Journal Paper

“Principal component analysis technique for early fault detection”

Journal of Intelligent & Fuzzy Systems

2021

Kumari Sarita,
Electrical Engineering Department, BIT Sindri, Dhanbad, Jharkhand, India

Ramesh Devarapallia
Electrical Engineering Department, BIT Sindri, Dhanbad, Jharkhand, India
ramesh.ee@bitsindri.ac.in

Sanjeev Kumarb
Tata Steel, Jamshedpur, Jharkhand, India

Hasmat Malik
BEARS, University Town, NUS Campus, Singapore

Fausto Pedro García Márquez
Ingenium Research Group, Universidad de Castilla-La Mancha
FaustoPedro.Garcia@uclm.es

Pankaj Raia
Electrical Engineering Department, BIT Sindri, Dhanbad, Jharkhand, India

Cite as: Sarita, K., Devarapalli, R., Kumar, S., Malik, H., Márquez, F. P. G., & Rai, P. (2021). Principal component analysis technique for early fault detection. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-12.

DOI: 10.3233/JIFS-189755

Principal component analysis technique for early fault detection

Kumari Sarita[a], Ramesh Devarapalli[a*], Sanjeev Kumar[b], Hasmat Malik[c], Fausto Pedro García Márquez[d], and Pankaj Rai[a]

[a] *Electrical Engineering Department, BIT Sindri, Dhanbad, Jharkhand, India*
[b]*Tata Steel, Jamshedpur, Jharkhand, India*
[c]*BEARS, University Town, NUS Campus, Singapore*
[d] *Ingenium Research Group, University of Castilla-La Mancha, Spain. FaustoPedro.Garcia@uclm.es*

E-mail: *ramesh.ee@bitsindri.ac.in

Abstract. Online condition monitoring and predictive maintenance are crucial for the safe operation of equipments. This paper highlights an unsupervised statistical algorithm based on principal component analysis (PCA) for the predictive maintenance of industrial induced draft (ID) fan. The high vibration issues in ID fans cause the failure of the impellers and, sometimes, the complete breakdown of the fan-motor system. The condition monitoring system of the equipment should be reliable and avoid such a sudden breakdown or faults in the equipment. The proposed technique predicts the fault of the ID fan-motor system, being applicable for other rotating industrial equipment, and also for which the failure data, or historical data, is not available. The major problem in the industry is the monitoring of each and every machinery individually. To avoid this problem, three identical ID fans are monitored together using the proposed technique. This helps in the prediction of the faulty part and also the time left for the complete breakdown of the fan-motor system. This helps in forecasting the maintenance schedule for the equipment before breakdown. From the results, it is observed that the PCA-based technique is a good fit for early fault detection and getting alarmed under fault condition as compared with the conventional methods, including signal trend and fast Fourier transform (FFT) analysis.

Keywords: Machine Learning; Industry 4.0; PCA; Condition Monitoring; Predictive Maintenance; Preprocessing.

I. Introduction

Generally, three types of maintenance are implemented in industries: reactive, preventive, and predictive. Reactive maintenance is the maintenance that is done once there is a problem detected, preventive maintenance is the doing maintenance at a regular rate, and predictive maintenance is related to forecasting when problems will arise. Predictive maintenance helps to predict the health condition of the equipment using its current and historical data and forecasts the need for maintenance of the equipment. There are various benefits of doing predictive maintenance : (i) It increases the uptime and safety which results in the reliability of the equipment and the overall industry, (ii) It minimizes the maintenance costs which results in the cost of ownership, and (iii) It optimizes the supply chain which results in the reputation of the industry, etc. In literature, many researchers have discussed the concept and benefits of predictive maintenance of industrial equipment [11-14]. The authors [11] have used the industrial internet of things (IIOT) for implementing the proactive model. The industrial equipment generally degrades due to high vibration, temperature, and pressure issues. To avoid or reduce this degradation rate of the equipment, and condition monitoring is done. The condition monitoring of equipment is done in some cases to diagnose the root cause of any fault in the equipment. For reliable condition monitoring, predictive and preventive maintenance plays an important role. Preventive and predictive maintenance help to predict the future coming faults of the equipment which can be avoided, and the sudden breakdown of the equipment is also avoided. For such reliable maintenance, machine learning and data science are implemented to develop a predictive model [1-3]. Horrell et al. [3] have discussed various analytical problems, reliability and predictive maintenance of industrial equipment using the data of the equipment by implementing data analysis. Vibration is an important parameter for condition monitoring of any rotating equipment because it indicates the working condition and performance of large rotating equipment and also highly correlated with other parameters of the equipment . Gholap and Jaybhaye [4] have implemented FFT method for fault diagnosis (FD) and for this the authors have used vibration-based condition monitoring. The vibration signal of the equipment gives the condition of system and also ables to predict the life of the equipment which helps in implementing predictive maintenance of the equipment. Wang et al. [5] have implemented wavelet transform (WT) for FD of wind turbine gearbox using vibration signal of the gearbox. These techniques [4, 5] are very accurate to indicate the fault condition but indication is found when the fault has occurred but not before. In reference [6], time domain, frequency domain and time-frequency domain techniques were used for vibration-based condition monitoring are discussed. The authors observed that the wavelet-based methods and time-frequency analysis method along with machine learning technique are most effective techniques for FD. The authors also demostrated that the vibration-based FD technique gives better results of condition monitoring. In industries, vibration spectrum analysis based preventive and predictive maintenance technology is used . The peak level of vibration indicates fault if increases beyond threshold value and the frequency spectrum indicate particular fault type (unbalance, misalignment, looseness, or damage in any part) that has occurred in the equipment.

