) defined respectively in Eq. (14) - (16)

$$Mon = \text{mean}\left(\left|\frac{\#\text{ pos } d/dx}{n-1} - \frac{\#\text{neg } d/dx}{n-1}\right|\right) \tag{14}$$

where n is the number of observations in a particular history.

$$Prog = \exp\left(-\frac{\text{std(fail values)}}{\text{mean}(|\text{ fail value} - \text{start value }|)}\right) \tag{15}$$

$$Trend = \min\left(|\text{ corrcoef}_{ij}\right) | \tag{16}$$

These key metrics each return values between 0 and 1 whereby a value of 1 indicates a very high score while a value of 0 implies otherwise. Ideally, a threshold is required for selection; above which, any feature with a score equal to or higher than the threshold value is selected, while the rest are ignored. Automating the selection process can be achieved by defining a fitness function– Z as a weighted sum of the three metrics using Eq. (17).

$$Z = w_m\, Mon\, + w_p Prog + w_t Trend \tag{17}$$

TABLE 1. Multi-domain features and definitions.

Domain	Feature	Mathematical Model
Time-Domain	Root Mean Square	$\mathrm{X}_{rms} = \sqrt{\frac{\sum_{i=1}^{n}(x_i)^2}{n}}$
	Kurtosis	$\mathrm{X}_{kurt} = \frac{1}{N}\Sigma\left(\frac{(x_i-\mu)^3}{\sigma}\right)$
	Skewness	$\mathrm{X}_{\mathrm{skew}} = E\left[\left(\frac{(x_i-\mu)^3}{\sigma}\right)\right]$
	Standard Deviation	$\mathrm{X}_{sd} = \sqrt{\frac{\sum_{i=1}^{n}\lvert x-\mu\rvert^2}{N}}$
	Mean	$\mu = \frac{1}{N}\Sigma x_i$
	Max	$\mathrm{X}_{\max} = \max(x_i)$
	Min	$\mathrm{X}_{\max} = \min(x_i)$
	Variance	$\mathrm{X}_{var} = \frac{1}{N}\Sigma\left((x_i-\mu)^2\right)$
	Crest Factor	$\mathrm{X}_{CF} = \frac{x_{\max}}{x_{rms}}$
	Shape Factor	$\mathrm{X}_{SF} = \frac{x_{RMS}}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$
	Impulse Fcator	$\mathrm{X}_{IF} = \frac{x_{\max}}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$
	Peak-to-peak	$\mathrm{X}_{p-p} = x_{\max} - x_{\min}$
Frequency-Domain	Spectral Centroid	$FC = \frac{\sum_{i=2}^{N} x_i' x_i}{2\pi\sum_{i=1}^{N} x_i^2}$
	Root mean square frequency	$RMSF = \sqrt{\frac{\sum_{i=2}^{N}(x_i')^2}{4\pi^2\sum_{i=1}^{N} x_i^2}}$
	Spectral kurtosiss	$SK = \frac{2\sum_{k=0}^{\bar{B}_L/2-1}\left(\lvert X(k,n)\rvert-\mu_{\lvert X\rvert}\right)^4}{B_L\cdot\sigma_{\lvert X\rvert}^4} - 3$
	Spectral Skewness	$ss = \frac{2\sum_{k=0}^{B_L/2-1}\left(\lvert X(k,n)\rvert-\mu_{\lvert X\rvert}\right)^3}{B_L.\sigma_{\lvert x\rvert}^3}$

Subsequently, the FastICA algorithm is employed for fusing the highly prognosible features to form a single comprehensive HI for prognostics.

C. MOGA-LSTM FOR RUL PREDICTION

For prognostics, several evaluation metrics abound for exploration with each metric unique to its its pros and cons [37]; however, most research studies on these metrics assess prognostics algorithms based on algorithm's prediction accuracy, speed/cost of computation, and economy-centered metrics. In this study, the metrics of interest are those which assess prediction accuracy of the model and this prompts the need for multiple objective functions for the MOGA algorithm for finding a solution which satisfies these criteria. Consequently, the MOGA algorithm aims to discover the best LSTM parameters that return the:

- Minimum root mean square error (RMSE) of the RUL prediction using Eq. (18).
- Maximum relative accuracy (RA) of the RUL prediction using Eq. (19).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\left(x_i - x_i^*\right)^2} \tag{18}$$

where x_i and x_i^* are the actual and predicted index values respectively.

$$RA_{\lambda}^{l} = 1 - \frac{\left|r_*^l(i_\lambda) - \left\langle r^l(i_\lambda)\right\rangle\right|}{r_*^l(i_\lambda)} \tag{19}$$

where:

λ is the time window modifier such that $t_\lambda = t_P + \lambda\left(t_{EoL} - t_P\right)$

l is the l^{th} index for unit under test (UUT),

$r*(i_\lambda)$ is the ground truth RUL at time index i_λ

$\left\langle r^l(i_\lambda)\right\rangle$ is an appropriate central tendency point estimate of the predicted RUL distribution at time index i_λ

As proposed by Bentley and Wakefield [38], the sum of weighted objectives (SWO) is a very effective technique for multi–objective optimization problems whereby each objective is weighted to specify its relative importance; the fitness SWO fitness function is then produced by summing the multiple objectives using Eq. (20)

$$\text{fitness} = \sum w_i f_i \tag{20}$$

where f_i is the fitness function for the i^{th} objective.

Fig. 5 illustrates the optimization process of the MOGA for a 3–layer LSTM network using a set of training data (HI from start to time–to–start–prediction (TSP)) for discovering the best solution to a problem. The update process of the MOGA is bounded by the red dotted lines. If the best solution is met, the test data is used for performance evaluation (RUL prediction) in an unsupervised manner. In practice, choosing the appropriate weight values can be an uphill task (especially when there are many objectives); however, for dual–objective cases like the one presented in this study— *minimum RMSE* and *maximum RA*, an equal weighting is most ideal.

V. EXPERIMENTAL VALIDATION

This section presents the experimental validation of the proposed RUL prediction framework on a run-to-failure test on an electromagnetic pump.

A. EXPERIMENTAL SETUP AND DATA ACQUISITION

No past data-driven prognostics methodology had been reported for VSC63A5 electromagnetic Pumps nor any other solenoid pumps. This work, being the first attempt, aims at exploring vibrational and pressure signals for the RUL prediction. Following earlier studies on failure causes in pumps [34], [39], and failure diagnostics of electromagnetic pumps in particular [26], filter clogging has shown to be one of the most critical failure modes of solenoid pumps; hence the motivation behind this work— to simulate a natural operating condition of the pump under fluid contamination till full filter clogging occurs.

