

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337025820>

# Health prediction of Hydraulic Cooling Circuit using Deep Neural Network with Ensemble Feature Ranking Technique

Article · November 2019

DOI: 10.1016/j.measurement.2019.107225

---

CITATIONS

8

READS

31

2 authors, including:



Jatin Prakash

Indian Institute of Technology Indore

6 PUBLICATIONS 11 CITATIONS

SEE PROFILE



# Health prediction of hydraulic cooling circuit using deep neural network with ensemble feature ranking technique



Jatin Prakash, P.K. Kankar \*

*System Dynamics Lab, Discipline of Mechanical Engineering, Indian Institute of Technology Indore, Indore 453552, India*

## ARTICLE INFO

### Article history:

Received 19 August 2019

Received in revised form 11 October 2019

Accepted 29 October 2019

Available online 5 November 2019

### Keywords:

Hydraulic system

Feature ranking

XGBoost

Deep neural network

Activation function

## ABSTRACT

Health prediction of the hydraulic systems is of utmost importance as any breakdown may lead to severe losses. In the present manuscript, the emphasis is on developing an artificially intelligent model using a deep neural network to predict the working behaviour of the cooling circuit in the hydraulic system. Overall, four different models have been proposed and compared for their performance. Features are calculated from the pressure signals. The capabilities of XGBoost and ReliefF have been compared as a feature ranking technique and the implications of two different activation function “tanh” and “relu” have been analysed. Features shortlisted through XGBoost gives higher performance with “tanh” activation function. The result reveals that the deep neural network model can be effectively used to predict the health of hydraulic cooling circuit.

© 2019 Elsevier Ltd. All rights reserved.

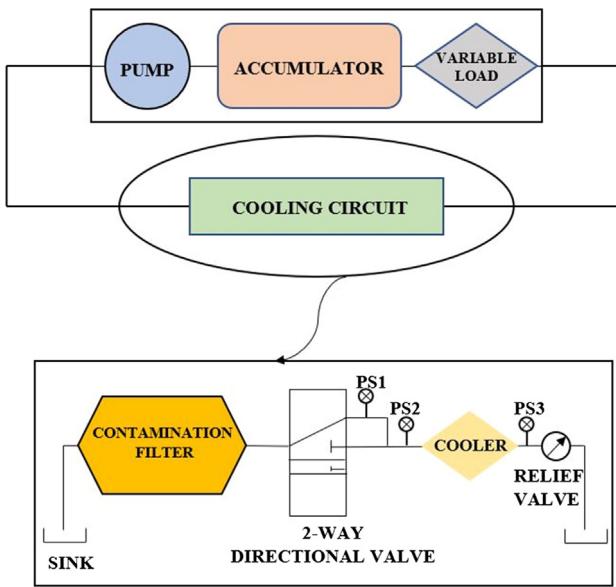
## 1. Introduction

Hydraulic systems are one of the major sub-systems widely used in industries [1]. They can transmit high loads in lower efforts. The proper functioning of such systems are highly essential. Nowadays, the monitoring of mechanical systems can automatically be performed employing various machine learning algorithms [2]. The condition monitoring technique uses parameters of the system to identify the substantial change that may be indicative of any developing fault [3]. Identification of such critical parameters is a challenging task for maintenance engineer and essential for predictive maintenance. Briefly, the fault detection technique can be classified into three categories as data-driven fault detection technique, model-based fault detection technique and hybrid fault detection technique [4–7]. The model-based fault detection technique requires the mathematical model of the system. A mathematical model has been developed to determine the failure of the single-stage pin-mounted hydraulic cylinder [8]. Sometimes, the actual system is too complex for mathematical modelling, thus data-driven techniques are preferred [4]. Few of the predominant data-driven health monitoring techniques include wavelets [9,10], regression [11] and statistical methods [12]. Fuzzy logic is another popular technique used for prognostics [13]. In these data-driven techniques, data is acquired continu-

ously from sensors and processed to predict healthy and faulty conditions. The vibration-based analysis technique is most widely used in condition-based monitoring and is classified as a non-destructive technique [14–15]. Literatures suggest different approaches like decision tree (DT), artificial neural network (ANN) and support vector machine (SVM) methods for the fault diagnosis [16–17]. Besides, the hybrid condition monitoring technique combines the model-based analysis with various neural network models for classification and detection of the faults [18]. Feature extraction from the raw signal is one of the essential steps in order to determine the useful information indicating the health of the system. Features classified in various categories such as time-domain features, frequency domain features, wavelet-based features etc. have been proposed and employed successfully for the health monitoring of the rotating components [19]. In order to enhance the computational efficiency of the model, different feature ranking methods have been used to reduce the dimensionality by selecting the optimal features [17]. Feature ranking techniques such as Fisher score, Laplacian Score, Mutual Information, genetic algorithm, Wilcoxon ranking, Chi-square, ReliefF have been employed for bearing condition monitoring [19,20]. These features were used to train the model and for predicting the behaviour of the system. In the context of the condition monitoring of hydraulic systems, studies are limited to actuator internal leakage and valve spool blockage using ANN [22,23]. In another study, a model for the internal leakage detection is proposed using intrinsic mode function combined with random forest classifier for the condition

\* Corresponding author.

E-mail addresses: [phd1901103007@iiti.ac.in](mailto:phd1901103007@iiti.ac.in) (J. Prakash), [\(P.K. Kankar\).](mailto:pkankar@iiti.ac.in)



**Fig. 1.** Schematic diagram of the hydraulic system consisting of cooling circuit.

monitoring of the hydraulic cylinder by monitoring the inlet and outlet pressure [24].

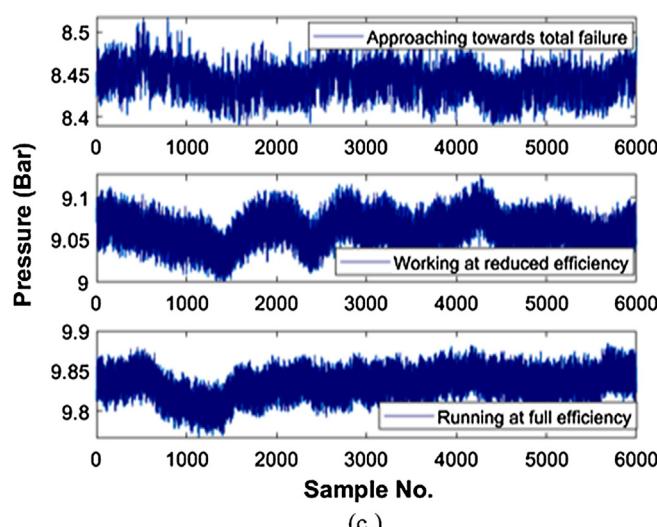
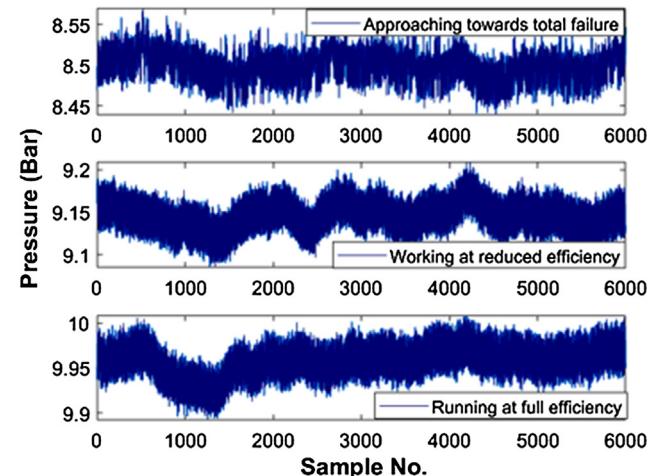
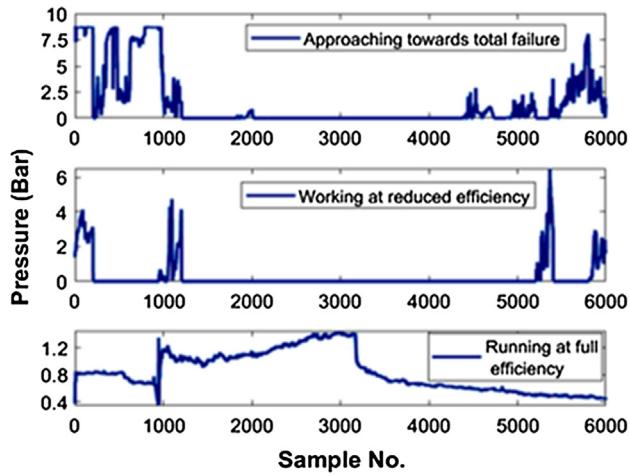
In this work, deep neural network (DNN) models are developed for the prediction of the cooling system in the hydraulic system. The dimensionality of feature vectors is reduced by the feature selection using XGBoost and ReliefF criteria. The feature importance plots are shown for both criteria. Two parallel DNN models have been analysed using two different activation functions. Thus, in total four distinct models have been proposed in this study and their accuracies are compared.

## 2. Methodology

### 2.1. Datasets

The generalized form of condition monitoring of a hydraulic system begins with the extraction of data using the sensors. The dataset in this study is obtained from the University of California (UCI) Machine Learning Repository [25]. The test rig consists of two circuits: the hydraulic circuit and the cooling circuit as shown in Fig. 1.