The maintenance and operation of equipment and plant should be cost-effective. The various commonly used techniques of condition monitoring have been explained by the authors [7-9]. [9] Fahmy [9] has explained reliability-centred maintenance (RCM), which is a methodology for deciding whether the maintenance is needed or not in the equipment. In condition-based maintenance , the three decisions are needed: (i) parameter selection, (ii) inspection interval, (iii) setting the warning limit. The first depends upon the type of equipment to be monitored and the parameters available, the second depends upon the frequency of fault occurrence and time between consequent failures which is determined and inspected from the historical data of the equipment. and the third one depends upon the threshold limit set according to the vibration severity charts of International Organization for Standardization (ISO) standard. Various decision models have also been studied by the authors [7, 8]. There are three types of decisions which need to be made in condition-based maintenance: (1) selecting the parameters to be monitored; (2) determining the inspection frequency; and (3) establishing the warning limit (the trigger). Once the fault is detected, or vibration is increased beyond the set limit, the machine learning-based algorithm and deep learning as discussed in [10]. Other than vibration, the temperature is also an essential indicator of the structural condition of the equipment. If there is any fault, corroded connecting part, or damaged part, then there would be temperature rise in some portions of the equipment.

The recent developments and facilities in big data collection, cloud storage and computation and Internet of Things (IOT) help to make industry 4.0 more popular. In predictive maintenance of industrial equipment, the first step is the collection of data for the data-driven maintenance. The collected data is further used in different predictive algorithms using statistics and machine learning techniques. The digital transformation of industries results in industry 4.0 revolution with the help of artificial intelligence, data science, machine learning, cloud computing, and IOT [15,16]. The digital transformation is making industries smart and more reliable. The authors [15, 16] have explained the process and need of movement from IOT to IIOT. This facilitates the movement of industries towards the revolution of industry 4.0 and makes it smart. Predictive maintenance is not only developing machine learning algorithms, but it is the complete process of data collection, cloud storage, data processing, result visualization of condition monitoring and giving a report of the real-time status of the equipment and predicting future faults. The digital transformation of complex equipment system is a complex work, therefore, to make sure that the developed predictive monitoring system is reliable and to obtain optimal system operation, optimization techniques play an important role. Preventive maintenance scheduling using optimization technique is discussed in [17]. This increases the production efficiency of the industry. The authors have used the historical data, including the various loadings of the equipment, manufacturing data, availability and cost. The supervised machine learning techniques are applicable when the fault data of the equipment is available. The major difficulty in supervised machine learning methods is in the feature extraction process. In the literature, there are various techniques proposed to achieve effective features with the least possible data loss [18, 19]. Alweshah et al. [18] have highlighted the wrapper genetic programming (WGP) approach for feature extraction. In comparison with other fuzzy logic-based approaches, WGP is found to give better results. The size of attributes or data-set affects the accuracy of the classifier. To boost the accuracy of classifier, various algorithms are available in the literature, as discussed in [20-22].

1.1. Principal Component Analysis:

PCA is a type of unsupervised machine learning technique. This technique is preferred when one does not have failure data of the equipment. This is also useful in reducing the overfitting problem when there is a large number of parameters in the data set. The number of principal components (PCs) can be less than or equal to the number of attributes. Among the principal components, the first principal component (PC1) is given the highest priority and goes on decreasing for PC2, PC3, and so on. The PCs are independent of each other, and they are orthogonal.

It is a challenge for industries to select the best method among the new algorithms and methodologies of different learning techniques, therefore, the fault detection can be performed with accuracy. The most critical step in predictive maintenance for industries is to detect the fault accurately and early. The authors [23, 24] have explained the benefits of PCA. It is one of the most prominent and important algorithms that are commonly used these days. In literature, there are other similar techniques also available which uses PCA for fault detection, including Kernel PCA, Q-statistic, T2-statistics, independent component analysis (ICA), and multiscale neighbourhood normalization-based multiple dynamic principal component analysis (MNNMDPCA) method. Clustering method [25] is another commonly used method. The methods based on subtractive clustering, and K-means are discussed in [25]. The authors have discussed different algorithms such as PCA with T2 statistic, Hierarchical clustering, K- Means clustering, CMeans, and Model-based clustering for fault detection by the authors in [27-29]. Among the methods available in the literature, PCA is comparatively simple, has less computational complexity and capable of predicting the fault condition accurately as compared with the other unsupervised techniques mentioned above in some cases. The simplest method to find PCs is to find the covariance matrix of the given data set, and after that, eigenvalues and eigenvectors of the covariance matrix are found from the characteristic equation. The eigenvalues are observed to find the highest eigenvalue and corresponding eigenvector, which will become PC1 of the given data set.

1.2. Problem Formulation and contribution

From generating plants to distribution system, industries or a place where electrical machines (motors, generators, Transformers) are the important parts of the production or generation, the efficiency and performance of the overall organization depends upon the reliability and performance of these equipment. Therefore condition monitoring of this equipment is necessary. There are various techniques for the condition monitoring of the equipment, one of them is the process in which the operator or the maintenance team do the maintenance of the equipment in a particular set time schedule of an interval (monthly/ six months/ yearly). Another technique that is generally used is to monitor the health condition of the equipment in real-time by fetching the sensors data of the equipment online, and the action for maintenance is taken when the sensor data of the equipment shows some abnormality means if the level of a particular sensor data crosses its threshold limit set by the operator or the maintenance team. These are not reliable as in the first technique the equipment may get broken down due to some internal faults that occur in between the scheduled maintenance time intervals and, in the second technique, one can know the fault condition of the equipment when the fault has started damaging the equipment and affecting the process, and also it will take some more time to diagnose the root cause of the fault. Therefore, there is need of preventive and predictive maintenance of these types of equipment which will be helpful in predicting the faults that would occur in future and also helpful in the diagnosis process to detect the root cause of the fault.

The techniques used for predictive maintenance usually include supervised machine learning techniques such as support vector machine (SVM) and artificial intelligence-based techniques such as a fuzzy and neural network. These techniques need proper learning from the historical data and a strong feature extraction technique. In many cases, the industries do not have historical data or fault data. In these cases, supervised machine learning techniques do not give accurate result and accuracy due to false learning. For such situations, the PCA-based unsupervised machine learning approach gives better performance for fault detection and localization using the raw data of the equipment to be monitored.