Fig. 6 shows a picture of the actual test bed while Fig. 7 illustrates the experimental setup. The solenoid pump, after being firmly mounted on a table vice, was connected to a constant AC supply rated 220V, 60Hz from an automatic voltage regulator. High sensitivity Piezoelectric accelerometers were

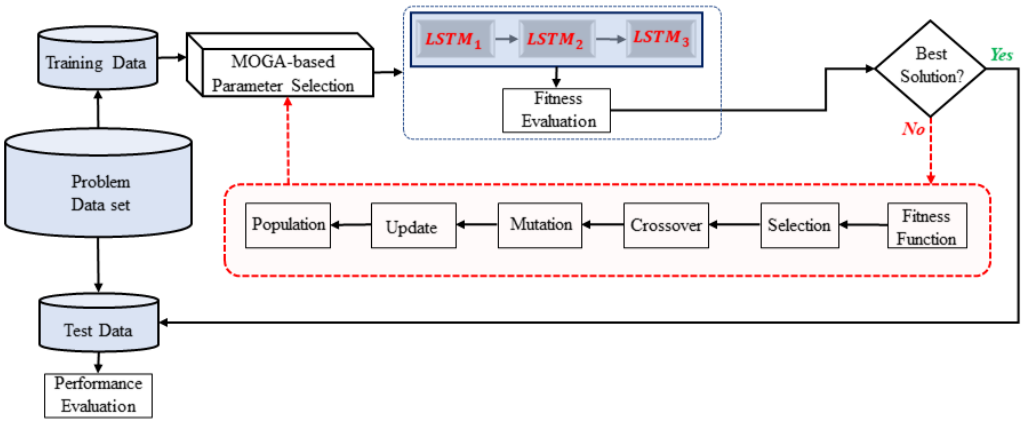


FIGURE 5. Flowchart of MOGA-LSTM.

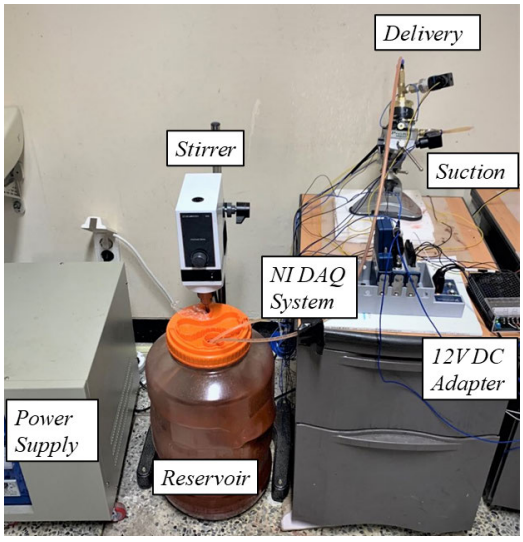


FIGURE 6. A picture of the actual testbed.

connected at the Z and X axes as shown in Fig. 7 to capture vibrational signals through a NI 9234 module at 20KHz sampling rate while two WIKA A10 transmitters (powered by a 20V DC adapter) were connected at the suction and delivery points of the pump to capture pressure measurements in the form of current signals $(4 - 20mA)$ via an NI 9228 current module at 20KHz. The temperature signals were captured at a sampling rate of 1KHz with an RTD thermocouple affixed to the pump's solenoid casing via an NI 9214 module and were only used for visual monitoring. The modules were connected to a NI Compact DAQ 9178 data acquisition system through which the digital signals were assessed on a LabView environment and stored in ''.csv'' file format.

As illustrated in Fig. 7, the mode of operation starts with a suction process via a flexible pipe inserted into the reservoir which contains 5 Litres of diesel and 10 grams of Iron(III) oxide (Fe_2O_3) which is constantly stirred. As the pump operates, the plunger oscillates vertically as a result of magnetization of the solenoids and this results in the pumping process— fluid delivery through the delivery port and return to the reservoir by a 1.0 GPH nozzle.

B. EXPERIMENTAL RESULTS

Following a daily operation of approximately 8 hours each day, the whole experiment lasted for 2,448 hours (102 days) with a significant decline in the delivery pressure and increased vibration caused by a clogged suction filter as shown in Fig. 8.

From the commencement of the experiment to about 1100 hours (first 46 days), a healthy *suction−delivery* process was observed with little sedimentation on the filter without any noticeable variation in the pressure sensors outputs (refer to Fig. 8(b)). This implies that even though the little filter clogging may be in place, the amount of clogging wasn't enough to cause any deviation from the normal operation of the pump. From the 46^{th} day however, cavitation was observed along the transparent delivery pipe with a fluctuation in the pressure signals. This signals a fault in the pump (a decline in the performance of the pump) and as shown in Fig. 8(c), the partial clogging must have caused the cavitation. This continued till the last day, when there was no more fuel delivery from the pump as a result of full filter clogging as shown in Fig. 8(d).

Fig. 9 and Fig. 10 show the vibrational and pressure signals for the whole run-to-failure time of the experiment.

C. DWT DECOMPOSITION, FEATURE ENGINEERING AND HI CONSTRUCTION

As a pre-processing step to capture both dependent and independent characteristics from the sensors, the signals were standardized and an ensemble–averaged output was computed for the respective sensor measurements. In addition to the four sensor signals, an ensemble–averaged (X and Z)–axis vibrational signal and an ensemble–averaged (suction and delivery)–axis pressure signal were computed. From these signals, DWT decomposition was conducted up to the 4^{th} level (*sym*5 wavelet returned the maximum *E-SER*). From the

