A Naive Bayes Classifier for Fault Detection and Classification Using Dimension Reduction Technique

Nasar Aldian Ambark Shashoa
Department of Electrical and Computer Engineering
Libyan Academy for Postgraduate Studies
Tripoli, Libya
naser.shashoa@academy.edu.ly

Omer S.M. Jomah
Department of Electrical and Computer Engineering
Libyan Academy for Postgraduate Studies
Tripoli, Libya
omer.jomah@academy.edu.ly

Omar Abusaeeda
Department of Computer Networks
University of Tripoli
Tripoli, Libya
omar.abusaeeda@uot.edu.ly

Musa Faneer
Department of Electrical and Computer Engineering
Libyan Academy for Postgraduate Studies
Tripoli, Libya
musa.faneer@gmail.com

***Abstract*—Fault detection and classification is critical to the reliability of modern control systems in different industries, where detecting and classifying faults in operational processes are very important things while failure to detect and classify them, may cause irreparable damage. In this paper, fault detection and classification approach is presented. The first step, multi stage recursive least squares parameter estimation approach for controlled autoregressive autoregressive moving average systems (CARARMA) is developed with a view to estimate the parameters of the system, additionally, improve the effectiveness of the computation. By means of multi stage approach, the (CARARMA) system is decomposed into three simple identification models, and the parameters of each simple model is identified one by one. These parameters estimated by this approach are referred to as features, and not all of them have the same useful data about the system. In order to select the valuable features and improve a classification accuracy, the Linear Discriminant Analysis (LDA) approach based on scattering matrices is applied for dimension reduction. The classification between these reduced classes is done based on the Naive Bayes classifier. Finally, the obtained results explain the performance of this proposed approach.**

Keywords— fault detection and classification, Parameter Estimation, Multi-Stage, Linear Discriminant Analysis, Naive Bayes classifier.

I. INTRODUCTION

For industrial operations to promote system reliability and safety, fault detection techniques are increasingly necessary. Therefore, the importance of fault detection and isolation has increased along with the requirement for system reliability and safety, which is being pushed by both economic incentives and environmental factors. For the systems to identify and isolate problematic components, some form of redundancy is required. Hardware redundancy can be used to provide redundancy in situations with very high security requirements. Hardware redundancy offers excellent performance and reliability, however there are disadvantages to the strategy, such as the price of additional hardware, the weight and size of the device, and the fact that some components cannot be replicated. Analytical redundancy can take the place of hardware redundancy. In this technique, a mathematical model's redundancy can be used in place of additional hardware. Model-based technique is the term used to describe the analytical redundancy. An important method of model-based approach is parameter estimation, which involves estimating the parameters using the real data [1]. For many diverse systems, LS and RLS algorithms algorithm are a well-known and significant model-based techniques [2]. Although the LS and RLS algorithms are capable of estimating system parameters, they are constrained by their inability to estimate the parameter of the noise model and their poor computing efficiency [3]. There are many modern method to overcome these problems, and in this paper, multi stage RLS identification algorithm for (CARARMA) systems is chosen. By using this method, the noise model parameters can be estimated and computational load is reduced. The estimated parameters produced by this methodology are called as features and not all of these features have the same level of informational usefulness. Therefore, a few useful parameters can accurately describe the process' behavior. It is important to choose the informative parameters for the classification mission, to improve classification accuracy. A dimension reduction approach is utilized to solve this case. In this work, it is decided to use (LDA) approach based on scattering matrices. This technique is effective for reducing dimensions [4]. Finally, the classification is done using the Naive Bayes classifier. It is depends on Bayesian probability model. If there is a class offered, this classifier is based on the idea that the value of each feature is independent of each other. It is depends on the mathematical principle of conditional probability. The remainder of this article is structured as: in section 2 the system description and identification model is introduecd. In section 3 multi stage recursive least square identification algorithm is derived. Section 4 presents dimension reduction techniques. Section 5 naive bayes classifier is introduced. Section 6 a simulation

results are given to illustrate the capability of the proposed approach. The conclusions are displayed in section 7.

