Fault detection Automation in Distributed Control Systems using Data-driven methods: SVM and KNN

Ahmadi S.H., Khosrowjerdi M.J.

Abstract-- **Fault diagnostic methods with fuzzy logic methods, SVM, KNN and artificial intelligence systems have been used in complex systems such as wind turbines, gas turbines, power distribution systems, power transformers and rotary machines, but in the specific field of distributed control systems, the vacancy of this topic is strongly felt. Due to the need of the industry to detect faults quickly and in a timely manner in all modes of sensors, actuators, outputs and control logics to maintain expensive, valuable resources, important and complex equipment, it is very necessary to enter this topic. In this paper, a suitable theoretical and practical basis for diagnosing various types of faults in the DCS of a gas refinery is done. The fact that the operator quickly identifies the area and the cause of the fault can avoid huge losses in terms of downtime. Automation of fault diagnosis in DCS has not been explicitly mentioned in any article or book, and here the plan is presented for the first time. In this design, we connect MATLAB classification apps to the industrial system like DCS, then data are analyzed by SVM and KNN methods to detect faults. The results show that faults can be detected with a probability of more than 85% accuracy without the need for on-site expert force and with much less time.**

Note to Practitioners:

According to the fact that so far no specific method for fault diagnosis has been used in Distributed control systems (DCS) and there is still a need for expert and their presence on site, so we have tried to automate fault diagnosis in dcs systems. Let's. Of course, the use of data-driven methods such as SVM and KNN has been used in many cases, such as turbines, power transformers, rotary machines, to diagnose and classify faults, but in this particular case, DCS, is reported for the first time. The advantage of using this method is reducing the time of fault diagnosis and also not needing an on-site specialist, since in complex and expensive industries such as oil and gas industries, the issue of time and amount of production loss is very important. Finding faults of systems in a short time can be very important. Using this method is recommended even for other cases of fault detection in large scale systems where there are many variables such as Space Industries ,environmental issues and economic issues.

Index Terms— **Distributed Control Systems (DCS), Fault detection, SVM, KNN**

I.   INTRODUCTION

Work on fault detection began academically at MIT around the 1970s, during which time models and measurements were used to diagnose faults in one or more parameters or system response faults. In Fig. 1, fault f enters the process and is detected through residual theory by comparing the output of the exact model with the output of the actual system [1]. With the development of computers in the following years, the availability of data from systems became possible in large quantities and quickly, and this development also affected the methods of fault diagnosis. New data-driven methods were developed with the help of computer algorithms and machine learning to extend fault detection methods.
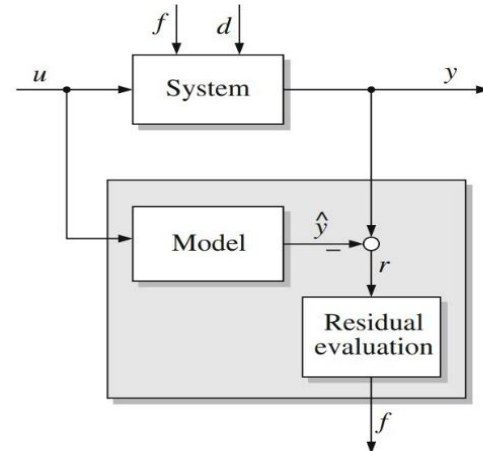

Fig. 1. Model-based fault diagnosis by the residual method

Model-based methods are very fast and have good accuracy, provided that we have a very accurate model of the system, which of course is usually not available to researchers in industrial conditions, or the exact model due to complexity and other there is no linearity of the system [2].

In these cases, data-based methods are used. To examine data-based diagnostic methods, it is first necessary to examine the tools needed to do so. Artificial intelligence now has a special place in the field of control, and from time-to-time new methods appear in the articles that are suitable for optimization, data analysis, and extraction of results from the mass of

Ahmadi S.H. is with the Sahand University of Technology, Tabriz, Iran (e-mail: sh_ahmadi98@sut.ac.ir).

Khosrowjerdi M.J. is with the Department of Electrical Eng., Sahand University of Technology Tabriz, Iran (e-mail: khosrowjerdi@sut.ac.ir).

scattered data. These methods include fuzzy logic, genetic algorithm, ant colony algorithm, SVM, KNN and etc.