The hydraulic circuit consists of various components like pump, accumulator and the mechanism for variable load application. The



**Fig. 2.** (a.) Raw pressure signal variation for pressure sensor 1 for three different conditions (b.) Raw pressure signal variation for pressure sensor 2 for three different conditions (c.) Raw pressure signal variation for pressure sensor 3 for three different conditions.

enhanced view of the cooling circuit is also shown in Fig. 1. A contamination filter is attached to avoid the entry of contaminant in the flow of the hydraulic fluid. The fluid is passed through 2-way directional valve followed by the flow-through the cooler and then via relief valve to the fluid sink below. Two pressure sensors namely PS1 and PS2 are installed between the 2-way directional valve and the cooler. Pressure Sensor PS3 is installed between the cooler and the relief valve to measure the pressure of the fluid flow. The data is captured at the sampling frequency of 100 Hz for 60 s. Total of 2205 instances of data is recorded. The signals obtained from the pressure sensors PS1, PS2 and PS3 belong to three different categories i.e. *Approaching towards total failure*, *Working at reduced efficiency*, and *Running at full-efficiency*. The raw signal profile for all three pressure sensors, PS1, PS2 and PS3 for three different categories is shown in Fig. 2.

## 2.2. Feature extraction

Statistical Features like mean, skewness, kurtosis, shape indicator, clearance indicator etc. are commonly used for fault diagnosis

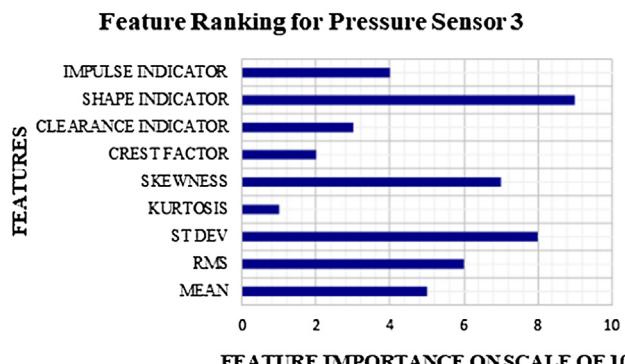
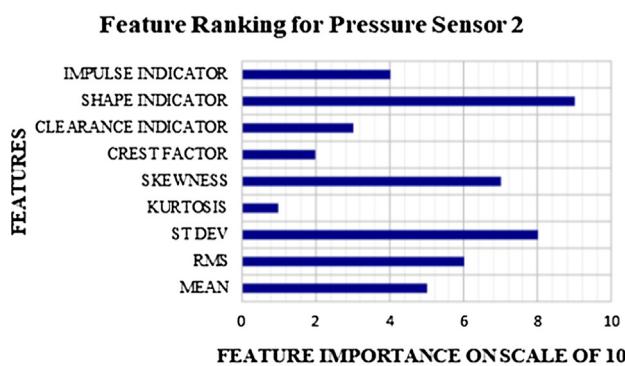
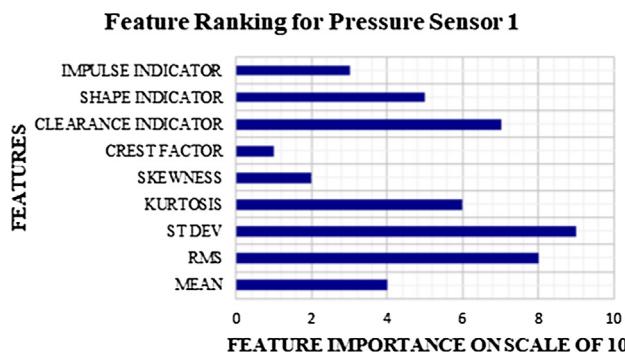


Fig. 3. Feature Importance plot considering ReliefF.

of mechanical systems [10]. A wide number of statistical features are extracted from pressure signals as described below [26].

- Mean: Mean is demarcated as the ratio of summation of all data to the number of elements of the data.

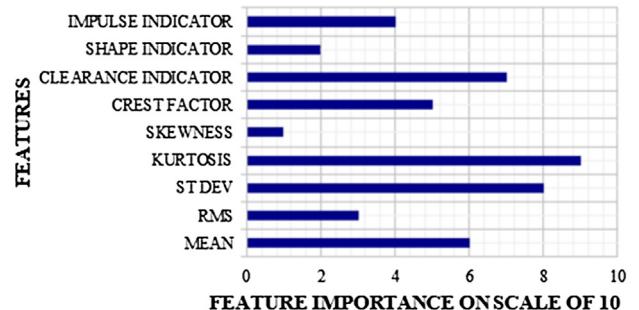
$$\text{Mean}(\bar{x}) = \frac{\sum_{i=0}^n x_i}{n} \quad (1)$$

- Root Mean Square (RMS): RMS, the quadratic mean is the square root of the arithmetic mean of the square of the data.

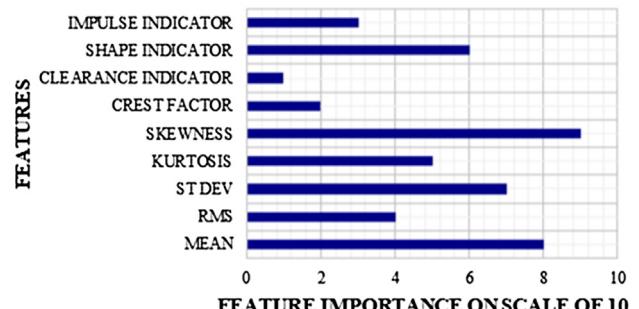
$$\text{RMS} = \sqrt{\frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2)} \quad (2)$$

- Standard Deviation: Standard deviation infers the scattering of discrete data about the mean value of the dataset.

## Feature Ranking for Pressure Sensor 1



## Feature Ranking for Pressure Sensor 2



## Feature Ranking for Pressure Sensor 3

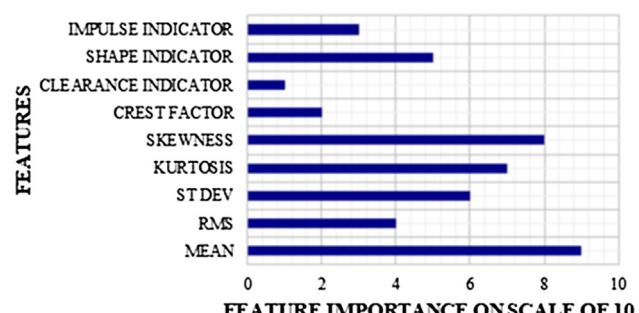
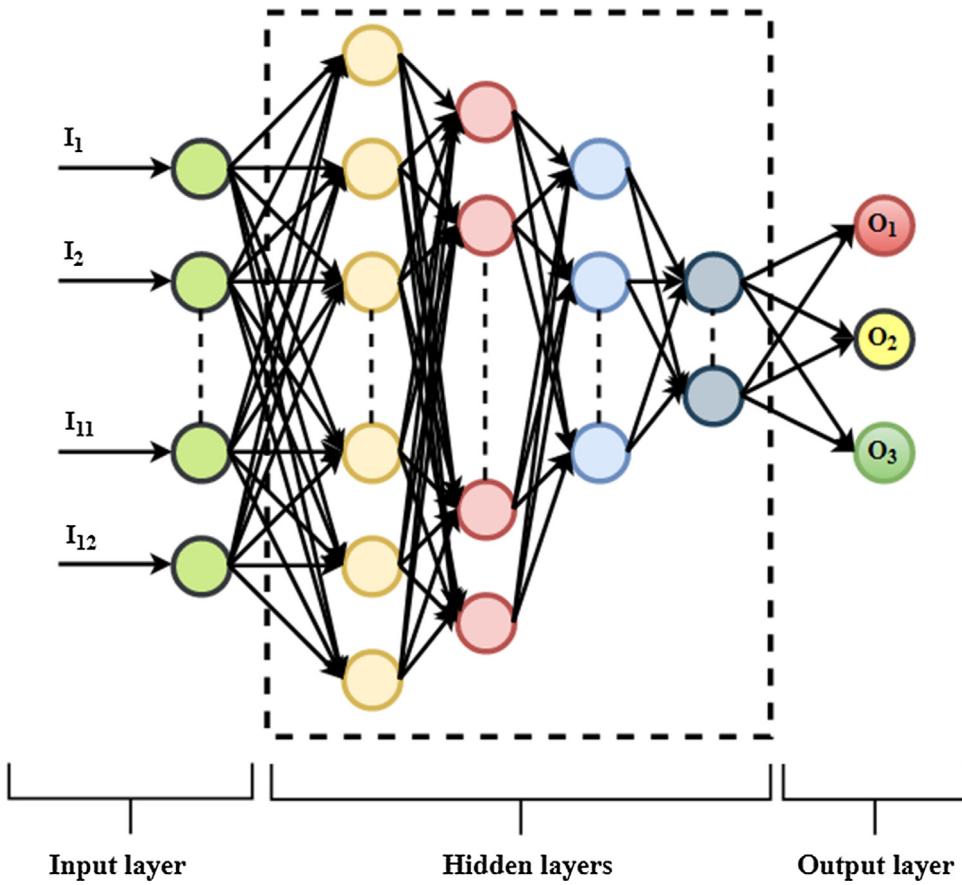


Fig. 4. Feature Importance plot considering XGBoost.