PCA technique has been used as a data reduction technique till now in most of the literature of fault diagnosis along with the other fault diagnosis techniques. This has not been used alone for the fault diagnosis of equipment.

In this paper, the unsupervised machine learning technique, PCA-based fault diagnosis algorithm is proposed. This proposed technique monitors three fan-motor systems alone and has the capability to monitor more equipment together. For industries having no historical data or faulty data can be benefited with this algorithm. The novelty of the work is focused on the uniqueness of the fault diagnosis system which is very simple, having less computation burden, having self-capability of extracting features as compared to the other techniques which demand feature extraction technique, data smoothing technique, and having computational complexity. The PCA technique is implemented alone for the fault diagnosis of industrial ID fans and motors using vibration signals which have not been implemented in the literature previously.

The main finding of the paper are as follows:

- The score vectors of PCA indicate the health condition of the equipment under consideration accurately.
- The movement of data in a particular direction indicates the degradation in health. Once maintenance is done, the initial value and final values collapse at the same place in the space of formed between the selected principal components indicating no failure in the equipment.
- The lifetime or time gain after maintenance can be evaluated using this predictive algorithm. This helps in achieving the safe operation of equipment and process.

1.3 Paper Organisation:

The paper is organised as follows: in Section II, the mathematical formulation of the PCA-based algorithm and the number of PCs are discussed; The computational procedure followed in this paper is explained in Section III, which includes data collection, pre-processing, selection of PCs, and data analysis; The results obtained from the PCA-based algorithm are discussed in Section IV; The work is done and results obtained concluded in Section V, and; The scopes for future works are highlighted in Section VI.

II. PCA Algorithm:

The authors in [23] have explained the PCA algorithm stepwise with the mathematical expression for finding the PCs. A dataset 'A' which is a non-square matrix is considered to find the PCs of matrix A. The required steps in finding PCs of any dataset A are as follows:

$$\text{A= dataset matrix=}[A]_{p\times q} \tag{1}$$

where, p and q are representing rows and columns of the matrix A, respectively. The first step is to subtract the mean from each dimension.

$$[A]_q - [\bar{A}]_q \tag{2}$$

Now, the covariance matrix is to be formed using the formula: $[C] = \sum_{i=1}^{p} \frac{(a_i-\bar{a})(a_i-\bar{a})^T}{p}$; where, $\bar{a} = mean\ value\ of\ a_i$.

The eigenvalues and eigenvectors are calculated, and eigenvectors are stored in a matrix [B], which contains the loading vectors, according to equation (4),

$$([C]_{q\times q} - I_{q\times q} \lambda)\{A\}_{q\times 1} = \{0\}; \tag{3}$$

where, I= identity matrix

$$[B]_{q\times q} = [\{A_1\}\{A_2\}\{A_3\} \dots \dots \dots \dots \dots \{A_q\}] \tag{4}$$

The eigenvalues are put in a diagonal matrix $[\text{eig}]_{q\times q}$, which contains the eigenvalues corresponding to the PCs. Then, eigenvalues are ranked in decreasing order, and top 'a' vectors are chosen $[\text{eig}]_{a\times a}$ to get the required PCs. The eigenvectors corresponding to the chosen eigenvalues are retained as given in equation (5).

$$[B]_{q\times a} = [\{A_1\}\{A_2\}\{A_3\} \dots \dots \dots \{A_a\}] \tag{5}$$

Finally, the PCs [P] are calculated, and then these are projected in the dataset matrix [A]:

$$[A]_{p\times q}[B]_{p\times a} = [P]_{q\times a} \tag{6}$$

For identification and classification of faults, the PCA method is a good fit that helps in processing the large datasets containing parameters of a large number that are distorted and highly correlated. The PCA technique is used to convert the high dimensional datasets into lower-dimensional, which is the space covered by the PCs [30-32]. In [30], Li et al. have explained the mathematical model of the PCA technique, which has been used in this paper.

For any dataset $A_{p\times q}$, which contains p observations of q variables (p>q), the PCA explains the variance of dataset A in terms of a new set of independent variables (PCs). When the PCA method is applied, the dataset A is written as the linear combination of orthogonal vectors, and these orthogonal vectors are in the direction of the PCs according to equation (7).

$$A = SB^T + R = s_1 b_1^T + s_2 b_2^T + \ldots\ldots + s_k b_k^T + R (7)$$

where, S=matrix formed by the scores of the principal components, B= matrix of loading for PCA, R= residual matrix(random error)

Generally, k should be much less than q so that there is no significant parameters information left in R. $|b_i| = 1$, as vectors of loading matrix are orthogonal to each other. i. e. $b_i^T b_j = 0\ (i \neq j)$ and, $b_i^T b_j = 1\ (i = j)$.

Each observation a_i is located on the principal component space by score vectors $s_i (i = 1,2, \ldots, k)$.

The distances of the origin of the space along with each PC from the elements of the score vector. The PC scores are calculated as: s_i =product of the vector of loading matrix b_i and the actual observation, according to equation (8)

$$\Rightarrow s_i = Ab_i\ (i = 1,2, \ldots. k) \qquad (8)$$

The maximum variance of the dataset is explained by the first PC, and it decreases for the second, third and so on. In the PCA method, it is assumed that a particular structure is formed with a large variance in the first few PCs space and the other remaining PCs cause a small loss of accuracy, which is negligible. It is required to cover the larger variation of the data (>85%) with the help of a specific number of PCs that have been chosen. There are various techniques in the literature which is used to decide the number of PCs to be selected. One most commonly used technique for this is the cross-validation technique [33]. It has been seen that first, second and third PCs are sufficient enough to cover the larger variation of the data. One more popular method for deciding the number of PCs is the singular value decomposition technique.