Most of these methods have relatively complex algorithms, but in 1995 Mr. Vapnik [3] proposed a method that, while simple, has a solid foundation. This method is called SVM or support vector machine.

This method has been used in a variety of applications because its implementation is relatively simple and also the improvement of its algorithm has continued since that year.

SVM has a special algorithm very similar to the function of perceptron in neural network that can distinguish, separate and classify information. This is done by first learning a machine with a predefined data set and then asking the machine to isolate new data for healthy or faulty data based on the learned algorithm.

This method is for separating healthy and faulty data or two different groups, which is also called binary classification.

This method is especially useful where the system model is very complex or not available. Without a precise model for highly sensitive systems, such as aircraft or spacecraft and complex industrial systems, it is recommended to use the SVM method to detect faults in system model parameters, sensors and actuators.

In this method, data related to two different classes are simply separated by a line or a hyperplane. For example, in Fig. 2, the data are separated into two different groups.
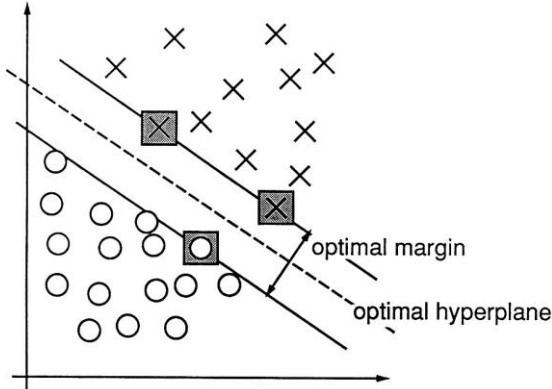

Figure 2. An example of two-dimensional separation is that the gray squares are support vectors

Even if the data are not separated by a line, for example in Fig. 3, they can be separated and classified by a curved line with a simple technique called Kernel [4].
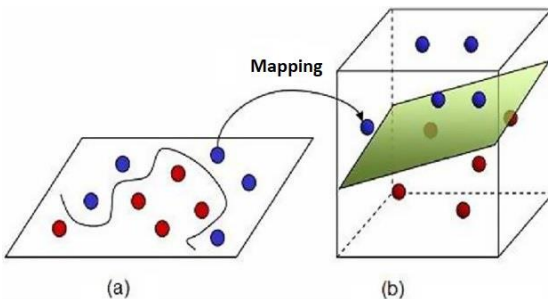

Figure 3. Using kernels to Classify data that cannot be separated by a line

A. Number of recent articles on fault diagnosis with SVM

From years 2000 to 2021 articles on fault detection with SVM in the IEEE database are including 590 articles, which 103's is journal and about 479's are conference papers. However, given that there are about 42,392's articles on fault detection, a small percentage seems to be devoted to this topic, about 2% but in recent years, about 10%. In ScienceDirect, a total of 75,827's articles and about 5184 articles have been worked on the fault detection by SVM method, which is about 4%, and in recent years, 12% of the total articles (Table I.). Therefore, there are signs of growth and more interest in working on this issue.

TABLE I. NUMBER OF ARTICLES ON THE TOPIC FD AND FD WITH SVM

Number of Papers 2000-2021	Fault Detection (FD)	FD with SVM
IEEE	42000	600
ScienceDirect	75000	5000

B. Applications of Fault detection with SVM method

Data-based methods have been used in many systems. For example, in fault detection in wind turbines, power distribution networks, power transformers, etc., of course, in distributed control systems, in cases such as adjusting the parameters of PID controllers or obtaining a proposed model for a real system such as a boiler [5], But in this paper, we want to use this method in fault diagnosis to complete the puzzle of using these methods in large scale control systems like DCS.

We first introduce some important articles that have done good work in this field of troubleshooting with the SVM method, then describe our purpose with an example. In Widodo article [6] the SVM method is introduced and then a review of the articles in their method of dealing with fault diagnosis has been done. To diagnose a fault, data processing and preprocessing is neccessary. Also, in SVM method, to extract the fault, features must first be extracted. It should be noted in what characteristics healthy data and unhealthy data are different.