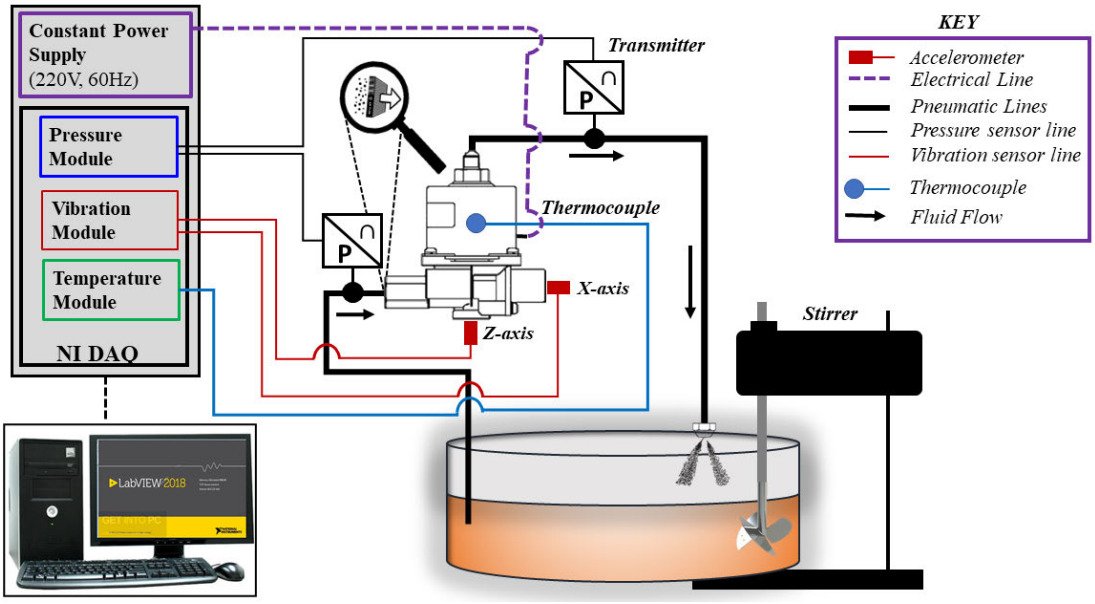


FIGURE 7. Experimental setup with illustrations for sensor placement.

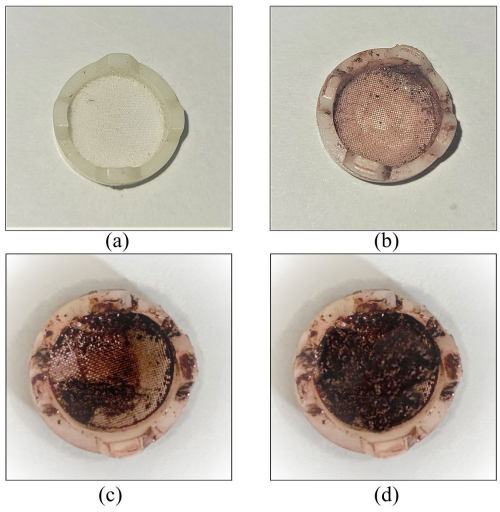


FIGURE 8. Pictures of the suction filter at different life stages (a) before experiment, (b) initial healthy running condition, (c) partial clogging resulting to cavitation, and (d) filter clogging failure at end of experiment.

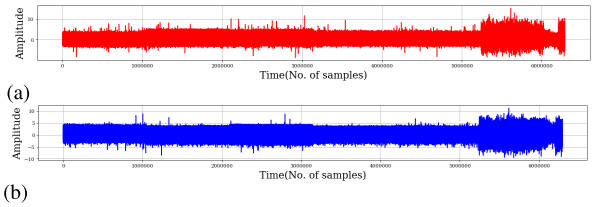


FIGURE 9. Vibrational signals for the run-to-failure test (a) X-axis, (b) Z-axis.

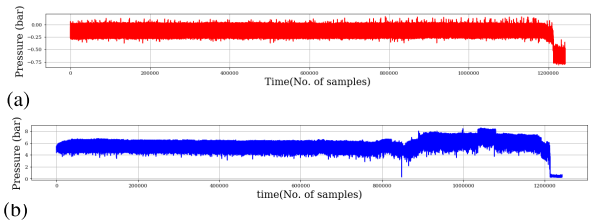


FIGURE 10. Pressure signals for the run-to-failure test (a) Suction, (b) Delivery.

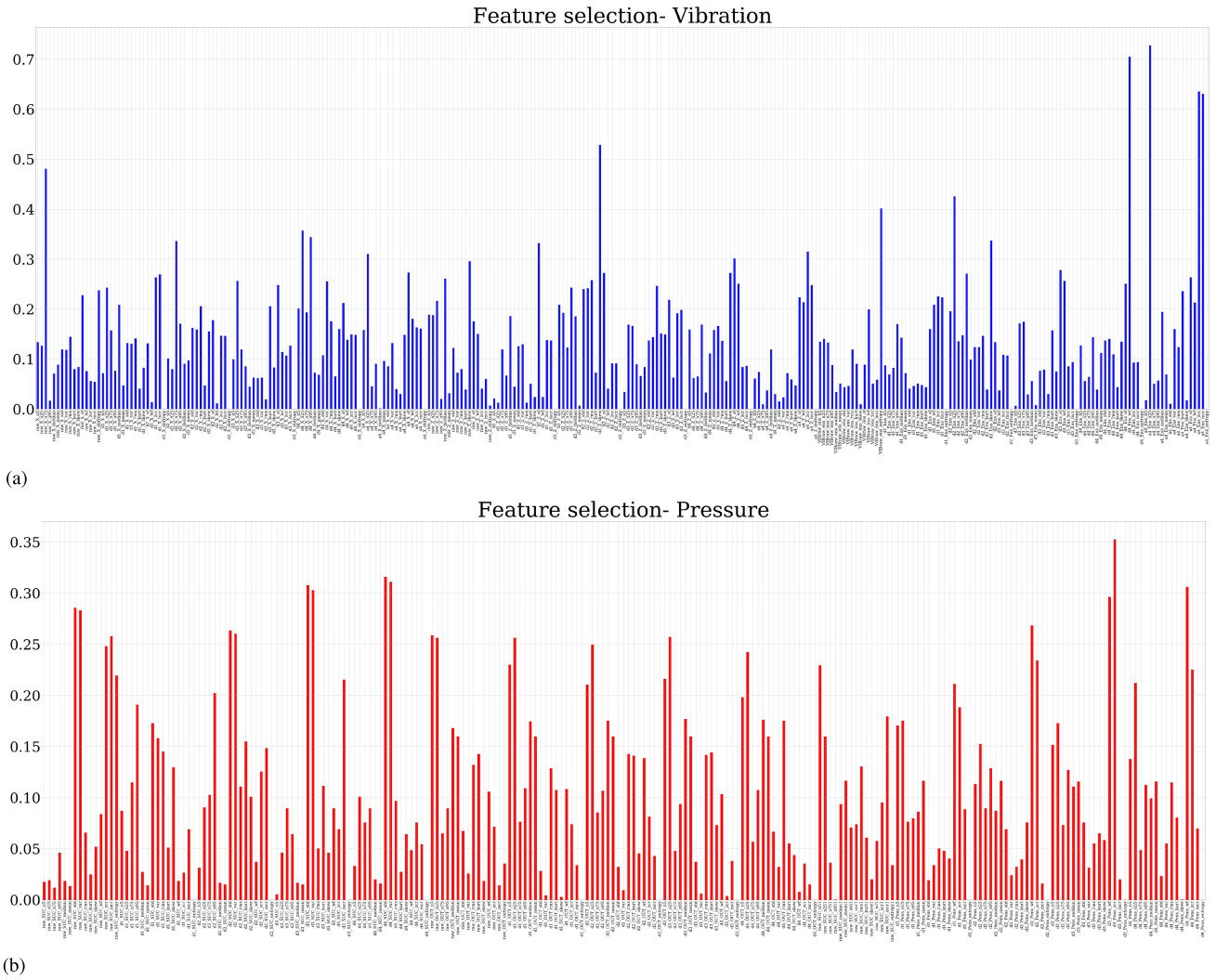
respective levels, the statistical features defined in Table 1 were extracted, smoothed, and tested for monotonicity, trendability, and prognosibility. Fig. 11 shows the the feature ranking results based on Z scores of vibrational and pressure signal features respectively.