II. System Description and Identification Model

In this study, the discrete-time, linear, time-invariant system is described by (CARARMA) model and a drawing of it is shown in fig. 1

$$A(z)y(n) = B(z)u(n) + \frac{D(z)}{C(z)}\xi(n) \quad (1)$$

Where

$$\mathrm{A(z)} = A(z) = 1 + \sum_{i=1}^{n_a} a_i z^{-i}$$

$$B(z) = \sum_{i=1}^{n_b} b_i z^{-i}$$

$$C(z) = 1 + \sum_{i=1}^{n_c} c_i z^{-i}$$

$$D(z) = 1 + \sum_{i=1}^{n_c} d_i z^{-i}$$

$u(\mathrm{n}), \mathrm{y(n)}$ are the system input and output, $\xi(\mathrm{n})$ is the noise, and a_i , b_j , c_i and d_i are the constants.

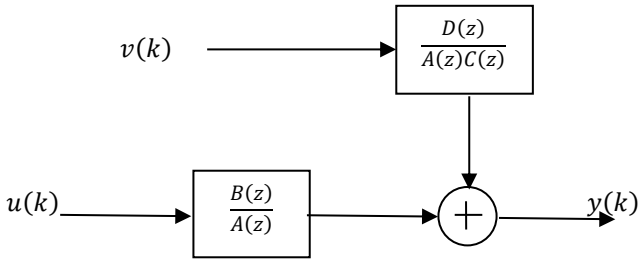

Fig. 1. The CARARMA system

The inner variable is initially defined as

$$v(n) = \frac{D(z)}{C(z)}\xi(n), \quad (2)$$

As a result, It is possible to rewrite equation (1) as

$$\mathrm{A(z)y(n)} = B(z)u(\mathrm{n}) + \mathrm{v(n)} \quad (3)$$

Using a linear regression form, the following explanation can be made:

$$y(\mathrm{n}) = {Z_1}^T(n)\theta_1 + {Z_2}^T(n)\theta_2 + v(n) \quad (4)$$

Where,

$$\mathrm{Z_1(n)} = [-\mathrm{y(n-1)}, \dots, -\mathrm{y(n-n_a)}]^{\mathrm{T}} \in \mathbb{R}^{\mathrm{n_a}},$$

$$\mathrm{Z_2(k)} = [\mathrm{u(n-1)}, \dots, \mathrm{u(n-n_b)}]^{\mathrm{T}} \in \mathbb{R}^{\mathrm{n_b}},$$

The information vectors.

$$\theta_1 := [\mathrm{a_1}, \dots, \mathrm{a_{n_a}}]^{\mathrm{T}} \in \mathbb{R}^{\mathrm{n_a}},$$

$$\theta_2 := [\mathrm{b_1}, \dots, \mathrm{b_{n_b}}]^{\mathrm{T}} \in \mathbb{R}^{\mathrm{n_b}},$$

The parameter vectors. Equation (2) can be changed as

$$v(k) = [1 - C(z)]v(k) + D(z)\xi(k)$$
$$= -c_1 v(n-1) - \dots - c_{n_c} v(n-n_c) +$$
$$d_1\xi(n-1) + \cdots + d_{n_d}\xi(n-n_d) + \xi(n) \quad (5)$$

It can alternatively be stated as a linear regression with the following formula:

$$v(k) = {Z_3}^T(n)\theta_3 + \xi(n) \quad (6)$$

Where

$$\mathrm{Z_3(n)} = [-\mathrm{v(n-1)}, \dots, -\mathrm{v(n-n_c)}, \xi(\mathrm{n-1}), \dots, \xi(\mathrm{n-n_d})]^{\mathrm{T}}$$

$$\theta_3 := [\mathrm{c_1}, \dots, \mathrm{c_{n_c}}, \mathrm{d_1}, \dots, \mathrm{d_{n_d}}]^{\mathrm{T}}$$

$$\theta = [\theta_1\ \theta_2\ \theta_3]^{\mathrm{T}}$$

Substitute equation (6) into equation (4), the result is given by the equation below [5],

$$y(n) = {Z_1}^T(n)\theta_1 + {Z_2}^T(n)\theta_2 + {Z_3}^T(n)\theta_3 + \xi(n)$$
$$= \mathrm{Z}^T(n)\theta + \xi(n) \quad (7)$$

$$\mathrm{Z}^T(n) = \left[{Z_1}^T(n)\ {Z_2}^T(n)\ {Z_3}^T(n)\right]$$

$$\theta = \begin{bmatrix}\theta_1\\ \theta_2\\ \theta_3\end{bmatrix}$$

The system's (CARARMA) identification model is represented by this outcome.