III. Computational Procedure:

The proposed technique and the process followed in this paper is illustrated in Figure1.

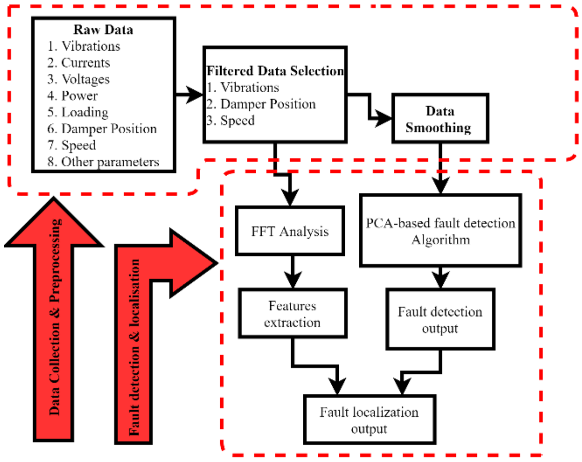

Fig. 1. Block diagram representation of the proposed system.

The procedure followed in PCA algorithm for fault detection is explained in this Section. The Fast Fourier Transform (FFT) based technique is used in parallel with the PCA technique with the raw data for FD if any fault is detected from the PCA algorithm. The peak value of vibration and corresponding frequency give the indication of type, and PCA shows the location of the fault. The features extracted from FFT include the root mean square (RMS) value of vibration and frequency of fundamental and multiple harmonics.

The block diagram representation of the PCA-based FD algorithm is shown in Figure 2. The vibration dataset matrix is used to form the loading matrix of the PCA technique. From the loading matrix formed, the PCs are evaluated. The number of PCs are equal to the number of attributes used in the vibration dataset. From PCs, effective PCs are selected. In this case, the first two PCs are found effective, which cover 99% of the data variance. The selected PCs and the dataset are analysed using the concepts of data analysis to find centroid, mean, standard deviation, mean path followed, first and last points, and range of scores of PCs which cover different range of dataset. This analysis are followed in the proposed algorithm to detect the fault in the fan-motor system.

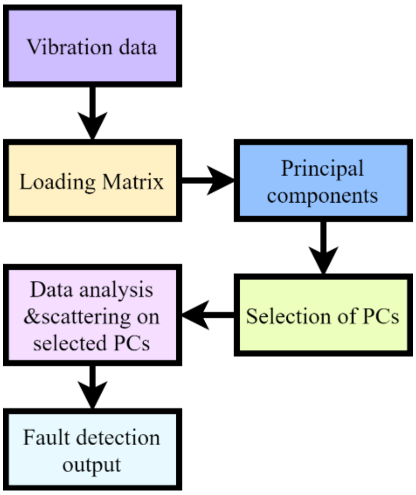

Fig. 2. PCA algorithm for fault detection

A. Data Collection and Preprocessing:

In this paper, the raw vibration data has been collected from three exhaust fans-motors system, and unsupervised learning algorithm using the PCA technique is fitted. The vibration signals to be monitored are smoothened first, therefore, the data analysis is done easily and smoothly. Figure3 shows the vibration of the motor and fan bearings of driving and non-driving ends and the smoothed signals are shown in Figure4.

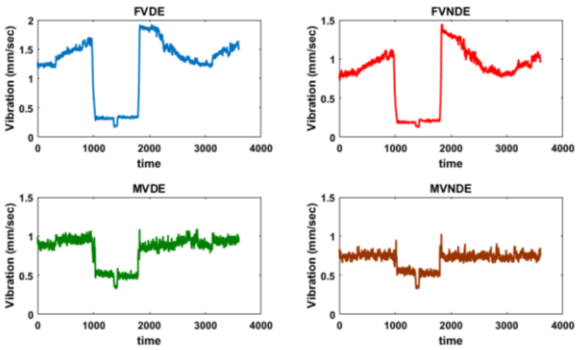

Fig.3. Vibration signals of fan and motor bearings of driving and non-driving; x-axis time (seconds); y-axis: vibrations(mm/sec)

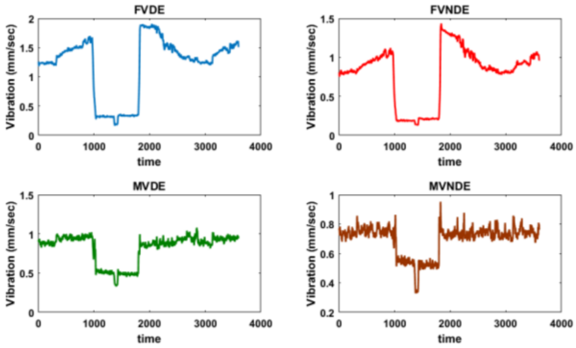

Fig. 4. Smoothed vibration data of fan and motor bearings of driving and non-driving ends; x-axis time (seconds), y-axis: vibrations(mm/sec)

In Figures 3 and 4, FVDE stands for fan vibration of driving end, FVNDE for fan vibration of the non-driving end, MVDE stands for motor vibration of driving end and MVNDE for motor vibration of the non-driving end.

B. *Selection of Principal Components:*

The data variance covered by each individual and cumulative PCs is analyzed. PC1 always covers the maximum data variance, and then it decreases for second, third principal components… After looking at the percentage of data covered by each component, one can select a few components, therefore, 95% of data or more can be covered. Here, the first two components are covering 99% dataset cumulatively. Therefore, only these two components are considered for further procedure. The percentage of data variance covered by individual PCs and cumulative components is shown in Figure5.