Coble [6], prescribed a threshold of ($Z \geq 0.3$); as a result, highly prognosible features were selected accordingly and using the FastICA algorithm, the HI was constructed. See Fig. 12.

D. TSP/ALARM POINT DETERMINATION

From a prognostics viewpoint, early fault detection is necessary since most times, it may be too late to schedule maintenance. This is because the remaining time available before a complete failure could be short. This is where the need for an early initial degradation/fault detection approach arises.

Although the FastICA (and other dimensionality reduction methods) are effective for reliable HI construction, identifying degradation stages and TSP/alarm point from the newly constructed HI can be difficult since the new feature space dimension is dependent on (1) the number of constituent features, (2) the characteristics (variance) of constituent features, and (3) the structure/theoretical nature of the fusion algorithm; nevertheless, several approaches for TSP/alarm



(a)

(b)

FIGURE 11. Ranking of features based on Z score (a) features from Vibrational signals, (b) features from pressure signals.

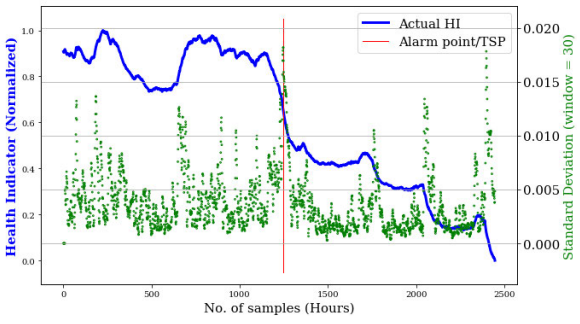


FIGURE 12. FastICA-generated HI for the whole run-to-failure time, STD of HI over time, and TSP/alarm point at 5σ.

point detection have been proposed including the use of clustering algorithms like the K-Means, fuzzy c-means, etc. for unsupervised partitioning of a HI based on cluster points [10], [41], [42], artificial neural networks, and intuition/experience of the engineer/analyst; however, from a statistical viewpoint for real-time monitoring (as is intended in this study), these approaches are quite limited in performance. interestingly, initial fault detection can be achieved in real-time by monitoring the standard deviation (STD) of the HI in real-time. As proposed by [43], a statistical alarm (or TSP) could be set at a window of 5 times the standard deviation of the healthy state, i.e. 5σ.

Fig. 12 shows the normalised HI in blue while the STD (a window of 30 samples) is shown in green dots. As shown, the HI reflects the time–dependent properties—healthy region, partial clogging (fault region), and full clogging (failure region). As shown, the STD spiked above the mean STD at healthy condition (0.003) at around the 46^{th} day when cavitation was actually observed, thus very reliable.

E. RUL PREDICTION

With the HI ready for prognostics with a monotonically decreasing trend, the failure threshold is set at the minimum ($HI = 0$) which indicates the the EOL. Fig. 13 shows the

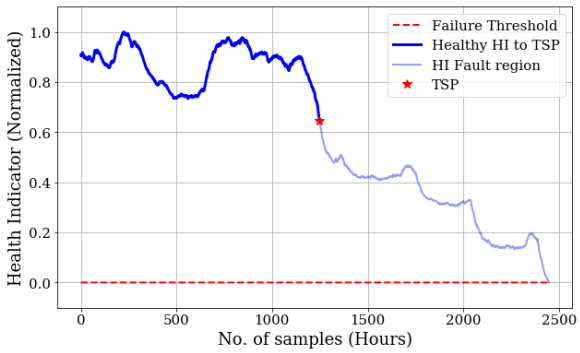


FIGURE 13. HI illustrating the TSP, EOL and Failure threshold.

TABLE 2. Parameters used in the experiment.

Description	Symbol	Value
Number of LSTM layers	-	3
Number of units in the input layer	p	4
Number of units in the first LSTM layer	m_1	*Given by GA*
Number of units in the Second LSTM layer	m_2	*Given by GA*
Number of units in the third LSTM layer	m_3	*Given by GA*
Iteration times of LSTM	T_2	50
Batch size of LSTM	b	10
Population size	N	20
Iteration times of GA	t_1	10
Crossover probability	p_c	0,7
Mutation probability	p_m	0.03
Gene length	l_g	10
Activation between LSTM layers	$f(x)$	$ReLu$
Activation between to output layer	$f(y)$	$Linear$

HI with the TSP and failure threshold at day 46 and 102 respectively.

1) MOGA-LSTM INITIALIZATION AND PARAMETERIZATION

For the RUL prediction using DL methods like the MOGA-LSTM predictor requires model training and parameter initialization. In this work, only healthy data (HI until the 46^{th} day) were used for training a 3–layer LSTM network with the parameters summarised in Table 2. A 3–layer LSTM with *ReLu* activation function between hidden layers and a *linear* activation function in the output layer (with *adam* optimization) has high predictive efficiencies. The *ReLu* activation function is quite popular for regression problems considering their comparative advantage over other activation functions— *sigmoid*, *tanh*, *softmax* for avoiding vanishing gradient problems and sparsity, and with the MOGA integrated for parameter optimization, minimal estimation errors can be achieved. Also, choosing a reasonable amount of iteration times for the LSTM (and the GA) algorithm ensures that an exhaustive search for optimal hyper-parameters is achieved.