**Fig. 5.** Schematic representation of DNN model.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3)$$

iv. Kurtosis: Kurtosis, the word derived from the Greek word "Kurtos" means the distribution of the data in either of the tails in the distribution plot.

$$\text{Kurtosis} = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})}{n}}{\sigma^4} \quad (4)$$

v. Skewness: Skewness is defined as the asymmetry in the statistical distribution. It defines the extent to which a particular distribution differs from the normal distribution.

$$\text{Skewness} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n} \quad (5)$$

vi. Crest Factor: Crest factor is the ratio of the absolute of maximum value in the data to the root mean square of the data.

$$\text{Crest Factor} = \frac{|x_{\max}|}{\text{RMS}} \quad (6)$$

vii. Clearance Indicator: It is the ratio of the maximum absolute value of the signal and the square root amplitude value of the signal.

$$\text{Clearance Indicator} = \frac{\max(|x_i|)}{\left(\frac{1}{n} \sum_{i=1}^n |x_i|\right)^2} \quad (7)$$

viii. Shape Indicator: It is the ratio of RMS value of the signal to the mean of the signal.

$$\text{Shape indicator} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2}}{\text{mean}} \quad (8)$$

ix. Impulse Indicator: Impulse indicator is the ratio of the maximum absolute value of the signal and the absolute mean amplitude value of the signal.

$$\text{Impulse Indicator} = \frac{\max(|x_i|)}{\frac{1}{n} \sum_{i=1}^n |x_i|} \quad (9)$$

### 2.3. Feature ranking techniques

It is known that features contain information. In general perception, more is information, better is the discriminative power. However, this is not always true in data analysis because some of the features may be irrelevant. To improve the performance of practical classifiers, correlated or irrelevant features are required to be removed [26]. This reduction in the dimensionality of the feature matrix saves computation effort to develop DNN models. Thus, in this work, two different feature ranking methods, namely XGBoost and ReliefF criteria have been employed to reduce the dimensionality of the feature matrix. The reduced feature vector consists of four features obtained by reducing the extracted feature

vector by 55% for each sensor. These four features are chosen on the basis of their ranking in decreasing order considering the corresponding feature ranking technique treated to them. The two different feature ranking techniques applied in this study are briefly explained as follows:

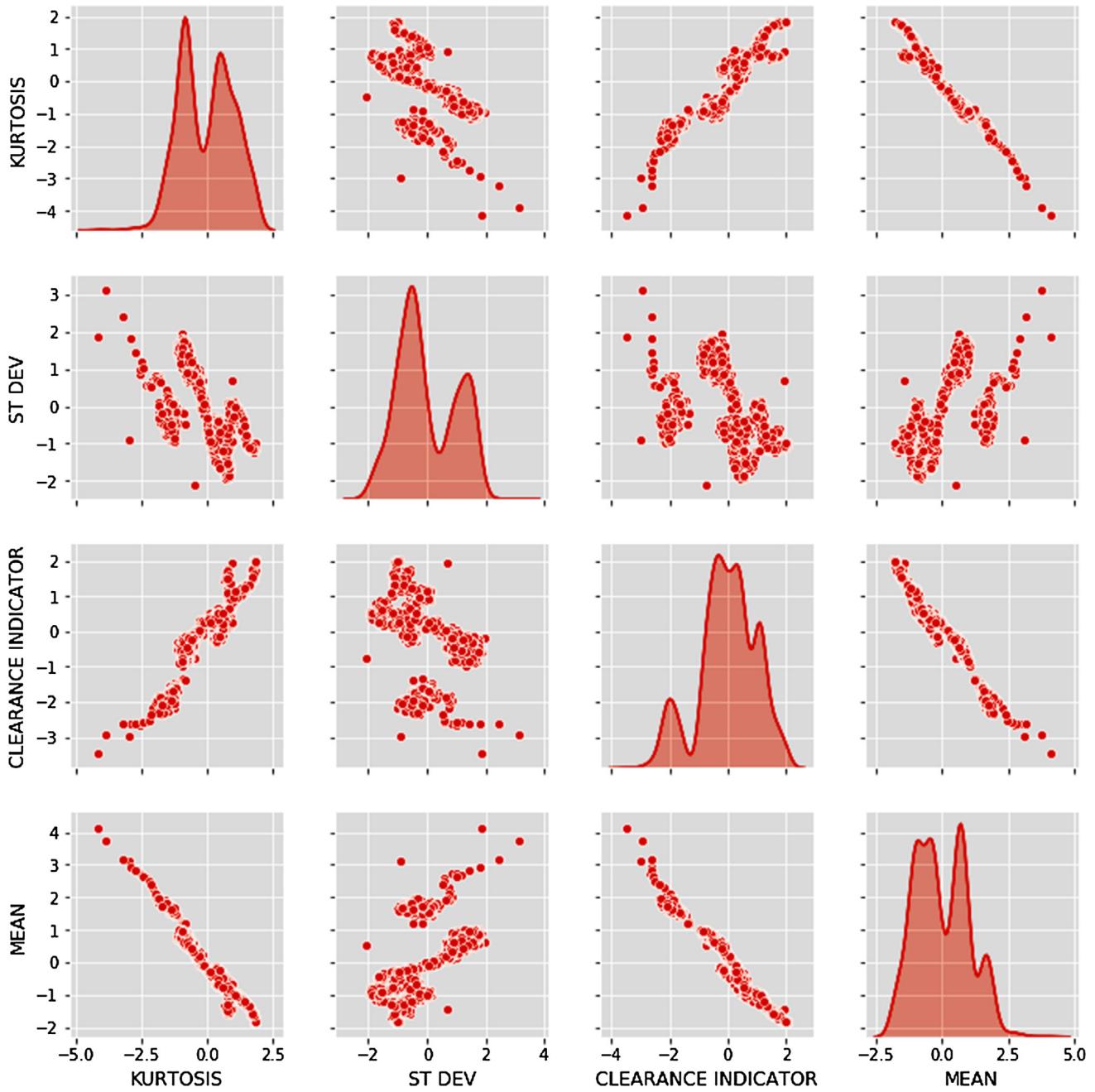
### 2.3.1. ReliefF

ReliefF is the specific evaluation filter algorithm that can detect the dependencies among features. It employs the concept of the nearest neighbour to derive the feature statistics inspired by Instance-Based learning [27,28].

ReliefF calculates the feature score by assigning a weight to variables based on how well they separate samples from their

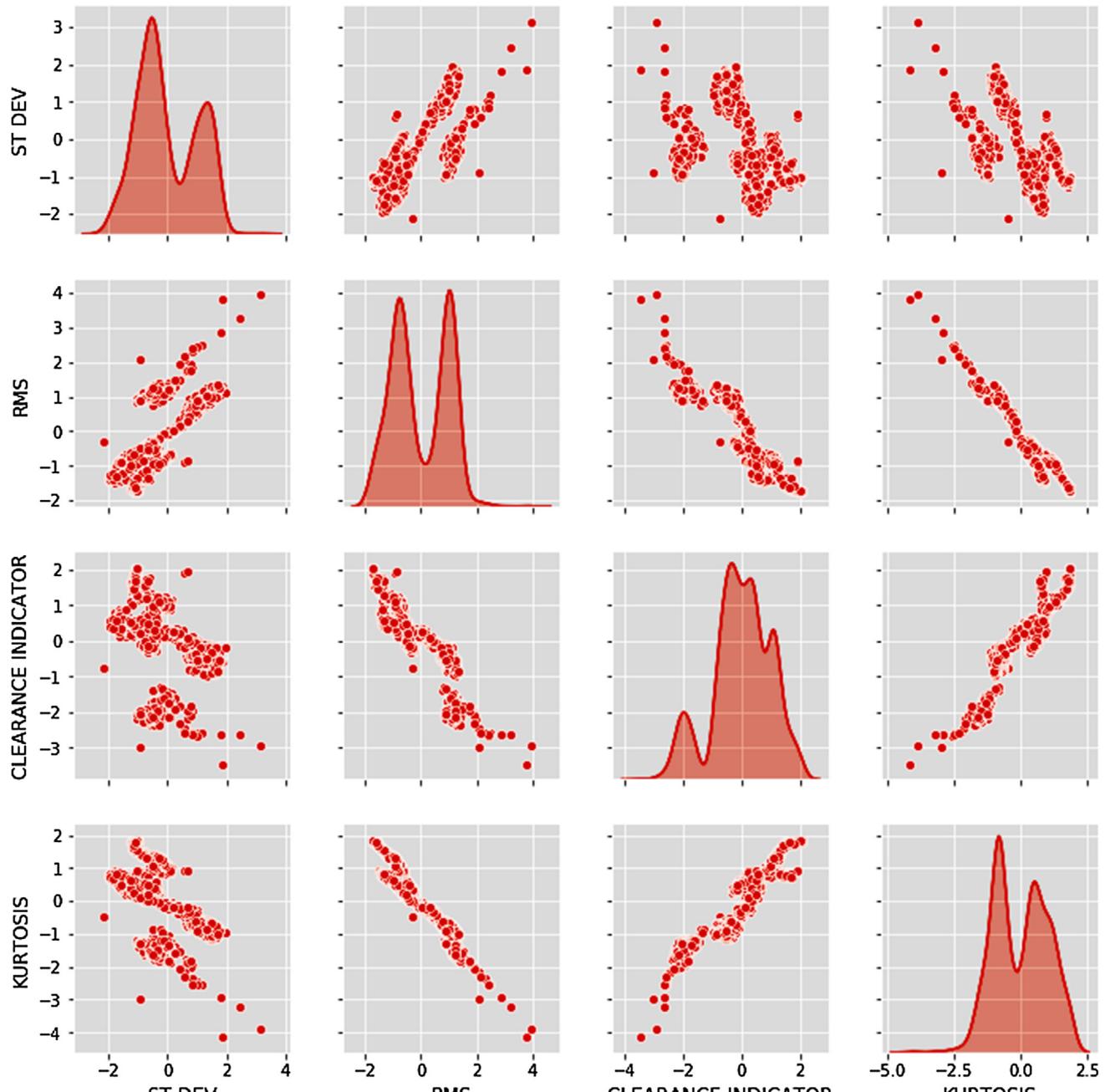
nearest neighbours from the same and from the opposite class. The algorithm casually picks an occurrence ( $R_i$ ) and then it hunts for  $k$  of its nearest neighbours within the identical class, known as nearest hits ( $H_j$ ) and from each of the different classes, known as nearest misses  $M_j(C)$ . The quality estimation  $W[A]$  is initially assigned to zero and updated [26,27]. For  $m$  number of iterations,  $W[A]$  is as mentioned in Eq. (10):

$$W[A] = W[A] - \sum_{j=1}^k \frac{diff(A, R_i, H_j)}{(m.k)} + \sum_{c \neq \text{class}(R_i)} \times \frac{\left[ \frac{P(c)}{1-P(\text{class}(R_i))} \sum_{j=1}^k diff((A, R_i, M_j(C))) \right]}{(m.k)} \quad (10)$$