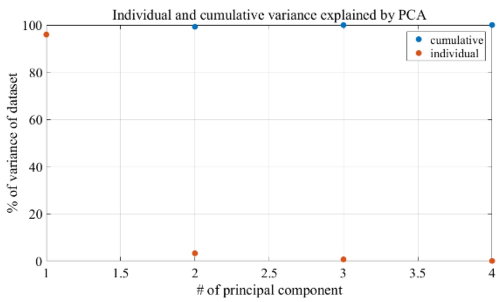

Fig. 5. Percentage of data explained by PC1 and PC2

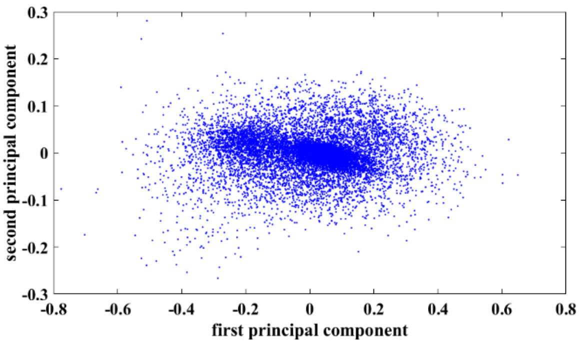

Fig. 6. Vibrations data projected into the PC1 and PC2

After selection of PCs, the dataset is scattered over the space formed by the selected PCs as shown in Figure 6. First and last data points give an indication of the direction of movement of data that is the direction of degradation. The first and last data point will collide if the maintenance has been done after failure or no failure has occurred, and the equipment is in normal condition. But if any degradation has occurred, then the last data point will move away from the first data point. Three identical fans data has been used for the analysis of the PCA technique to identify the faulty equipment. The data set of all three fans are scattered in PC space, as shown in Figure7. Their first and last data points are also highlighted to detect the direction of movement. The first and last data point of all three fans together is shown in Figure7(d).

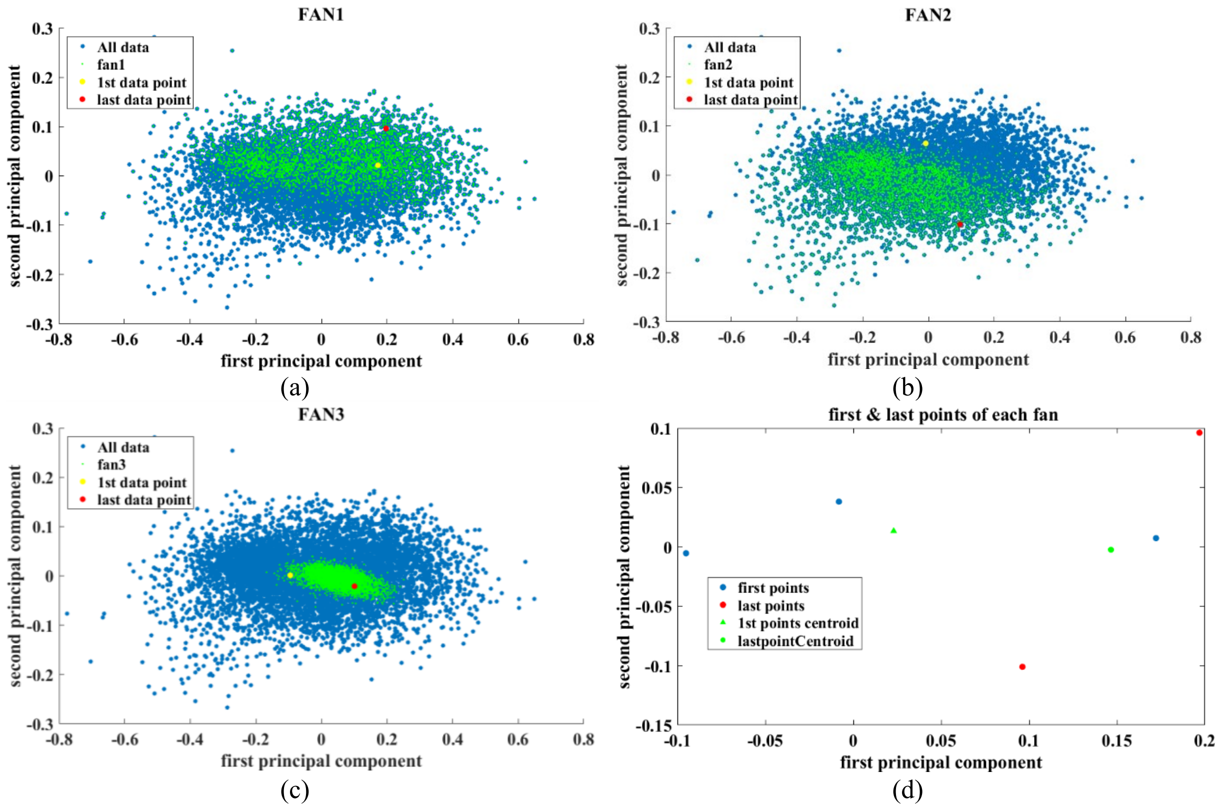

Fig. 7. First and last data points of fans: (a) fan1; (b) fan2; (c) fan3; (d) each fan

The range of vibration under normal operating condition is found to be in the range of 0.4 to 1.6 mm/sec, but if there is any problem with these fans, then the vibration will increase. Therefore, a threshold limit of 1.8mm/sec is set, and it is checked that how many data points are above this limit and how long the data is sustaining above this limit. If the data points are sustaining longer than five minutes, then the system needs to take action to prevent any failure and diagnose the problem’s root cause. Through programming, higher vibration points (>1.8) will be found out from the current data set, then it will be analyzed that which fan has these high vibration points. Therefore, it is necessary to find which fans are having these high vibration points so that prior maintenance can be given to that particular fan. This is analyzed through programming, and then after detecting the fan number, an automatic email is sent to the maintenance team to take action quickly. This system of maintenance is not reliable and, therefore, there is need of smart alarm system which can detect the fault in minimum possible time and can also be able to forecast the problem.

The FFT plot of the raw vibration data is shown in Figure 8. From FFT, the features are selected and used for the FD. These features include RMS value of vibration and multiple frequencies of the fundamental frequency at which peak values of vibration are detected. These help in detecting the type of fault. The location of the fault is detected from the proposed PCA algorithm. On combining the outputs of PCA and FFT analysis, the complete FD is performed.