III. Multi Stage Recursive Least Square Identification Algorithm

The decomposition technique is the fundamental concept for the multi-stage RLS algorithm. Hence, the (CARARMA) System is divided into three subsystems as a result. First, three intermediate variables should be defined.

$$y_1(n) = y(n) - {Z_2}^T(n)\theta_2 - {Z_3}^T(n)\theta_3, \quad (8)$$

$$y_2(n) = y(n) - {Z_1}^T(n)\theta_1 - {Z_3}^T(n)\theta_3, \quad (9)$$

$$y_3(n) = y(n) - {Z_1}^T(n)\theta_1 - {Z_2}^T(n)\theta_2, \quad (10)$$

It is possible to separate the system in equation (5) into the multiple fictitious subsystems indicated below [6].

$$y_i(n) = Z_i^T(n)\theta_i + \xi(n), \quad i = 1,2,3. \quad (11)$$

Three criterion functions should be defined,

$$J_i(\theta_i) := \sum_{j=1}^{t}\left[y_i(j) - {Z_i}^T(j)\theta_i\right]^2, \quad i = 1,2,3.$$

Using the partial derivatives of the solutions to the multi-optimization problems is

$J_i(\theta_i), i = 1,2,3.$ with regards to θ_i be zero gives

$$\left.\frac{\partial J_i(\theta_i)}{\partial\theta_i}\right|_{\theta_i=\hat{\theta}_i(t)} = -2Z_i(j)\sum_{j=1}^{t}[y_i(j) - {Z_i}^T(j)\hat{\theta}_i] = 0, \quad i = 1,2,3. \quad (12)$$

Considering the estimated parameters

$$\hat{\theta}(k) := \theta = \left[\hat{\theta}_1(n)\ \hat{\theta}_2(n)\ \hat{\theta}_3(n)\right]^{\mathrm{T}}$$

From (12), we may obtain the RLS algorithm for computing at time n as:

$$\hat{\theta}_i(n) = \hat{\theta}_i(n-1) + L_i(n)\left[y_i(n) - Z_i{}^T(k)\hat{\theta}_i(n-1)\right] \quad (13)$$

$$L_i(n) = P_i(n-1)Z_i(n)\left[+Z_i{}^T(n)P_i(n-1)Z_i(n)\right]^{-1},$$

$$P_i(n) = \left[I - L_i(n)Z_i{}^T(n)\right]P_i(n-1),$$

$$P_i(0) = p_0 I, \qquad i = 1,2,3$$

With i=1,2,3, and substitute equations (8)-(10) into equation (13) results in:

$$\hat{\theta}_1(n) = \hat{\theta}_1(n-1) + L_1(n)\left[y(n) - Z_2{}^T(n)\theta_2 - Z_3{}^T(n)\theta_3 - Z_1{}^T(n)\hat{\theta}_1(n-1)\right] \quad (14)$$

$$\hat{\theta}_2(n) = \hat{\theta}_2(n-1) + L_2(n)\left[y(n) - Z_1{}^T(n)\theta_1 - Z_3{}^T(n)\theta_3 - Z_2{}^T(n)\hat{\theta}_2(n-1)\right] \quad (15)$$

$$\hat{\theta}_3(n) = \hat{\theta}_3(n-1) + L_1(n)\left[y(n) - Z_1{}^T(n)\theta_1 - Z_2{}^T(n)\theta_2 - Z_3{}^T(n)\hat{\theta}_3(n-1)\right] \quad (16)$$