(a)

**Fig. 6.** (a.) Dependency plot of XGBoost ranked features for Pressure Sensor 1. (b.) Dependency plot of ReliefF ranked features for Pressure Sensor 1. (c.) Dependency plot of XGBoost ranked features for Pressure Sensor 2. (d.) Dependency plot of ReliefF ranked features for Pressure Sensor 2. (e.) Dependency plot of XGBoost ranked features for Pressure Sensor 3. (f.) Dependency plot of ReliefF ranked features for Pressure Sensor 3.



(b)

Fig. 6 (continued)

The feature ranking output of three different pressure sensors using ReliefF is shown in Fig. 3. The bar chart shown in Fig. 3 is used to determine the four highest-ranked features according to ReliefF criteria. The x-axis contains the feature importance ranked on the scale of 10 whereas y-axis denotes extracted features. As shown in Fig. 3, the standard deviation shows higher relevancy for all three pressure sensors. Statistically, the standard deviation is the measure of variability. Thus, it indicates that the variation in the measured data is of much significance to classify faults.

### 2.3.2. XGBoost

Extreme gradient boosting (XGBoost) is an Ensembled technique that means combining multiple models either in parallel or in sequential. The very difference between Random Forest (RF)

and XGBoost method lies in the structure of decision trees (DT). In RF, DTs are made independently while in XGBoost trees are made in compliance with the already-built ones. Let the assumed data “ $D' = \{(x_i, y_i) : i = 1 \dots n, x_i \in R^m, y_i \in R\}$ ” denotes  $n$  observations and  $m$  features with corresponding variable  $y$ . For this dataset, a tree-ensembled method employs  $K$  additive functions to predict the output [29].

$$\hat{y} = \emptyset(x_i) = \sum_{k=1}^K f_k(x_i) \quad (11)$$

where,  $f_k \in \text{Regression Tree}$  and  $(q : R^m \rightarrow T)$ .  $q$  characterizes the structure of each tree mapping to the corresponding leaf index and  $T$  is the number of leaves in each tree. In order to choose the

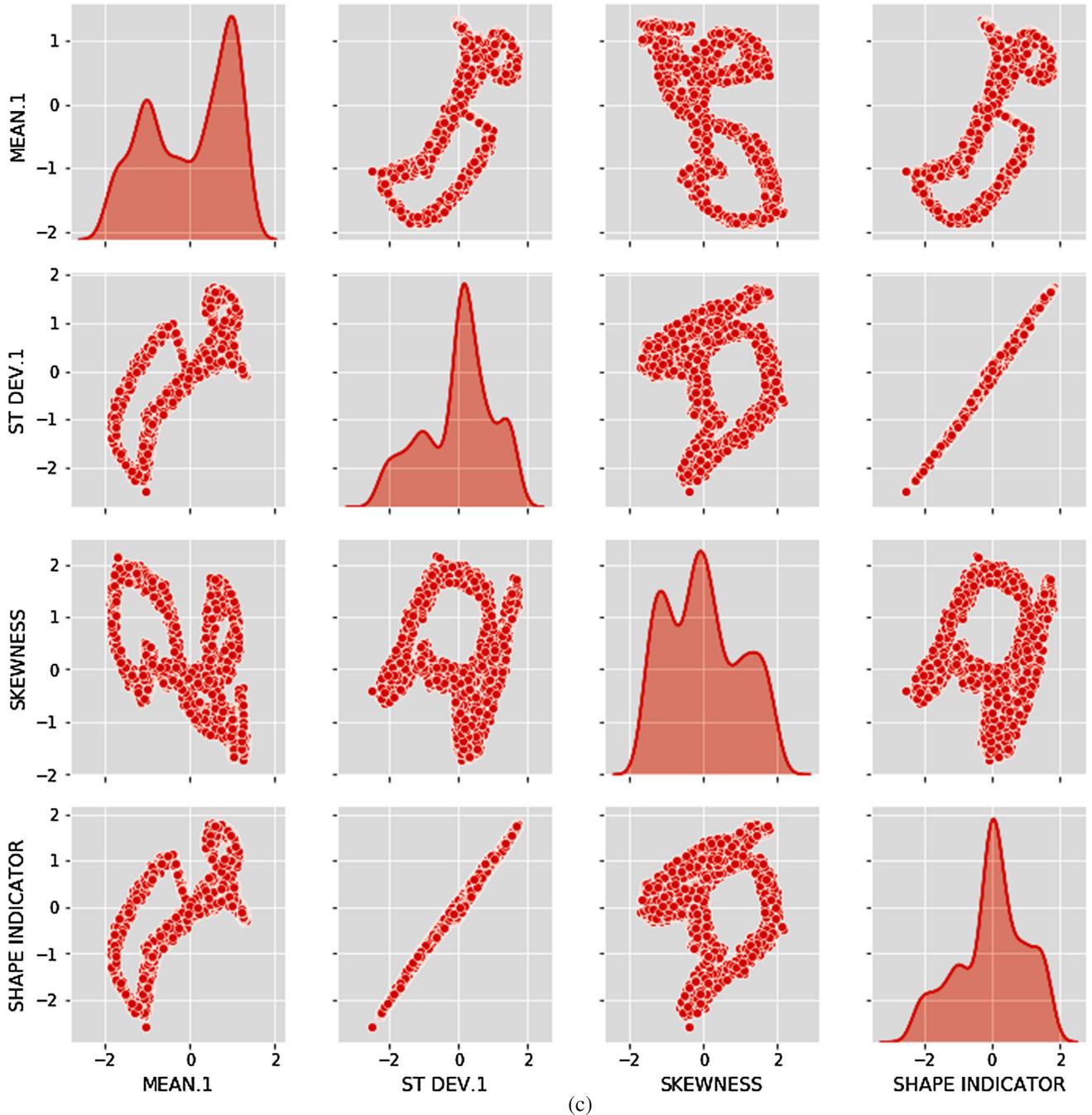


Fig. 6 (continued)

function  $f_k$ , regularised objective function  $L(\emptyset)$  is minimised, where  $L(\emptyset)$  is given by:

$$L(\emptyset) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega f_k \quad (12)$$

here,  $l$  is the differentiable convex loss function which measures the difference between the target value and predicted value and  $\Omega$  is the model penalty given by

$$\Omega f_k = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (13)$$

Model Penalty is the unique feature of the XGBoost which enhances the final weights during learning to avoid the overfitting of the model. To optimize the objective function in Eq.

(12), the traditional approach fails as it is the combination of parameters as well as the function. For  $n^{\text{th}}$  iteration, an additional term  $f_n$  is added for minimization. Thus, for  $n^{\text{th}}$  iteration:

$$L^n = \sum_i l(y_i, \hat{y}_i^{n-1} + f_n(x_i)) + \sum_k \Omega f_k \quad (14)$$

Second-order approximation and Taylor expansion are used to simplify Eq. (14) leading to loss reduction formula after tree split. Thus, XGBoost provides two additional features that prevent over-fitting, i) Weighing of each tree ii) Column sampling similar. The feature ranking output of three different pressure sensors using XGBoost is shown in Fig. 4.