Consequently, a one–step–ahead regression technique was employed for a supervised training process using the training data (1248 training examples) with the weighted sum of the RMSE and RA between the predicted future index values and the actual index values serving as the fitness function of the MOGA. Fig. 14 shows the training process of the model over

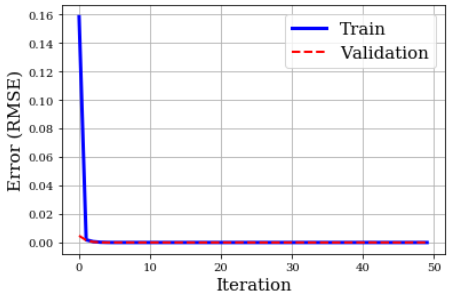


FIGURE 14. Training process of GA-LSTM over 50 iterations.

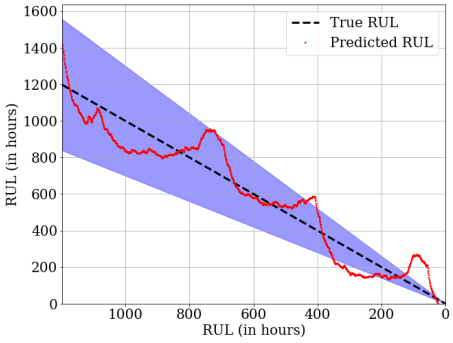


FIGURE 15. RUL prediction results by MOGA-LSTM at TSP (46^{th} day).

50 iterations. Usually, monitoring the iteration process for convergence after training provides insight on the predictive capability of the model since a convergence of the validation error towards zero indicates accurate learning— mapping of input variables to the output; however, it is worthy to note that convergence does not imply model perfection since there could be possibilities of under–fitting or over–fitting; therefore, predictive performance can only be ascertained by its performance on test data.

2) RESULTS

Using the trained model, the one–step ahead prediction is done at TSP followed by the RUL prediction. The RUL prediction is achieved by first computing the ground truth (the actual RUL) and with the trained model, the RUL from TSP to EOL is estimated. The RUL prediction results by the MOGA-LSTM predictor at TSP is shown in Fig. 15 while Fig. 16(a) shows the one-step ahead prediction.

As shown, the proposed method is suitable for learning the complex dynamic behaviour of the HI as the predicted trend follows with great similarity, the actual test data (test HI).

3) PERFORMANCE EVALUATION

As achieved in the previous subsection, the MOGA-LSTM predictor has shown highly reliable prognostic capabilities at TSP; however, this section aims to further assess its performance by making predictions on the 65^{th}, 75^{th}, and 85^{th} days respectively and comparing its results using standard prognostics performance metrics— RMSE, MAE, and

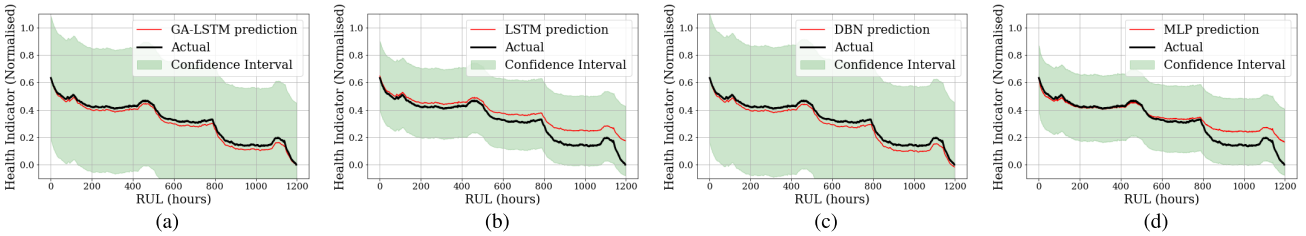


FIGURE 16. One-step ahead prediction results at TSP point (46^{th} day) by (a) GA-LSTM, (b) LSTM, (c) DBN, and (D) MLP.

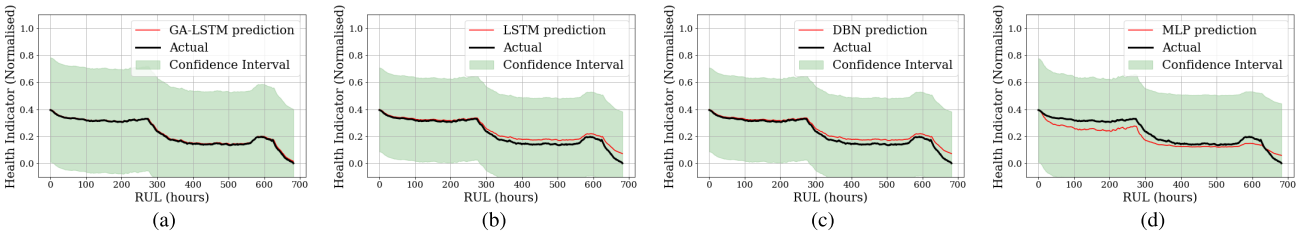


FIGURE 17. One-step ahead prediction results on 65^{th} day) by (a) GA-LSTM, (b) LSTM, (c) DBN, and (D) MLP.

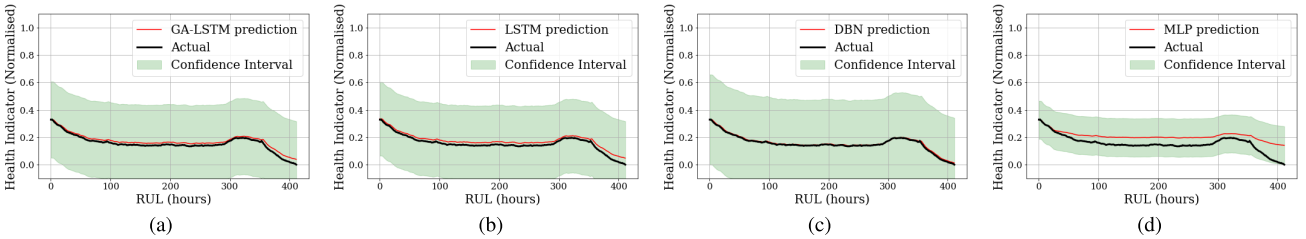


FIGURE 18. One-step ahead prediction results on 75^{th} day) by (a) GA-LSTM, (b) LSTM, (c) DBN, and (D) MLP.