The issue is that the right-hand sides of the equations from (14) – (16), include the vectors of unknown parameters θ_1, θ_2 and θ_3. In order to solve this issue, the replacement of these unknown parameters with their previous estimation is done $\hat{\theta}_i(n-1)$, , where, $i = 1,2,3$.

$$\hat{\theta}_1(n) = \hat{\theta}_1(n-1) + L_1(n)\left[y(n) - Z_2{}^T(n)\hat{\theta}_2(n-1) - Z_3{}^T(n)\hat{\theta}_3(n-1) - Z_2{}^T(n)\hat{\theta}_1(n-1)\right] = \hat{\theta}_1(n-1) + L_1(n)\left[y(n) - Z^T(n)\hat{\theta}(n-1)\right],$$

$$\hat{\theta}_2(n) = \hat{\theta}_2(n-1) + L_2(n)\left[y(n) - Z_1{}^T(n)\hat{\theta}_1(n-1) - Z_3{}^T(n)\hat{\theta}_3(n-1) - Z_2{}^T(n)\hat{\theta}_2(n-1)\right] = \hat{\theta}_2(n-1) + L_2(n)\left[y(n) - Z^T(n)\hat{\theta}(n-1)\right],$$

$$\hat{\theta}_3(n) = \hat{\theta}_3(n-1) + L_3(n)\left[y(n) - Z_1{}^T(n)\hat{\theta}_1(n-1) - Z_2{}^T(n)\hat{\theta}_2(n-1) - Z_3{}^T(n)\hat{\theta}_3(n-1)\right] = \hat{\theta}_3(n-1) + L_3(n)\left[y(n) - Z^T(n)\hat{\theta}(n-1)\right],$$

The inner variables $w(n-i)$ of the vector $\varphi_3(n)$ are unknown, and their estimates can be utilized in place of $w(n-i)$as:

$$\hat{Z}_3(n) := \left[-\hat{v}(n-1), \dots, -\hat{v}(n-n_c), \hat{\xi}(n-1), \dots, \hat{\xi}(n - \mathrm{n_d})\right]^T$$

$$\hat{Z}(n) = \left[Z_1 \;\; Z_2 \;\; \hat{Z}_3\right]^{\mathrm{T}}$$

Equation (5) provides the following:

$$v(n) = y(n) - Z_1{}^T(n)\theta_1 - Z_2{}^T(n)\theta_2,$$

Afterwards, $\hat{v}(n)$ can be determined by using:

$$\hat{v}(n) = y(n) - Z_1{}^T(n)\hat{\theta}_1 - Z_2{}^T(n)\hat{\theta}_2.$$

Hence, the multi-stage RLS identification algorithm is described as [6]:

$$\hat{\theta}_1(n) = \hat{\theta}_1(n-1) + L_1(n)\left[y(n) - \hat{Z}^T(n)\hat{\theta}(n-1)\right] \quad (17)$$

$$L_1(n) = P_1(n-1)Z_1(n)\left[1 + Z_1{}^T(n)P_1(n-1)Z_1(n)\right]^{-1} \quad (18)$$

$$P_1(n) = \left[I - L_1(n)Z_1{}^T(n)\right]P_1(n-1),$$

$$P_1(0) = p_0 I_{n_a} \quad (19)$$

$$Z_1(n) = [-y(n-1), -y(n-2), \dots, -y(n-n_a)]^T \quad (20)$$

$$\hat{\theta}_2(n) = \hat{\theta}_2(n-1) + L_2(n)\left[y(n) - \hat{Z}^T(n)\hat{\theta}(n-1)\right] \quad (21)$$

$$L_2(n) = P_2(n-1)Z_2(n)\left[1 + Z_2{}^T(n)P_2(n-1)Z_2(n)\right]^{-1} \quad (22)$$

$$P_2(n) = \left[I - L_2(n)Z_2{}^T(n)\right]P_2(n-1),$$

$$P_2(0) = p_0 I_{n_\mathrm{b}}, \quad (23)$$

$$Z_2(n) = [\mathrm{u}(n-1), \mathrm{u}(n-2), \dots, \mathrm{u}(n-n_\mathrm{b})]^T \quad (24)$$