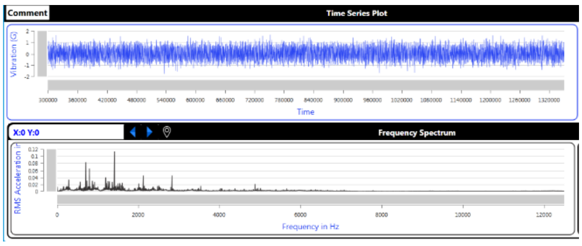

Fig. 8. FFT analysis of raw vibration data.

The FFT analysis is also performed using IBA Analyser software which can detect the fault condition of the equipment. The FFT of the fan vibration signal using IBA analyser is shown in Figure. 9.

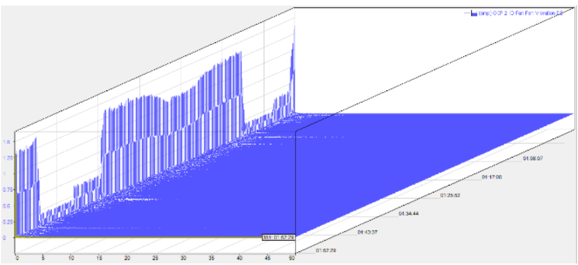

Fig. 9. FFT analysis using IBA Analyser software

IV. Results and Discussion:

In this paper, the PCA technique is used to make an algorithm for predictive maintenance of industrial equipment. For a case study, the fan-motor system is considered to use its historical data in the algorithm. This technique is applicable to other equipment of the plant whose failure data is not available. The alarm system generated using PCA technique is shown in Figure10: if the data points are lying in the green region, then the equipment is working in the normal range; If the data points are lying in the orange region, then there is some problem with the equipment and the system automatically give the alarm; If the data points are lying in the red region, then the system gives warning that the equipment is not in normal condition and some problem has occurred in the equipment. The coverage and boundary of different operation condition is not taken randomly, but it is calculated from the PCA algorithm using the score coefficients of the PCs. The score coefficients of PCs are important parameters for health monitoring of equipment [34].

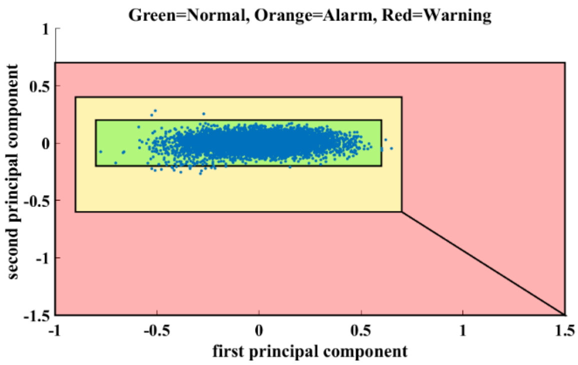

Fig. 10. Alarm system for Condition monitoring of ID fans

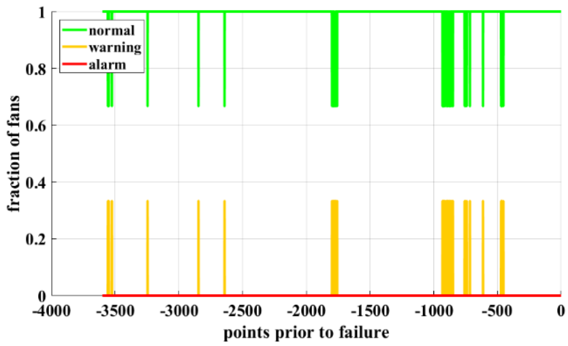

Fig. 11. Condition monitoring of ID Fans using PCA technique

Using this technique, it is also possible to see how many fractions of equipment are working under normal condition, warning condition and in alarm condition as shown in Figure11. It is also able to find how much it takes for the vibration of the equipment (fan/motor) to reach to the warning region from the alarm region assuming warning region to be the breakdown of the equipment. Therefore, it is simply the time to failure of the equipment or the time left for the equipment for maintenance before failure.

The condition monitoring system will be reliable when the maintenance team and the operator are given a health status update automatically, which is done using an alarm system along with email and text message. Whenever there is any fault condition is detected, the system automatically sends an email to the maintenance team members regarding the same, and also gives an appropriate alarm. In this paper, the additional uptime gained by doing maintenance upon first alarm and first warning has been calculated for the forecasting and scheduling the predictive and preventive maintenance of the equipment as shown in Table1. The comparison of the proposed work with the existing conventional methods are provided in Table 2.

Table1. Output of the PCA algorithm regarding the health status of the equipment

Fans having vibration >1.8	Fan1, Fan2
Maximum Vibration/ Fan	1.9274/Fan1
Cycles gained by doing maintenance upon first alarm signal(in %)	10800.0
Cycles gained by doing maintenance upon first warning signal(in %)	9797.0
Alarm message to maintenance team	Fan1 and Fan2 vibrations are >1.8, Fan1 vibration is higher, so it needs prior maintenance.

Table . 2. Comparitive analysis of proposed method with the existing methods

Existing methods for FD			PCA-based FD technique
Conventional method	IBA Analyser	FFT [4]	
Online signal trend is analysed. If the peak value of vibration crosses a threshold value, alarm is generated.	Signals are analysed online along with the FFT analysis available in IBA Analyser.	Fault is detected when it occurs and indication is observed from the FFT of the raw vibration signals.	Fault is detected before occurring as the degradation is indicated using the porposed technique.
No option to find the time to failure.	No option to find the time to failure.	No option to find the time to failure.	Time to failure is possible to find and also the life cycle gained after maintenance is possible to find.
Fault localization is not possible.	Fault localization is possible using FFT.	Fault localization is possible.	Fault localization is possible.