The bar chart shown in Fig. 4 is used to determine the four highest-ranked features according to XGBoost criteria. The  $x$ -axis

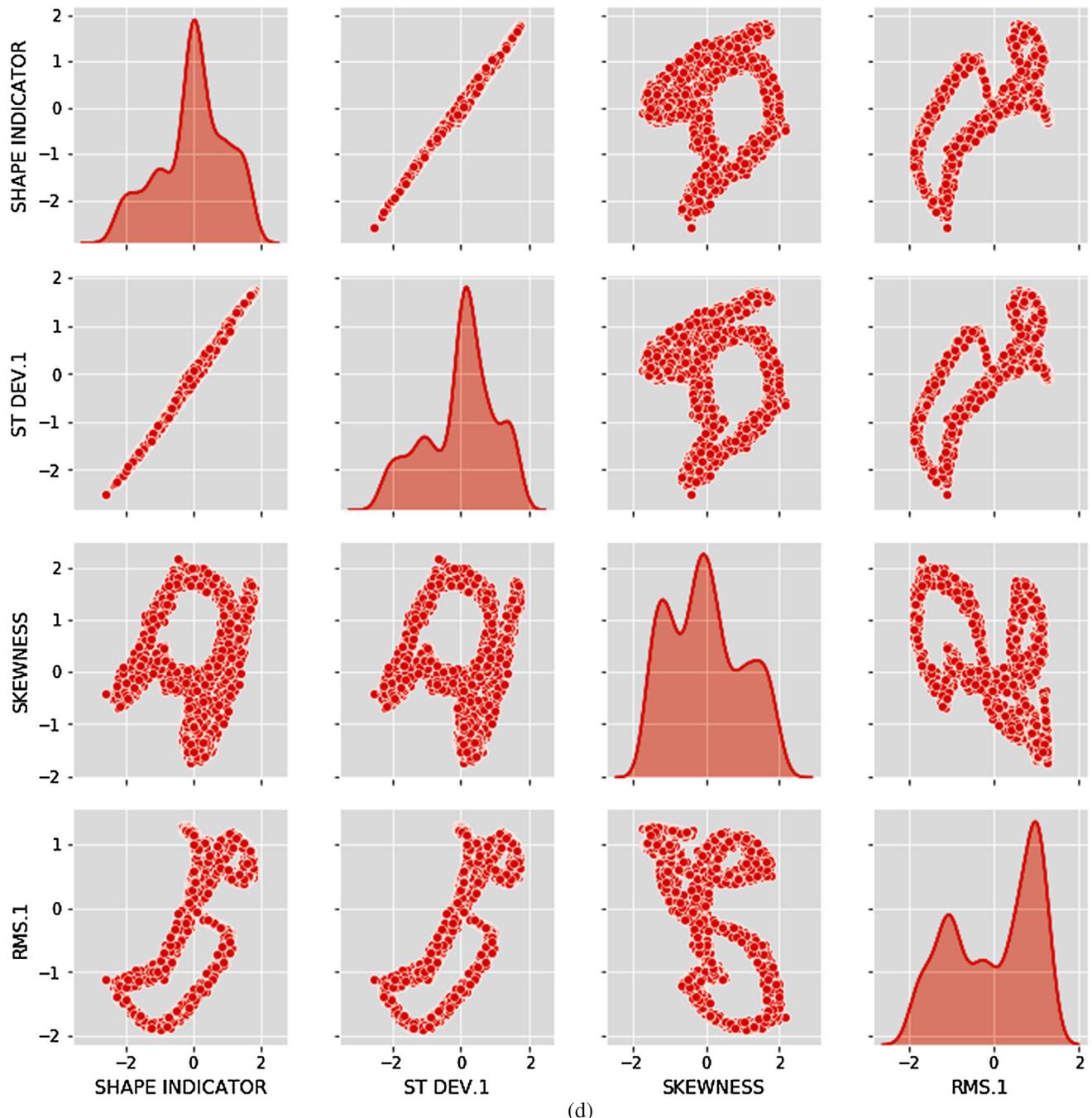


Fig. 6 (continued)

contains the feature importance ranked to the scale of 10 whereas *y*-axis denotes extracted features.

Skewness is the third statistical moment of the continuous random variable. It is the measure of the symmetry of the distribution of the random variable about its mean value [21]. As shown in Fig. 4, in the Pressure sensor 1, skewness is found to be of relatively lower significance, as the pressure values measured are not much distributed around the mean value whereas for pressure sensor 2 and pressure sensor 3, values are distributed in the close vicinity of the mean value of the pressure data. Thus, in pressure sensor 2 and pressure sensor 3, skewness is among the selected feature whereas for pressure sensor 1, skewness is not in the list of selected top four features.

### 3. Deep neural network for predicting cooler condition

Deep learning is the technique to learn underlying features in data using neural networks. A DNN is a feed-forward ANN with a certain level of complexity, having two or more than two hidden layers. It employs mathematical modelling in order to process data in complex ways by using a set of algorithms. High-level abstractions in data can be modelled using DNN architectures composed of multiple non-linear transformations. The multiple layers in DNN provide the way for feature extraction and transformation. Respectively successive layer receives output from the previous layer as an input. The learning of the deep structured and unstructured representation of data can be performed using DNN to

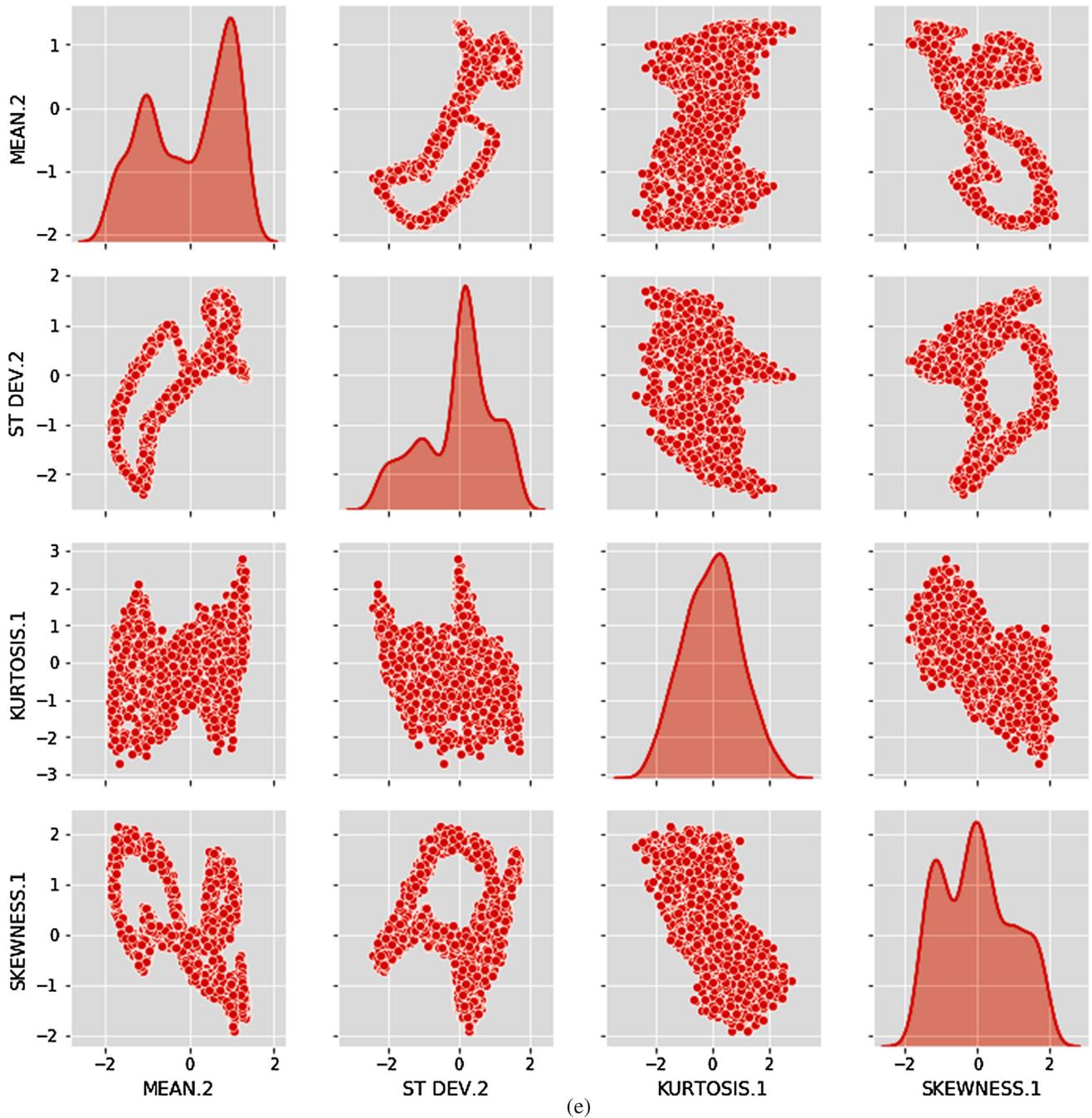


Fig. 6 (continued)

achieve an optimized solution using a machine learning algorithm to solve the problems. On the other hand, ANN, based on the neural structure of the brain are relatively crude electronic models which basically learns from experience. Such modelling promises a less technical way to develop machine learning solutions. Moreover, for a large dataset, the DNN is proved to be more efficient than ANN.

The schematic view of a generalized DNN model is shown in Fig. 5. In this study, a DNN with one input layer (12 input nodes), four different hidden layers consisting of 200, 150, 100 and 50 nodes respectively and one output layer consisting of 3 nodes for three different output as “Approaching towards total failure” ( $O_1$ ), “Working at reduced efficiency”( $O_2$ ) and “Running at full-efficiency”

( $O_3$ ) respectively has been proposed. Each hidden layer consists of a defined number of nodes and the logistic function, also called an *Activation function*.

Before training the model, input feature data is segregated as the training data and the testing data in the ratio of 4:1. The training feature matrix is fed as input to the fully connected (dense) DNN model. In the output layer, the output yields one of three values ( $O_1$ ,  $O_2$  and  $O_3$ ). Once the model is trained, testing data (signal) is fed to the model for predicting the condition of the system.