$$\hat{\theta}_3(n) = \hat{\theta}_3(n-1) + L_3(n)\left[y(n) - \hat{Z}^{\mathrm{T}}(n)\hat{\theta}(n-1)\right] \quad (25)$$

$$L_3(n) = P_3(n-1)\hat{Z}_3(n)\left[1 + \hat{Z}_3{}^T(n)P_3(n-1)\hat{Z}_3(n)\right]^{-1} \quad (26)$$

$$P_3(n) = \left[I - L_3(n)\hat{Z}_3{}^T(n)\right]P_3(n-1),$$

$$P_3(0) = p_0 I_{n_c}, \quad (27)$$

$$\hat{v}(n) = y(n) - Z_1{}^T(n)\hat{\theta}_1 - Z_2{}^T(n)\hat{\theta}_2 \quad (28)$$

$$\hat{Z}_4(n) := \left[-\hat{v}(n-1), \dots, -\hat{v}(n-n_c), \hat{\xi}(n-1), , \dots, \hat{\xi}(\mathrm{n} - \mathrm{n_d})\right]^T \quad (29)$$

$$\hat{\mathrm{Z}}(\mathrm{n}) = \left[\mathrm{Z}_1^{\mathrm{T}}(n), \mathrm{Z}_2^{\mathrm{T}}(n), \hat{\mathrm{Z}}_3^{\mathrm{T}}(n)\right] \in \mathbb{R}^{\mathrm{n}}, \quad (30)$$

$$\hat{\theta}(n) = \left[\hat{\theta}_1^{\mathrm{T}}(n), \hat{\theta}_2^{\mathrm{T}}(n), \hat{\theta}_3^{\mathrm{T}}(n)\right] \in \mathbb{R}^n \quad (31)$$

IV. DIMENSION REDUCTION TECHNIQUES

The term "dimension" describes a measurement of an object's specific aspect. The study of techniques to lower the number of dimensions describing an object is known as dimension reduction. Its main goals are to eliminate duplicate and unnecessary data In order to lower computational loads and prevent data over-fitting as well as to enhance data quality for effective data-intensive processing jobs [7]. For this study, (LDA) method depend on scattering matrices has been chosen from a variety of available dimension reduction techniques. In order to do this, a within-class matrix, between-class matrix, , and a mixed scattering matrix, , are defined:

$$S_W = \sum_{i=1}^{L} P_i E\{(X - M_i)(X - M_i)^T \,|\omega_i\} = \sum_{i=1}^{L} P_i \Sigma_i \quad (32)$$

$$S_B = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T \quad (33)$$

$$S_M = E\{(X - M_0)(X - M_0)^T\} = S_W + S_B \quad (34)$$

$$M_0 = E\{X\} = \sum_{i=1}^{L} P_i M_i \quad (35)$$

Where P_i class ω_i a priori probability, that is $P_i \cong {}^{n_i}\!/_{N}$, where n_i is the number of samples in Class ω_i, out of a total of N samples, M_i is the vector of mathematical expectation and Σ_i is the covariance matrix of *i*-th class. It is possible to

minimize the criterion in order to solve the challenge of finding the transformation matrix A.

$$J = tr(S_M^{-1} S_B) \tag{36}$$

In order to find the answer, the matrix's eigenvalues are searched for

$$(S_M^{-1} S_B)\varphi_i = \lambda_i \varphi_i; \quad i = 1,2,...,n; \quad \lambda_1 \geq ... \geq \lambda_n \tag{37}$$

Then, the matrix *A* (transformation matrix) is

$$\mathrm{A} = [\phi_1 \ . \ . \ . \ \phi_m]_{n*m} \tag{38}$$

Where φ_i, $i = 1,2,...,m$ indicate eigenvectors coinciding to the largest eigenvalues $\lambda_1,...,\lambda_m$