V. Conclusion

The proposed PCA-based algorithm for condition monitoring of industrial equipment is reliable in both the cases of unavailability and availability of fault data, as it can indicate the future occurring faults and can avoid unexpected breakdown of equipment. It is also helpful in forecasting the schedule of maintenance long before the fault occurs. Therefore, it is better than the existing methods including FFT, trend and wavelet analysis which give indication of fault only after occurance of fault. The existing methods mentioned are also not capable of evaluating the time to failure. The proposed technique is able to evaluate the time to failure for the equipment and also the life cycles gained after doing maintenance in the first alarm or warning of the fault. Thus, the plant efficiency and performance are improved using the proposed technique. This technique is also applicable for the condition monitoring and predictive maintenance of other equipment. An industry with larger number of identical equipment can use this technique to monitor all together using a single PCA-based monitoing technique.

VI. Future Scope of Work

The fault detection is the first step in the predictive maintenance of equipment. The proposed work has been carried out using vibration signals of the equipment. There are also other signals, e.g. temperature, pressure, lubrication, etc. which could be good predictors for predicting and monitoring the health status of industrial equipment. The accuracy of the proposed algorithm can be validated and compared with other existing unsupervised techniques including clustering. The fault condition and the corresponding faulty data can be collected using the proposed algorithm. It can also be used further in predictive maintenance using supervised learning-based algorithm which uses both faulty and normal operation data. This will be a step towards the digital transformation of the industry and help it to move towards the revolution of industry 4.0.

Nomenclature

PCA	Principal Component Analysis
ISO	International Standards Organization
RCM	Reliability-Centred Maintenance
SVM	Support Vector Machine
IRT	Infrared Thermography
IOTs	Internet of Things
PSO	Particle Swarm Optimization
PCs	Principal Components
MNNMDPCA	Multiscale Neighborhood Normalization-Based Multiple Dynamic Principal Component Analysis
MROC	Modified Rank Order Clustering
MVNDE	Motor Vibration of Non-Driving End
MVDE	Motor Vibration of Driving End
FVDE	Fan Vibration of Driving End
FVNDE	Fan Vibration of Non-Driving End
ID	Induced Draft
FD	Fault Diagnosis
FFT	Fast Fourier Transform
WT	Wavelet Transform
WGP	Wrapper Genetic Programming
IIOT	Industrial Internet of Things
RMS	Root Mean Square

Acknowledgement

The work reported herewith has been financially by the Dirección General de Universidades, Investigación e Innovación of Castilla-La Mancha, under Research Grant ProSeaWind project (Ref.: SBPLY/19/180501/000102) and the Spanish Ministerio de Economía y Competitividad, under Research Grant (DPI2015-67264-P).

The authors would like to express gratitude to Tata Steel, Jamshedpur and BIT Sindri, Jharkhand, India for providing research facilities to carry out this work, and thankful for the assistance received from the TEQIP CRS project ID 1-5766329561 program for the research work.

References

[1] Kaparthi, Shashidhar, and Daniel Bumblauskas. "Designing predictive maintenance systems using decision tree-based machine learning techniques." *International Journal of Quality & Reliability Management* (2020). https://doi.org/10.1108/IJQRM-04-2019-0131

[2] Saidy, Clint, Kaishu Xia, Anil Kircaliali, Ramy Harik, and Abdel Bayoumi. "The Application of Statistical Quality Control Methods in Predictive Maintenance 4.0: An Unconventional Use of Statistical Process Control (SPC) Charts in Health Monitoring and Predictive Analytics." In *Advances in Asset Management and Condition Monitoring*, pp. 1051-1061. Springer, Cham, 2020. https://doi.org/10.1007/978-3-030-57745-2_87

[3] Horrell, Michael, Larry Reynolds, and Adam McElhinney. "Data Science in Heavy Industry and the Internet of Things." *Harvard Data Science Review* (2020). https://doi.org/10.1162/99608f92.834c6595

[4] Gholap, A. B., and M. D. Jaybhaye. "Condition-Based Maintenance Modeling Using Vibration Signature Analysis." In *Reliability and Risk Assessment in Engineering*, pp. 111-122. Springer, Singapore, 2020. https://doi.org/10.1007/978-981-15-3746-2_10

[5] Wang, Tianyang, Qinkai Han, Fulei Chu, and Zhipeng Feng. "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review." *Mechanical Systems and Signal Processing* 126 (2019): 662-685. https://doi.org/10.1016/j.ymssp.2019.02.051

[6] Malla, Chandrabhanu, and Isham Panigrahi. "Review of condition monitoring of rolling element bearing using vibration analysis and other techniques." *Journal of Vibration Engineering & Technologies* 7, no. 4 (2019): 407-414. https://doi.org/10.1007/s42417-019-00119-y

[7] Li, Yang, Qirong Tang, Qing Chang, and Michael P. Brundage. "An event-based analysis of condition-based maintenance decision-making in multistage production systems." *International Journal of Production Research* 55, no. 16 (2017): 4753-4764. https://doi.org/10.1080/00207543.2017.1292063

[8] Quatrini, Elena, Francesco Costantino, Giulio Di Gravio, and Riccardo Patriarca. "Condition-Based Maintenance—An Extensive Literature Review." *Machines* 8, no. 2 (2020): 31. https://doi.org/10.3390/machines8020031

[9] Fahmy, Muhamad Noval. "Implementation of Maintenance Method on Steam Turbine Using Reliability-Centred Maintenance (RCM)." PhD diss., Institut Teknologi Sepuluh Nopember, 2020. http://repository.its.ac.id/id/eprint/80589

[10] Toh, Gyungmin, and Junhong Park. "Review of vibration-based structural health monitoring using deep learning." *Applied Sciences* 10, no. 5 (2020): 1680. https://doi.org/10.3390/app10051680