Further, in this study, the performance capabilities of two different activation functions namely “hyperbolic tangent (*tanh*)” and “Rectified Linear Unit (*relu*)” are compared on the DNN model using the same input data maintaining the same training to testing

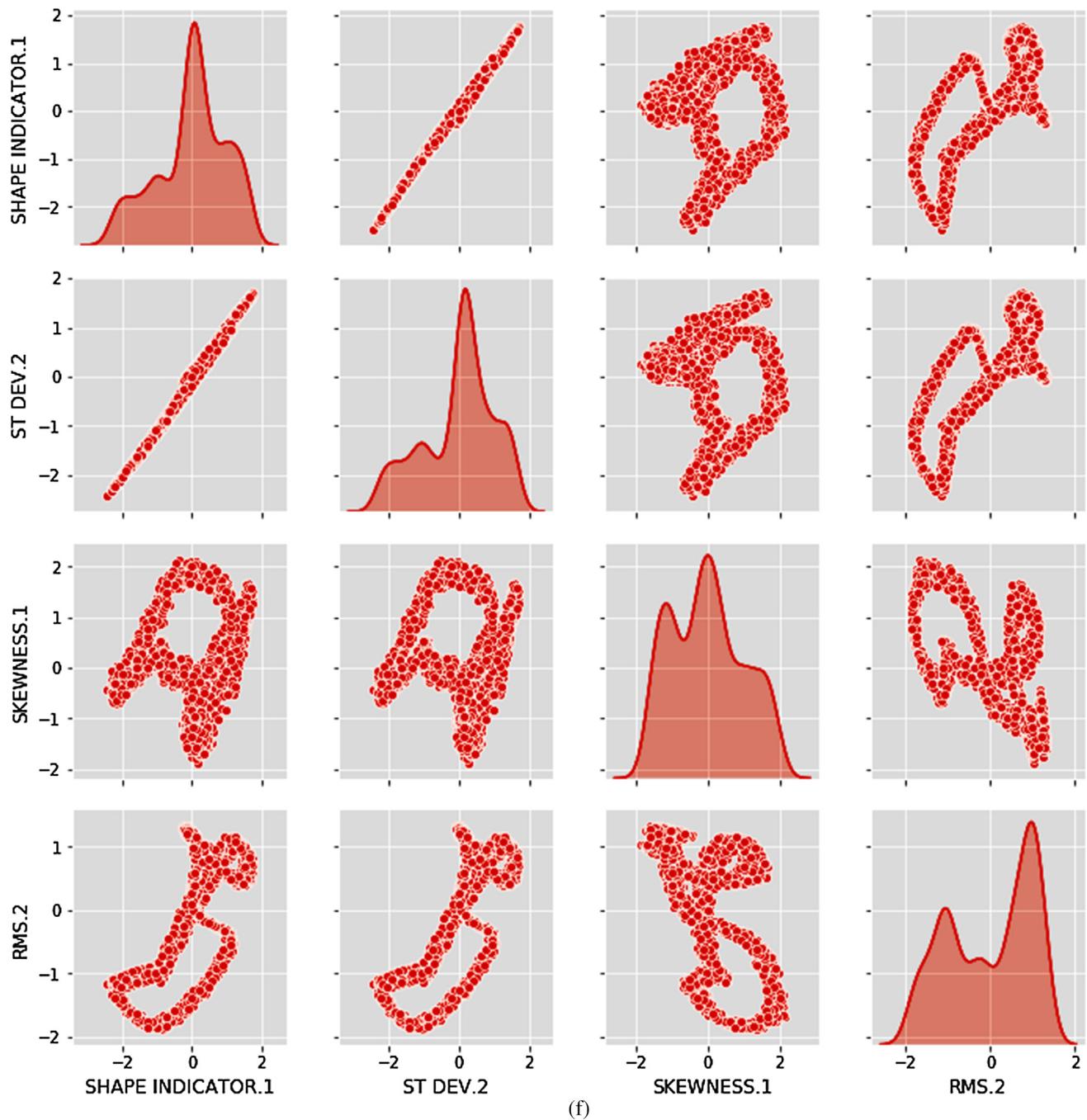


Fig. 6 (continued)

ratio. While compiling the model, the error of the model is calculated using the Mean squared error (*MSE*). The mathematical equation to estimate the *MSE* is given by:

$$MSE = \frac{1}{n} (y_i - \hat{y}_i)^2 \quad (15)$$

where,  $y_i$ =Predicted Value and  $\hat{y}_i$  = Expected Value

#### 4. Results and discussion

Fig. 6 shows the relationship among the top four ranked features according to XGBoost and ReliefF criteria for all three different pressure sensors respectively. Dependency plots of significant features are plotted relative to each other and to itself also. Observations include histograms when continuous data is compared to itself. Most of the histograms (except Kurtosis for Pressure Sensor 3) can be inferred as Bimodal histogram. It means that the data is distributed about two peak points in the range. Further, it can also be concluded that no such pre-defined relationship curve is obtained in most of the relative plots, so these chosen features are independent in nature and can affect the output of the DNN model. Fig. 6(a.) shows the dependency plots of four selected features namely kurtosis, standard deviation, clearance indicator and mean for the pressure sensor 1 in accordance with XGBoost criteria. The plot concludes either zero or very less dependency of features on each other, thus can be termed as independent to each other and can affect the output of the prediction using DNN model. Similarly, four features namely standard deviation, root

vations include histograms when continuous data is compared to itself. Most of the histograms (except Kurtosis for Pressure Sensor 3) can be inferred as Bimodal histogram. It means that the data is distributed about two peak points in the range. Further, it can also be concluded that no such pre-defined relationship curve is obtained in most of the relative plots, so these chosen features are independent in nature and can affect the output of the DNN model. Fig. 6(a.) shows the dependency plots of four selected features namely kurtosis, standard deviation, clearance indicator and mean for the pressure sensor 1 in accordance with XGBoost criteria. The plot concludes either zero or very less dependency of features on each other, thus can be termed as independent to each other and can affect the output of the prediction using DNN model. Similarly, four features namely standard deviation, root

**Table 1**

Summary of the model using Deep Neural Network for cooler health prediction.

Layer (Type)	Output Shape	Parameter
Dense_1	(None, 200)	2600
Dense_2	(None, 150)	30,150
Dense_3	(None, 100)	15,100
Dense_4	(None, 50)	5050
Dense_5	1	51
Total Parameters: 52,951		
Trainable Parameters: 52,951		
Non-Trainable Parameter: 0		

mean square, clearance indicator and kurtosis are ranked highest considering ReliefF criteria as shown in Fig. 6(b.). In Fig. 6(c.), mean, standard deviation, skewness and shape indicator are among the top selected features complying with XGBoost feature selection criteria for pressure sensor 2. In this, standard deviation and shape indicator show some little bit extent of linear dependency, rest of the features are highly independent. Similarly, ReliefF feature ranking technique filters shape indicator, standard deviation, skewness and root mean square as the four highest-ranked features as shown in Fig. 6(d.). Mostly all the features are highly independent except the petite interdependency between standard deviation and shape indicator. In Fig. 6(e.), mean, standard deviation, kurtosis and skewness are among the four highest-ranked features considering XGBoost featuring ranking technique for pressure sensor 3. The relative plot of kurtosis with respect to kurtosis shows unimodal histogram, it signifies that

the fourth moment of the pressure data for sensor 3 is distributed about the central mean line. Fig. 6(f.) illustrates the shape indicator, standard deviation, skewness and root mean square as the four carried forward features when treated with ReliefF feature ranking criteria.

In present work, two different DNN models have been built using “tanh” and “relu” for each feature ranking method. The DNN model summary is mentioned in Table 1. Table 1 describes the type of layers, output shape and parameters. The output shape shows “none” and the number of nodes in the corresponding layers. Here, “none” signifies dynamic dimension of the batch, i.e. model can be fitted with any batch size.

#### 4.1. DNN using activation function “tanh”

Hyperbolic tangent (*tanh*) is the solution of the differential equation  $f' + f^2 - 1 = 0$  subjected to  $f(0) = 0$  and nonlinear boundary value problem  $\frac{1}{2}f'' - f^3 + f = 0$  subjected to  $f(0) = f'(\infty) = 0$ . The activation function “tanh” is employed for each hidden layer of the DNN model.

The accuracy achieved with features ranked in accordance with ReliefF and XGBoost criteria using “tanh” as an activation function is shown in Fig. 7 and 8. These graphs show as the Epochs (Iterations) increases, the accuracy of the model enhances and after approximately 100 epochs it attains higher accuracy. When testing the data, training set and testing set tends to follow each other which signifies the good accuracy of the model in testing the data.

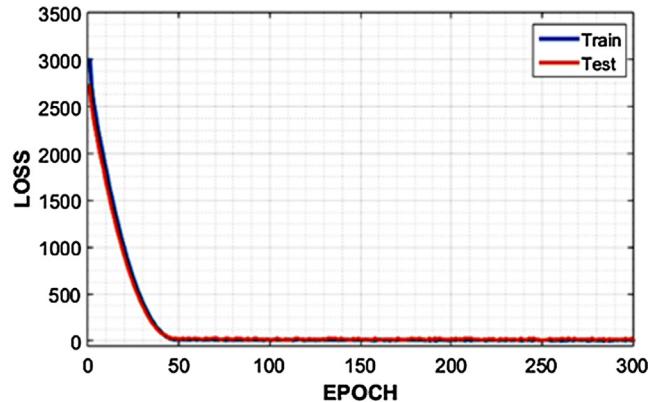
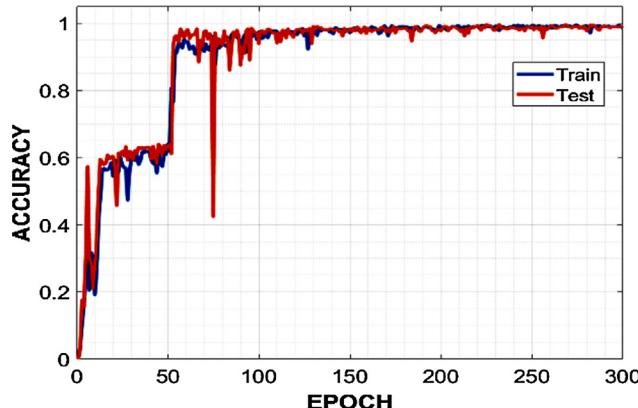


Fig. 7. Model accuracy and losses with tanh using ReliefF ranked features.