Then, reduced-dimension vectors are [7]:

$$X_{m*1} = \mathrm{A}_{n*m}^T X_{n*1} \tag{39}$$

V. NAIVE BAYES CLASSIFIER

The Naive Bayes classifier is depends on Bayesian probability model. If there is a class offered, The Naive Bayes classifier is based on the idea that the value of each feature is independent of each other. Its foundation is conditional probability, a mathematical concept. The Naive Bayes classifier makes 2n! Independent assumptions if n attributes are provided. A conditional probability model is provided as follows for the classifier above [9]:

$$P(\omega_i \,|x) \tag{40}$$

In this case, class variable ω is conditional on several features

Variable $x = x_1,...,x_n$

Equation (40) can be represented by applying the Bayes theorem:

$$P(\omega_i \,|x) = \frac{p(x \,|\omega_i) P(\omega_i)}{p(x)} \tag{41}$$

Where, $p(x)$ is the probability density function of x .

If we focus on the two-class case ω_1, ω_2, currently, the Bayes classification rule can be expressed as:

$$\begin{aligned} &p(x \,|\omega_1) P(\omega_1) > p(x \,|\omega_2) P(\omega_2) \Rightarrow x \in \omega_1 \\ &p(x \,|\omega_1) P(\omega_1) < p(x \,|\omega_2) P(\omega_2) \Rightarrow x \in \omega_2 \end{aligned} \tag{42}$$

$$\begin{aligned} &l(x) = \frac{p(x \,|\omega_1)}{p(x \,|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \Rightarrow x \in \omega_1 \\ &l(x) = \frac{p(x \,|\omega_1)}{p(x \,|\omega_2)} < \frac{P(\omega_2)}{P(\omega_1)} \Rightarrow x \in \omega_2 \end{aligned} \tag{43}$$

$l(x)$ is called the likelihood ratio and $\frac{P(\omega_2)}{P(\omega_1)}$ is referred to as the decision's likelihood ratio's threshold value.

Now, the current objective is to design the Bayesian classifier. Given that the related densities have an exponential form. Working with discriminant functions is recommended, in which the (monotonic) logarithmic function is involved $\ln(.)$:

$$\begin{aligned} &g(x) = -\ln l(x) = -\ln p(x \,|\omega_1) + \ln p(x \,|\omega_2) > \ln \frac{P(\omega_1)}{P(\omega_2)} \\ &\Rightarrow x \in \omega_1 \\ &g(x) = -\ln l(x) = -\ln p(x \,|\omega_1) + \ln p(x \,|\omega_2) < \ln \frac{P(\omega_1)}{P(\omega_2)} \\ &\Rightarrow x \in \omega_2 \end{aligned} \tag{44}$$

The Gaussian or normal density function is used in practice the most frequently, its tractability for computation and the fact that it effectively simulates a wide range of instances are the main drivers of its appeal [10].

$$\begin{aligned} &h(X) = -\ln l(X) = -\ln p_1(X) + \ln p_2(X) > \ln \frac{P_1}{P_2} \Rightarrow X \in \omega_1 \\ &h(X) = -\ln l(X) = -\ln p_1(X) + \ln p_2(X) < \ln \frac{P_1}{P_2} \Rightarrow X \in \omega_2 \end{aligned} \tag{45}$$

The decision rule of equation (45), in a particular scenario where $P(\omega_i \,|x)$ are Gaussian with expected vectors M_i and covariance matrices Σ_i , changes to [8]:

$$\begin{aligned} g(x) &= -\ln l(x) = \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) - \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) + \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} \\ &> \ln \frac{P(\omega_1)}{P(\omega_2)} \Rightarrow x \in \omega_1 \\ g(x) &= -\ln l(x) = \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) - \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) + \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} \\ &< \ln \frac{P(\omega_1)}{P(\omega_2)} \Rightarrow x \in \omega_2 \end{aligned} \tag{46}$$

VI. SIMULATION RESULTS

The following third order system is taken into account to demonstrates the efficacy of the offered approach

$$\mathrm{A(z)} = 1 + 0.52\mathrm{z}^{-1} + 0.61\mathrm{z}^{-2} + +0.34\mathrm{z}^{-3}$$
$$\mathrm{B(z)} = 0.42\mathrm{z}^{-1} + 0.31\mathrm{z}^{-2} + +0.68\mathrm{z}^{-3}$$
$$\mathrm{C(z)} = 1 + 0.45\mathrm{z}^{-1}$$
$$\mathrm{D(z)} = 1 + 0.75\mathrm{z}^{-1}$$