[11] A. Bousdekis and G. Mentzas, "A Proactive Model for Joint Maintenance and Logistics Optimization in the Frame of Industrial Internet of Things," in *Operational Research in the Digital Era–ICT Challenges*, Springer, 2019, pp. 23–45. https://doi.org/10.1007/978-3-319-95666-4_3

[12] B. O. Gombé *et al.*, "A SAW wireless sensor network platform for industrial predictive maintenance," *Journal of Intelligent Manufacturing*, vol. 30, no. 4, pp. 1617–1628, 2019. https://doi.org/10.1007/s10845-017-1344-0

[13] J. Wan *et al.*, "A manufacturing big data solution for active preventive maintenance," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2039–2047, 2017. doi: 10.1109/TII.2017.2670505

[14] W. Yu, T. S. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, "A Global Manufacturing Big Data Ecosystem for Fault Detection in Predictive Maintenance," IEEE Transactions on Industrial Informatics, 2019. doi: 10.1109/TII.2019.2915846

[15] Joyce, Jacob J., and Meshach W. Thamba. "Industrial internet of things (iiot)-an iot integrated services for industry 4.0: A review." *International Journal of Applied Science and Engineering* 8, no. 1 (2020): 37-42. http://dx.doi.org/10.30954/2322-0465.1.2020.5

[16] Munirathinam, Sathyan. "Industry 4.0: Industrial Internet of Things (IIOT)." In *Advances in Computers*, vol. 117, no. 1, pp. 129-164. Elsevier, 2020. https://doi.org/10.1016/bs.adcom.2019.10.010

[17] Li, Li, Yong Wang, and Kuo-Yi Lin. "Preventive maintenance scheduling optimization based on opportunistic production-maintenance synchronization. " *Journal of Intelligent Manufacturing* (2020): 1-14. https://doi.org/10.1007/s10845-020-01588-9

[18] M. Alweshah, O. A. Alzubi, J. A. Alzubi, and S. Alaqeel, " Solving Attribute Reduction Problem using Wrapper Genetic Programming". International Journal of Computer Science and Network Security, Vol. 16, No. 5, 2016. http://paper.ijcsns.org/07_book/201605/20160512.pdf

[19] Jafar A. Alzubi, "Optimal Classifier Ensemble Design Based on Cooperative Game Theory". Research Journal of Applied Sciences, Engineering and Technology, Vol. 11, No. 12, pp. 1336 -1343, 2015. DOI:10.19026/rjaset.11.2241

[20] Jafar A. Alzubi, "Diversity-Based Boosting Algorithm". International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016. https://www.academia.edu/download/48502471/Paper_70-Diversity_Based_Boosting_Algorithm.pdf

[21] O. A. Alzubi, J. A. Alzubi, S.Tedmori, H. Rashaideh, and O. Almomani "ConsensusBased Combining Method for Classifier Ensembles "The International Arab Journal of Information Technology, Vol. 15, No. 1, 2018. https://iajit.org/PDF/January%202018,%20No.%201/10845.pdf

[22] Jafar A. Alzubi "Diversity Based Improved Bagging Algorithm" In Proceedings of the International Conference on Engineering & MIS 2015 (ICEMIS '15). Istanbul - Turkey. https://doi.org/10.1145/2832987.2833043

[23] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance," in *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, 2018, pp. 355–361. doi: 10.1109/IEA.2018.8387124.

[24] A. Bakdi, A. Kouadri, and A. Bensmail, "Fault detection and diagnosis in a cement rotary kiln using PCA with EWMA-based adaptive threshold monitoring scheme," *Control Engineering Practice*, vol. 66, pp. 64–75, Sep. 2017. https://doi.org/10.1016/j.conengprac.2017.06.003

[25] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a K-means clustering approach," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2888–2911, Mar. 2011. https://doi.org/10.1016/j.eswa.2010.08.083

[26] J. Goldberger and S. T. Roweis, "Hierarchical Clustering of a Mixture Model," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. MIT Press, 2005, pp. 505–512. http://papers.nips.cc/paper/2585-hierarchical-clustering-of-a-mixture-model.pdf

[27] J. Yang, Y. Chen, and Z. Sun, "A real-time fault detection and isolation strategy for gas sensor arrays," in *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2017, pp. 1–6. doi: 10.1109/I2MTC.2017.7969906

[28] Y. Wang, F. Sun, and B. Li, "Multiscale Neighborhood Normalization-Based Multiple Dynamic PCA Monitoring Method for Batch Processes With Frequent Operations," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 3, pp. 1053–1064, Jul. 2018. doi: 10.1109/TASE.2017.2713800

[29] Z. Du, B. Fan, X. Jin, and J. Chi, "Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis," *Building and Environment*, vol. 73, pp. 1–11, Mar. 2014. https://doi.org/10.1016/j.buildenv.2013.11.021

[30] W. Li, T. Shi, G. Liao, and S. Yang, "Feature extraction and classification of gear faults using principal component analysis," *Journal of Quality in Maintenance Engineering*, Jun. 2003. https://doi.org/10.1108/13552510310482389

[31] Taguchi, Yh. "PCA Based Unsupervised FE." In *Unsupervised Feature Extraction Applied to Bioinformatics*, pp. 81-102. Springer, Cham, 2020. **DOI** https://doi.org/10.1007/978-3-030-22456-1_4

[32] Hemavathi, D., and H. Srimathi. "Effective feature selection technique in an integrated environment using enhanced principal component analysis." *Journal of Ambient Intelligence and Humanized Computing* (2020): 1-10. https://doi.org/10.1007/s12652-019-01647-x

[33] S. Wold, "Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models," *Technometrics*, Apr. 2012. https://doi.org/10.1080/00401706.1978.10489693

[34] Kellogg, Joshua J., Olav M. Kvalheim, and Nadja B. Cech. "Composite score analysis for unsupervised comparison and network visualization of metabolomics data." *Analytica Chimica Acta* 1095 (2020): 38-47. https://doi.org/10.1016/j.aca.2019.10.029