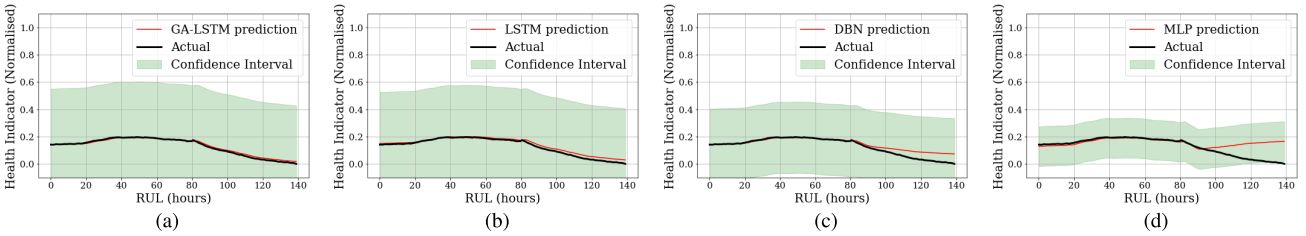


FIGURE 19. One-step ahead prediction results on 85^{th} day) by (a) GA-LSTM, (b) LSTM, (c) DBN, and (D) MLP.

MAPE [37] with other standard predictors— a single–layer LSTM predictor, a 3–layer deep belief network (DBN) predictor, and a 3–layer perceptron (MLP) neural network predictor were used for comparison.

Figs. 16–19 show the show the one-step ahead prediction results of the methods at TSP point (46^{th} day), 65^{th}, 75^{th}, and 85^{th} days respectively. In all the prediction points, the MOGA-LSTM's predictive prowess is comparatively more reliable as it learns the test data better than the other predictors. This is obviously associated with the following:

- Compared with the shallow single LSTM predictor, a multi-layer LSTM predictors are more likely to effectively capture characteristic information and learn the dynamic relationships hidden in the HI.
- With continuous-valued stochastic units, the MOGA-LSTM illustrates superior efficiencies in handling the input variables (HI) with complex non-linear characteristics.
- The integration of the MOGA further enhances predictive efficiencies of the deep LSTM model by finding optimal parameters for a minimal false alarm rate in RUL prediction.

Although the other algorithms also made reliable predictions as shown in Figs. 16–19, as also shown in Fig. 20, the proposed method outperforms the other competitors with

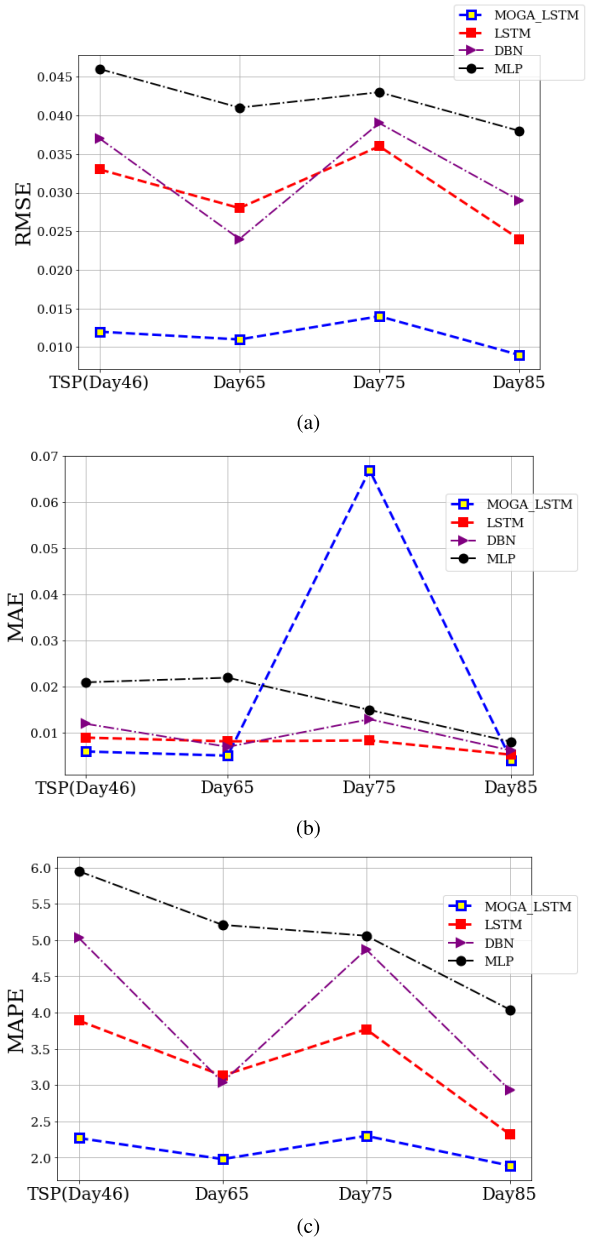


FIGURE 20. **RUL prediction performance comparison of predictors (a) RMSE, (b) MAE, and (c) MAPE.**

the least RMSE, MAPE, and MAE at the various prediction points.

VI. CONCLUDING REMARKS

As the first attempt on a data-driven approach to prognostics of electromagnetic pumps, this paper proposed a DL–based prognostics framework for solenoid pump prognostics by employing a sensor fusion technique. From vibrational and pressure signals, statistical features were extracted after a multi–level wavelet decomposition process from which, highly prognosible features/parameters were selected following prognosibility, monotonicity and trendability tests. The feature fusion technique— FastICA provides an effective method for fusing highly prognosible parameters/features for generating a comprehensive HI. This serves as input for the proposed MGA-LSTM prognostics model whose prowess for automated search for optimal hyper-parameters (by the MOGA) ensures a highly reliable prognostics performance by the LSTM model.

The proposed model was validated on a VSC63A5 Solenoid pump manufactured by Korea Control Limited under normal running conditions with an Fe_2O_3–contaminated diesel as the working fluid and its performance compared with other standard RUL prediction methods. The results show that the proposed method is more accurate and although more computationally expensive due to the exhaustive search by the MOGA, the proposed method better meets industrial demands which include real-time predictive capability, automation, accuracy, and mitigation of strenuous parameter tuning which usually lead to over–fitting and/or under–fitting.