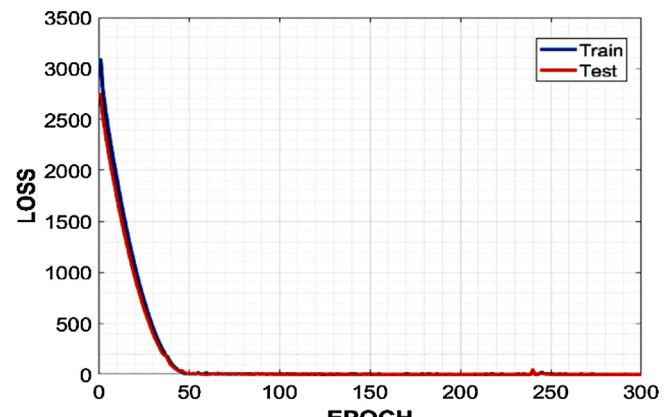
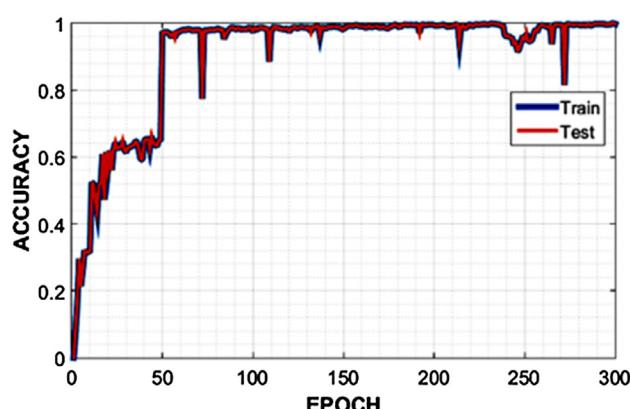
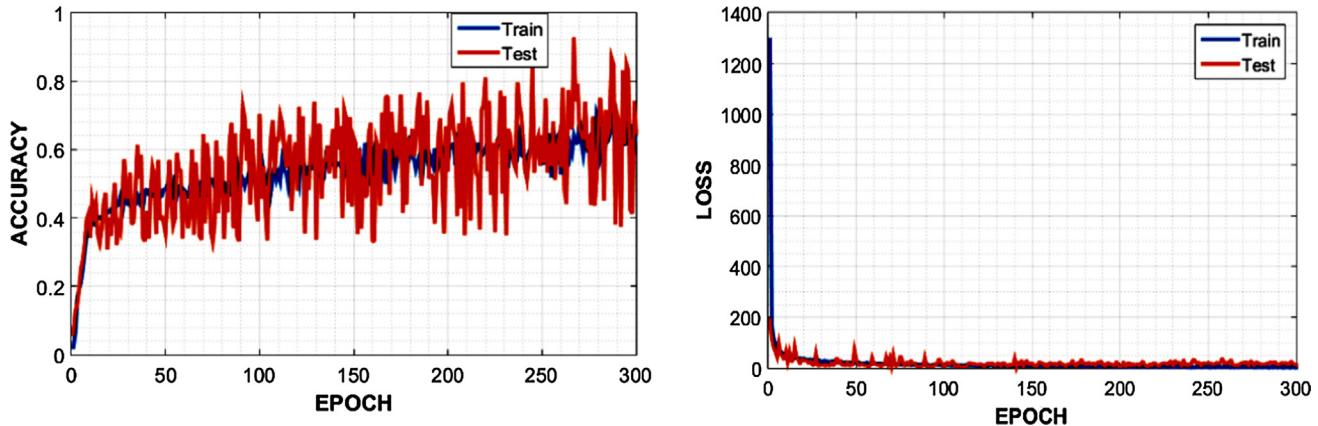
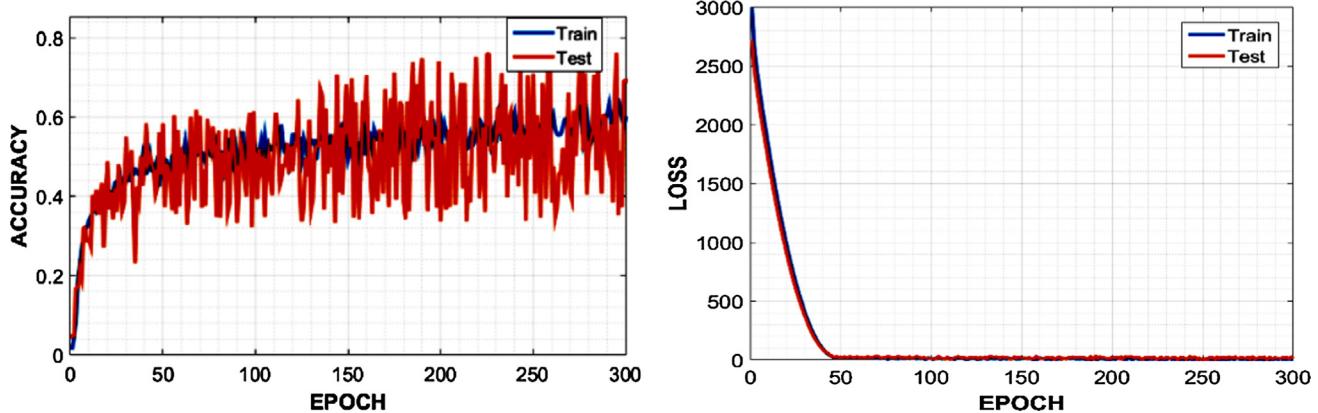


Fig. 8. Model accuracy and losses with tanh using XGBoost ranked features.



**Fig. 9.** Model accuracy and losses with *relu* using ReliefF ranked feature.



**Fig. 10.** Model accuracy and losses with *relu* using XGBoost ranked feature.

At the same time, the model loss is decreased and tends to zero, which signifies Mean squared error is minimised.

#### 4.2. DNN using activation function “*relu*”

The “*relu*” function represented by  $y = \max(0, x)$  is piece wise linear function that acts very firmly in most of the neural network but it discards the negative value of the input and produce the constant output as zero. The activation function “*relu*” is employed for each hidden layer of the DNN model. Fig. 9 and 10 shows the accuracy achieved by the model using “*relu*” as an activation function for models trained with features ranked in accordance to ReliefF and XGBoost criteria respectively.

Table 2 compares the accuracies achieved for four different models using two different activation functions combined with differently ranked features. It shows highest accuracy of 99.54% is achieved with XGBoost ranked features applied to DNN with “*tanh*” as an activation function and 97.95% accuracy is noticed for a model with ReliefF ranked features applied to DNN with “*tanh*” as an activation function. While modelling the neural network for health prediction of the cooling circuit, employing “*relu*” is not found to be useful as the pressure vary significantly. Thus, there is a need for an alternate activation function. Two different activation functions i.e. “*sigmoid*” and “*tanh*” show almost same behaviour, but the “*sigmoid*” function has the disadvantage of developing gradient descent at a higher rate than “*tanh*”, thus, “*sigmoid*” is not used. Moreover, activation function “*tanh*” maps the inputs in positive as well as negative with dense activation. The