$u(n)$ is produced as a white sequence that has $\mathrm{m} = 0$ and $\sigma^2 = 1^2$ $\xi(n)$ is produced a Gaussian white noise that has $\mathrm{m} = 0$ and $\sigma^2 = 0.1^2$. The validation of this proposed algorithm is tested utilizing Mean Bias Error (MBE) when $\sigma^2 = 0.1^2$ and $\sigma^2 = 0.6^2$ respectively. MBE is the mean of the difference between the estimated and actual output, and it calculated using the following equation:

$$MBE = \frac{1}{N}\sum_{k=1}^{N}(\hat{y}(n) - y(n)) \tag{47}$$

The preferred value of MBE is the small value (close to zero) [11]. Fig. 2 demonstrates MBE versus (n) when $\sigma^2 = 0.1^2$ and $\sigma^2 = 0.6^2$, the conclusion from the figure is that the model effectiveness increases when the noise level decreases.

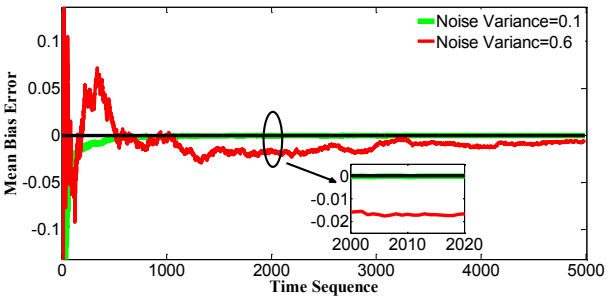

Fig. 2. Mean Bias Error With Two Different Noise Variances

The true output (Black Color), the estimated output (Red Color) and the residual (Blue Color), are plotted together as displayed in fig. 3. The figure illustrates that the estimated output of proposed algorithm is almost identical to the actual output and the residual is very close to zero. Now, assuming the fault has been happened during time $1000 \le k < 1600$, in order to show the model capability and examine the LDA efficiency for dimension reduction. Once the fault has been occurred, Changes will be made to the system's parameters and new assumptions of parameters have been made as: $\hat{\theta}_s(k) = [0.82, 0.91, 0.74, 0.62, 0.71, 0.88]^T$, $\hat{\theta}_n(k) = [0.85, 0.95]^T$

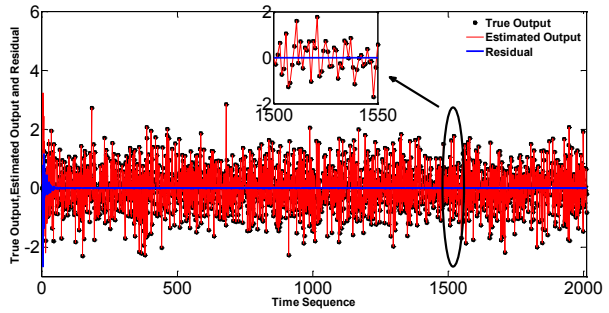

Fig. 3. The true output, the estimated output, and the residual.

These system parameters not provide the same level of informational value.

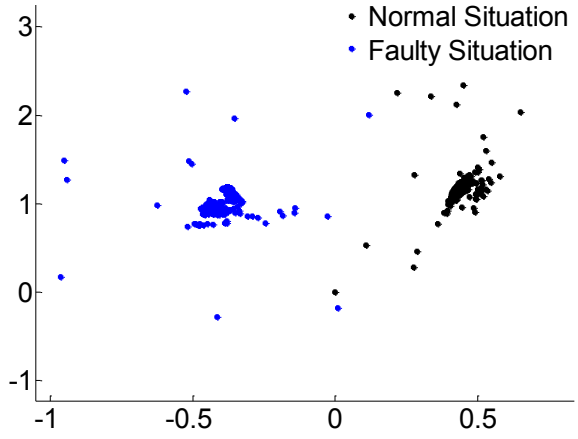

Fig. 4. two dimensions space for normal situation and faulty situation

Thus, dimension reduction is essential to select the parameters with the best system information, the parameters dimension reduction from n = 6 to m = 2 has been achieved. Fig.4 illustrates the reduced dimensions m = 2 for normal mode and faulty mode.The figure shows that the separations between two classes is very obvious. Fig. 5 illustrates the classification between these two classes e using Naive Bayes Classifier and the figure demonstrates that the Naive Bayes Classifier is very effective.