REFERENCES

[1] S. Yin, J. J. Rodriguez-Andina, and Y. Jiang, "Real-time monitoring and control of industrial cyberphysical systems: With integrated plant-wide monitoring and control framework," *IEEE Ind. Electron. Mag.*, vol. 13, no. 4, pp. 38–47, Dec. 2019.

[2] (Mar. 2019). *Ethiopian Airlines: No Survivors on Crashed Boeing 737*. [Online]. Available: https://www.bbc.com/news/world-africa-47513508

[3] W.-C. Feng, "Making a case for efficient supercomputing: It is time for the computing community to use alternative metrics for evaluating performance," in *Queue*, vol. 1, no. 7. New York, NY, USA: Association Computing Machinery, 2003, pp. 1542–7730, doi: 10.1145/957717.957772.

[4] X. Fei, C. Bin, C. Jun, and H. Shunhua, "Literature review: Framework of prognostic health management for airline predictive maintenance," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Hefei, China, Aug. 2020, pp. 5112–5117, doi: 10.1109/CCDC49329.2020.9164546.

[5] J. Guo, Z. Li, and M. Li, "A review on prognostics methods for engineering systems," *IEEE Trans. Rel.*, vol. 69, no. 3, pp. 1110–1129, Sep. 2020, doi: 10.1109/TR.2019.2957965.

[6] J. Coble, "Merging data sources to predict remaining useful life—An automated method to identify prognostic parameters," Ph.D. dissertation, Dept. Nucl. Eng., Univ. Tennessee, Knoxville, TN, USA, 2010.

[7] U. E. Akpudo and J. W. Hur, "A feature fusion-based prognostics approach for rolling element bearings," *J. Mech. Sci. Technol.*, vol. 34, no. 10, pp. 4025–4035, 2020.

[8] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Comput.*, vol. 11, no. 2, pp. 417–441, Feb. 1999.

[9] L. Jing, T. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, Feb. 2017.

[10] U. E. Akpudo and J. W. Hur, "Towards bearing failure prognostics: A practical comparison between data-driven methods for industrial applications," *J. Mech. Sci. Technol.*, vol. 34, no. 10, pp. 4161–4172, 2020.

[11] C. Sun, Z. He, H. Cao, Z. Zhang, X. Chen, and M. Jian Zuo, "A non-probabilistic metric derived from condition information for operational reliability assessment of aero-engines," *IEEE Trans. Rel.*, vol. 64, no. 1, pp. 167–181, Mar. 2015.

[12] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Evaluating algorithm performance metrics tailored for prognostics," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, Mar. 2009, pp. 1–13.

[13] S. Chen, M. Wang, D. Huang, P. Wen, S. Wang, and S. Zhao, "Remaining useful life prediction for complex systems with multiple indicators based on particle filter and parameter correlation," *IEEE Access*, vol. 8, pp. 215145–215156, 2020, doi: 10.1109/ACCESS.2020.3041682.

[14] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1996.

[15] M. D. L. D. Vedova, P. C. Berri, G. Bonanno, and P. Maggiore, "Fault detection and identification method based on genetic algorithms to monitor degradation of electrohydraulic servomechanisms," in *Proc. 4th Int. Conf. Syst. Rel. Saf. (ICSRS)*, Rome, Italy, Nov. 2019, pp. 304–311, doi: 10.1109/ICSRS48664.2019.8987734.

[16] S. Visalakshi and V. Radha, "A literature review of feature selection techniques and applications: Review of feature selection in data mining," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res.*, Coimbatore, India, Dec. 2014, pp. 1–6, doi: 10.1109/ICCIC.2014.7238499.

[17] W. Yuan, "Study on noise elimination of mechanical vibration signal based on improved wavelet," in *Proc. 12th Int. Conf. Measuring Technol. Mechatronics Autom. (ICMTMA)*, Phuket, Thailand, Feb. 2020, pp. 141–143, doi: 10.1109/ICMTMA50254.2020.00039.

[18] X. Chen, Y. Yang, Z. Cui, and J. Shen, "Wavelet denoising for the vibration signals of wind turbines based on variational mode decomposition and multiscale permutation entropy," *IEEE Access*, vol. 8, pp. 40347–40356, 2020, doi: 10.1109/ACCESS.2020.2975875.

[19] G. Chen, W. Xie, and Y. Zhao, "Wavelet-based denoising: A brief review," in *Proc. 4th Int. Conf. Intell. Control Inf. Process. (ICICIP)*, Beijing, China, Jun. 2013, pp. 570–574, doi: 10.1109/ICICIP.2013.6568140.

[20] S. Chatterjee, R. S. Thakur, R. N. Yadav, L. Gupta, and D. K. Raghuvanshi, "Review of noise removal techniques in ECG signals," *IET Signal Process.*, vol. 14, no. 9, pp. 569–590, Dec. 2020, doi: 10.1049/iet-spr.2020.0104.

[21] M. J. Roemer, G. J. Kacprzynski, and M. H. Schoeller, "Improved diagnostic and prognostic assessments using health management information fusion," in *Proc. IEEE Autotestcon IEEE Syst. Readiness Technol. Conf.*, Valley Forge, PA, USA, Aug. 2001, pp. 365–377, doi: 10.1109/AUTEST.2001.948984.

[22] L. Duan, F. Zhao, J. Wang, N. Wang, and J. Zhang, "An integrated cumulative transformation and feature fusion approach for bearing degradation prognostics," *Shock Vibrat.*, vol. 2018, pp. 1–15, Feb. 2018, doi: 10.1155/2018/9067184.

[23] L. Bagheriye, G. Ali, and H. G. Kerkhoff, "Life-time prognostics of dependable VLSI-SoCs using machine-learning," in *Proc. IEEE 26th Int. Symp. Line Test. Robust Syst. Design (IOLTS)*, Napoli, Italy, Jul. 2020, pp. 1–4, doi: 10.1109/IOLTS50870.2020.9159753.

[24] V. D. Calhoun and T. Adali, "Feature-based fusion of medical imaging data," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 711–720, Sep. 2009, doi: 10.1109/TITB.2008.923773.

[25] J. Weidong, "Fault diagnosis of gearbox by FastICA and residual mutual information based feature extraction," in *Proc. Int. Conf. Inf. Autom.*, Zhuhai, Macau, Jun. 2009, pp. 928–932, doi: 10.1109/ICINFA.2009.5205051.