**Table 2**  
Accuracies of various models.

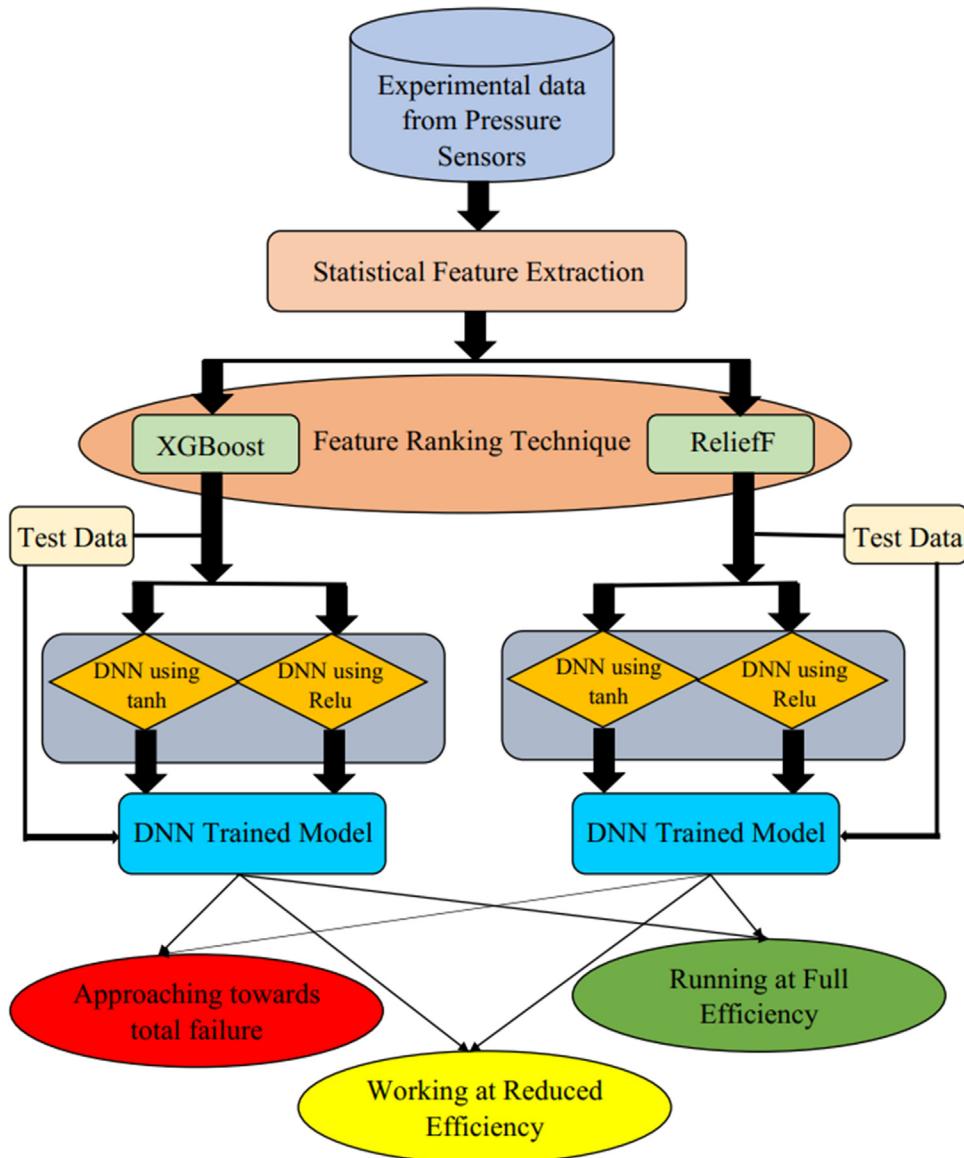
Feature Ranking Technique	Activation Function	Accuracy (%)
XGBoost	<i>tanh</i>	99.54
	<i>relu</i>	74.37
ReliefF	<i>tanh</i>	97.95
	<i>relu</i>	59.18
Total Parameters: 52,951		
Trainable Parameters: 52,951		
Non-Trainable Parameter: 0		

overall process of cooling circuit health prediction is shown in Fig. 11.

## 5. Conclusions

In this study, the performance of two different activation functions along with two different feature ranking techniques has been compared for signals received from installed three different pressure sensors. Following conclusions are drawn from this study:

- The health monitoring of the hydraulic cooling circuit can effectively be performed by analysing the pressure signals using a deep neural network.
- XGBoost can successfully be used as the feature ranking technique and found to be sorting more relevant features as compared to the features sorted by using ReliefF.



**Fig. 11.** Flow chart of methodology adapted for cooling circuit health prediction.

- c. Features ranked with higher importance using XGBoost criteria combined with activation function “*tanh*” yields the highest accuracy as seen in Fig. 8. The training and testing results almost overlap each other. It shows their agreement with each other and thus providing very high accuracy.
- d. The present study shows that the deep neural network with XGBoost ranked features can reduce the dimensionality of data without sacrificing the accuracy of fault detection in the hydraulic system components.
- e. The detailed analysis shows that for this type of data, “*tanh*” activation function yields better accuracy as it shows a close pattern to the pressure signal. Hence, model accuracy is enhanced as well as losses are minimised.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### References

- [1] F. Hoblit, Critical buckling for hydraulic actuating cylinders. Stress Engineer, Lockheed Aircraft Corporation. Product Engineering, 1950, 108–112.
- [2] A. Saxena, A. Saad, Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems, *Appl. Soft Comput.* 7 (1) (2007) 441–454.
- [3] A. Davies (Ed.), *Handbook of Condition Monitoring: Techniques and Methodology*, Springer Science & Business Media, 2012.
- [4] B.K.N. Rao (Ed.), *Handbook of Condition Monitoring*, Elsevier, 1996.
- [5] F.P.G. Márquez, A.M. Tobias, J.M.P. Pérez, M. Papaelias, Condition monitoring of wind turbines: techniques and methods, *Renewable Energy* 46 (2012) 169–178.
- [6] P.A. Higgs, R. Parkin, M. Jackson, A. Al-Habaibeh, F. Zorriassatine, J. Coy, November). A survey on condition monitoring systems in industry, in: ASME 7th Biennial Conference on Engineering Systems Design and Analysis, American Society of Mechanical Engineers Digital Collection, 2008, pp. 163–178.
- [7] M. Moradi, A. Chaibakhsh, A. Ramezani, An intelligent hybrid technique for fault detection and condition monitoring of a thermal power plant, *Appl. Math. Model.* 60 (2018) 34–47.
- [8] P. Jatin, N. Aniket, P.K. Kankar, V.K. Gupta, P.K. Jain, T. Ravindra, M. Ismail, Estimation of load carrying capacity for pin-mounted hydraulic cylinders, in: *Advances in Engineering Design*, Springer, Singapore, 2019, pp. 173–185.
- [9] Y. Gao, Q. Zhang, X. Kong, 2326681. Wavelet-based pressure analysis for hydraulic pump health diagnosis, *Trans. ASAE* 46 (4) (2003) 969–976.

- [10] P.K. Kankar, S.C. Sharma, S.P. Harsha, Fault diagnosis of ball bearings using continuous wavelet transform, *Appl. Soft Comput.* 11 (2) (2011) 2300–2312.
- [11] L.M. Hu, K.Q. Cao, H.J. Xu, Fault diagnosis for hydraulic actuator based on support vector regression, *J. System Simulation* (2007) 23.
- [12] R. Jegadeeswaran, V. Sugumaran, Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines, *Mech. Syst. Sig. Process.* 52 (2015) 436–446.
- [13] K. Mollazade, H. Ahmadi, M. Omid, R. Alimardani, An intelligent combined method based on power spectral density, decision trees and fuzzy logic for hydraulic pumps fault diagnosis, *Int. J. Intell. Syst. Technol.* 3 (4) (2008) 251–263.
- [14] E.P. Carden, P. Fanning, Vibration based condition monitoring: a review, *Struct. Health Monit.* 3 (4) (2004) 355–377.
- [15] S.W. Doebling, C.R. Farrar, M.B. Prime, A summary review of vibration-based damage identification methods, *Shock Vib. Digest* 30 (2) (1998) 91–105.
- [16] P.K. Kankar, S.C. Sharma, S.P. Harsha, Fault diagnosis of ball bearings using machine learning methods, *Expert Syst. Appl.* 38 (3) (2011) 1876–1886.
- [17] V. Vakharia, V.K. Gupta, P.K. Kankar, A comparison of feature ranking techniques for fault diagnosis of ball bearing, *Soft. Comput.* 20 (4) (2016) 1601–1619.
- [18] J. Liu, Y. Hu, B. Wu, C. Jin, A hybrid health condition monitoring method in milling operations, *Int. J. Adv. Manuf. Technol.* 92 (5–8) (2017) 2069–2080.
- [19] A. Sharma, M. Amarnath, P.K. Kankar, Novel ensemble techniques for classification of rolling element bearing faults, *J. Braz. Soc. Mech. Sci. Eng.* 39 (3) (2017) 709–724.
- [20] R. Upadhyay, P.K. Padhy, P.K. Kankar, A comparative study of feature ranking techniques for epileptic seizure detection using wavelet transform, *Comput. Electr. Eng.* 53 (2016) 163–176.
- [21] F. Honarvar, H.R. Martin, New statistical moments for diagnostics of rolling element bearings, *J. Manuf. Sci. Eng.* 119 (3) (1997) 425–432.
- [22] A. El-Betar, M.M. Abdelhamed, A. El-Assal, R. Abdelsatar, Fault diagnosis of a hydraulic power system using an artificial neural network, *Eng. Sci.* 17 (1) (2006).
- [23] R. Ahmed, M. El Sayed, S.A. Gadsden, J. Tjong, S. Habibi, Automotive internal-combustion-engine fault detection and classification using artificial neural network techniques, *IEEE Trans. Veh. Technol.* 64 (1) (2014) 21–33.
- [24] L. Li, Y. Huang, J. Tao, C. Liu, Internal leakage identification of hydraulic cylinder based on intrinsic mode functions with random forest, *Proc. Inst. Mech. Eng., C* (2019), 0954406219846148.
- [25] N. Helwig, E. Pignanelli, A. Schütze, Condition monitoring of a complex hydraulic system using multivariate statistics, in: 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, IEEE, 2015, pp. 210–215.
- [26] A. Sharma, M. Amarnath, P.K. Kankar, Use of feature ranking techniques for defect severity estimation of rolling element bearings, *Int. J. Acoust. Vib.* 23 (2018) 49–56.
- [27] R.J. Urbanowicz, M. Meeker, W. La Cava, R.S. Olson, J.H. Moore, Relief-based feature selection: introduction and review, *J. Biomed. Inform.* 85 (2018) 189–203.
- [28] Z. Wang, Y. Zhang, Z. Chen, H. Yang, Y. Sun, J. Kang, et al., Application of ReliefF algorithm to selecting feature sets for classification of high resolution remote sensing image, in: 2016 IEEE international geoscience and remote sensing symposium (IGARSS), 2016, IEEE, pp. 755–758.
- [29] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, ACM, pp. 785–794.