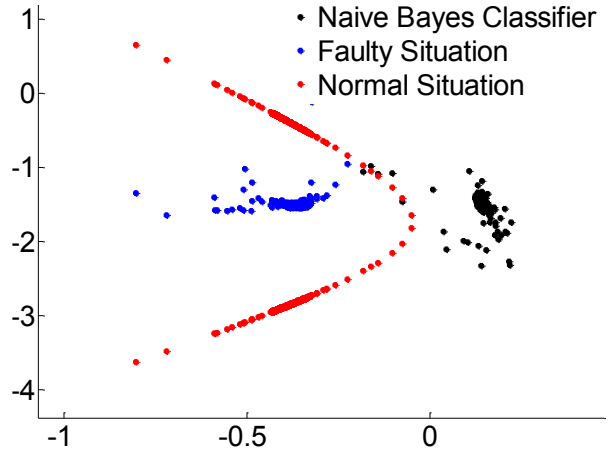

Fig. 5. The classification between fault-free mode and faulty mode based on Naive Bayes Classifier

VII. CONCUSIONS

In this work, Multi stage RLS algorithm has been derived. The validation of this model has been done based on mean bias error, the result shows this model is very effective. The estimation of the system parameters in both normal and faulty situations have been achieved as well as the dimension reduction of these parameters has been done in order to obtain their best important features utilizing LDA. The classification between these reduced classes has been performed using the Naive Bayes classifier. Finally, capability of this approach has been demonstrated with simulated data.

REFERENCES

[1] N. Shashoa , M. Hassan , A. Almukhtar,and A. Elmezughi , " Parameter Estimation and Residual Generation for (CARARMA) Algorithm Model Depend on DRGELS", 4rd IEEE Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT'2018), Istanbul, Turkey, April 18-19, 2018.

[2] N. A. Shashoa, W. A. Saheri, M. A. Arwin, and I. N. Jleta, "RLS Identification Model of Bearing System," 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences, 2021.

[3] Wang, D., Ding, F.: Input–output data filtering based recursive least squares identification for CARARMA systems. Digital Signal Processing 20 (2010).

[4] N. Shashoa, O. Jomah, O. Abusaeeda, and A. Elmezughi. "Feature Selection for Fault Diagnosis Using Principal Component Analysis." 2023 58th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), pp. 39-42. IEEE, 2023.

[5] J. Ding, "The hierarchical iterative identification algorithm for multi-input-output-error systems with autoregressive noise," Complexity, 2017.

[6] S. Wang and R. Ding, "Three-stage recursive least squares parameter estimation for controlled autoregressive autoregressive systems", Applied Mathematical Modelling. pp. 7489-7497, 2013.

[7] D. Manoranjan, "Dimensionality Reduction", Department of Information Systems, School of Computer Engineering, Nanyang Technological University, Nanyang Avenue, Singapore.

[8] K. Fukunaga, "Introduction to Statistical Pattern Recognition,", Academic Press,1990.

[9] D. Bhosale and R. Ade, "Feature Selection based Classification using Naive Bayes, J48 and Support Vector Machine", International Journal of Computer Applications. Vol. 99– No.16, August 2014.

[10] S. Theodoridis, "Pattern Recognition,", Academic Press,2003.

[11] E.O. Falayi, J.O. Adepitan and A.B. Rabiu, "Empirical Models for the Correlation of Global Solar Radiation with Meteorological Data for Iseyin, Nigeria," The Pacific Journal of Science and Technology, vol. 9. No. 2. November 2008.