[26] U. E. Akpudo and H. Jang-Wook, "A multi-domain diagnostics approach for solenoid pumps based on discriminative features," *IEEE Access*, vol. 8, pp. 175020–175034, 2020, doi: 10.1109/ACCESS.2020.3025909.

[27] V. S. Tormozov, A. L. Zolkin, and K. A. Vasilenko, "Optimization of neural network parameters based on a genetic algorithm for prediction of time series," in *Proc. Int. Multi-Conf. Ind. Eng. Mod. Technol. (FarEastCon)*, Vladivostok, Russia, Oct. 2020, pp. 1–4, doi: 10.1109/FarEastCon50210.2020.9271536.

[28] W. Bendali, I. Saber, B. Bourachdi, M. Boussetta, and Y. Mourad, "Deep learning using genetic algorithm optimization for short term solar irradiance forecasting," in *Proc. 4th Int. Conf. Intell. Comput. Data Sci. (ICDS)*, Fez, Morocco, Oct. 2020, pp. 1–8, doi: 10.1109/ICDS50568.2020.9268682.

[29] Y. Xiang, Z. Liu, and L. Wang, "Genetic-algorithm-optimization-based predictive functional control for chemical industry processes against partial actuator faults," *IEEE Access*, vol. 8, pp. 214586–214595, 2020, doi: 10.1109/ACCESS.2020.3041015.

[30] A. Starkey, H. Hagras, S. Shakya, and G. Owusu, "A comparison of particle swarm optimization and genetic algorithms for a multi-objective type-2 fuzzy logic based system for the optimal allocation of mobile field engineers," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Vancouver, BC, Canada, Jul. 2016, pp. 5068–5075, doi: 10.1109/CEC.2016.7748332.

[31] D. Wei, "Prediction of stock price based on LSTM neural network," in *Proc. Int. Conf. Artif. Intell. Adv. Manuf. (AIAM)*, Dublin, Ireland, Oct. 2019, pp. 544–547, doi: 10.1109/AIAM48774.2019.00113.

[32] Y. Yu, J. Cao, and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019, doi: 10.1109/ACCESS.2019.2946057.

[33] S. Chakraborty, J. Banik, S. Addhya, and D. Chatterjee, "Study of dependency on number of LSTM units for character based text generation models," in *Proc. Int. Conf. Comput. Sci., Eng. Appl. (ICCSEA)*, Gunupur, India, Mar. 2020, pp. 1–5, doi: 10.1109/ICCSEA49143.2020.9132839.

[34] *Condition Monitoring and Diagnostics of Machines Prognostics Part 1: General Guidelines. International Standards Organization*, Standard ISO 13381-1:2015, 2015.

[35] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.

[36] J. Coble and J. W. Hines, "Fusing data sources for optimal prognostic parameter selection," *Trans. Amer. Nucl. Soc.*, vol. 100, no. 1, pp. 211–212, 2009.

[37] A. Saxena, R. J. Celaya, B. Saha, S. Saha, and K. Goebel, "Metrics for offline evaluation of prognostic performance," *Int. J. Prognostics Health Manage.*, vol. 1, no. 1, pp. 4–23, 2010.

[38] P. J. Bentley and J. P. Wakefield, "An analysis of multiobjective optimization within genetic algorithms," Dept. Division Comput. Control Syst. Eng., Univ. Huddersfield, Huddersfield, U.K., Tech. Rep. ENGPJB96, 1996.

[39] M. Kristoffer, F. Gareth, M. Ilyas, E. Rodney, and H. Ian, "A review of major centrifugal pump failure modes with application to the water supply and sewerage industries," in *Proc. Asset Manage. Conf.* Huddersfield, U.K.: Univ. of Huddersfield, 2011, pp. 1–4.

[40] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, vol. 9, no. 8. Cambridge, MA, USA: MIT Press, 1997, doi: 10.1162/neco.1997.9.8.1735.

[41] D. Wu, Q. Yang, F. Tian, and D. X. Zhang, "Fault diagnosis based on K-means clustering and PNN," in *Proc. 3rd Int. Conf. Intell. Netw. Intell. Syst.*, Nov. 2010, pp. 173–176.

[42] J. Shi, Q. He, and Z. Wang, "GMM clustering-based decision trees considering fault rate and cluster validity for analog circuit fault diagnosis," *IEEE Access*, vol. 7, pp. 140637–140650, 2019.

[43] A. Klausen, H. Van Khang, and K. G. Robbersmyr, "Novel threshold calculations for remaining useful lifetime estimation of rolling element bearings," in *Proc. 13th Int. Conf. Electr. Mach. (ICEM)*, vol. 7, Sep. 2018, pp. 1912–1918.

UGOCHUKWU EJIKE AKPUDO (Graduate Student Member, IEEE) received the B.Eng. degree in mechanical and production engineering from the Enugu State University of Science and Technology, Nigeria, in 2012, and the M.Eng. degree in mechanical engineering from the Kumoh National Institute of Technology, South Korea. He is currently pursuing the Ph.D. degree with the Defense Reliability Laboratory, Mechanical Systems Engineering, Kumoh National Institute of Technology, Gumi, South Korea. While working at Maclisle Complex Limited, Enugu, Nigeria, from 2015 to 2019, he familiarized himself with a wide variety of efficient and effective ways of mechanical systems maintainability, general safety engineering, reliability engineering, and loss prevention. He is also a full-time Researcher with the Defense Reliability Laboratory, Mechanical Systems Engineering, Kumoh National Institute of Technology, Gumi, South Korea. His research interests include but not limited to prognostics and health management of mechatronic systems and reliability engineering.

HUR JANG-WOOK received the Ph.D. degree in mechanical engineering from the Tokyo Institute of Technology, Japan, in 1995. While serving in the Republic of Korea National Military, in 2011, he ranked a Colonel and took part in various projects funded by the Korean Government like the DAPA KHP Project. From 2015 to 2020, he led the project team in the RAM program aimed at optimizing reliability, availability, and maintainability in the Korean Defense Systems. He is currently the HOD of Mechanical Systems Engineering and the Director of Defense Reliability Laboratory (an Advanced Research Center sup-ported by the Korean Government) at the Kumoh National University of Science and Technology. He is currently the Vice President of the Korean Society for Prognostics and Health Management (KSPHM). His research interests are prognostics and health management, reliability engineering, and systems engineering.