In some articles, such as [7], vibration sensors are used to detect bearing faults. When the signal harmonics increase in speed, these harmonics can be identified and troubled from the acceleration spectrum as a feature. Of course, other types of feature extraction methods such as power spectrum, closed spectrum, classical spectrum, and autoregressive model can also be used. Statistical feature extraction is time and frequency that are extracted from vibration signals, [8] and in order for feature extraction and thus fault diagnosis to be done well; we need an appropriate amount of training data and otherwise detection, can not be done with great accuracy.

Most articles are about troubleshooting in mechanical rotating systems and machines, transmission systems, gearboxes, roller bearings, induction motor faults and turbo pumps, but in electromechanical machines, semiconductors, cooling systems, systems Air conditioning and chemical processes such as the Tennessee Eastman have also been used.

Design steps:

- Systems are described and the types of faults and their location are defined.

- Training begins in SVM.

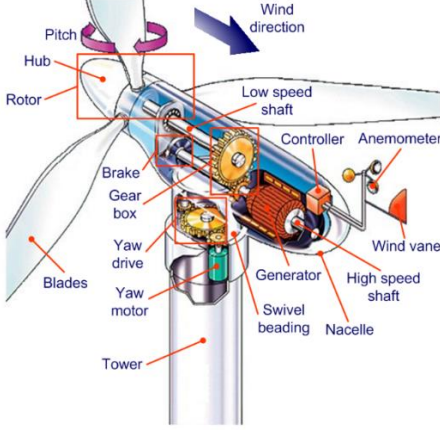- The simulation is done with a series of real data.

Fig 4. Wind turbine and its components [9]

"Fig. 4", shows a 3-blade wind turbine with 3 dual sensors on each to accurately measure the blade angle. This system is very complicated for modeling and we have to use data-driven methods instead of model-base methods for fault diagnosis.

In paper [10] Sa has been working on the detection of sensor defects, process and actuators of a wind turbine by SVM method.

In [11] want to examine how the definition of the feature vector, X is done. Increasing training data greatly reduces SVM efficiency. Increasing the X vector properties from 5 to 10 does not have much effect on the accuracy of the effect.

Detection and classification of sensor faults ensures the reliability and safety of systems and is therefore an important issue. This article examines the types of sensor faults, including irregular faults, thrust faults, elevation faults, impact faults, trapping faults or adhesions (Fig. 5.).
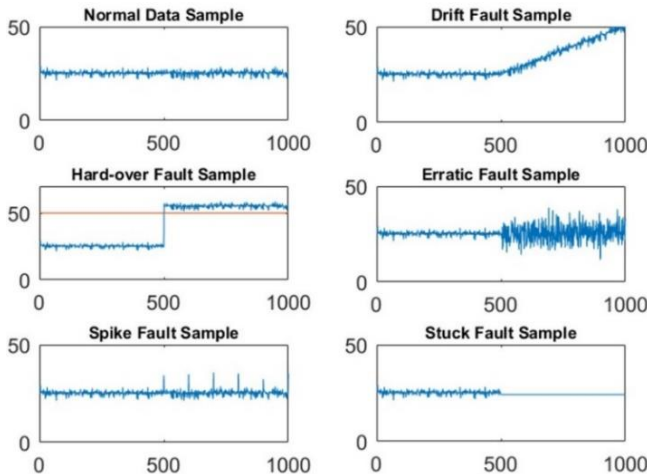

Fig 5. Types of sensor faults in normal and defective condition

An interesting task that has been done in this article is to bring the types of sensor faults, and it is specified that sensor faults can vary according to their response characteristics, and to simulate, implement and detect the fault, a feature must be Sensor output statistics should also be considered. For example, a sensor that has a spike fault cannot be identified in the way that the fault is of the overhead type, and the characteristics of these faults are different.

In book [12] the issue of fault diagnosis in rotating machines has been researched. Since the types of faults of these machines, including mechanical and electrical faults, sensor faults, etc., are very high, the use of practical and simple methods such as SVM is recommended.

The method of using this algorithm is summarized as follows. Each fault has a specific characteristic. Based on the characteristics of that fault, it can be related to a specific category. Linear or non-linear classification is based on the needs of the problem, and multi-class classification is implemented as one against all or one against one if needed.

If the data cannot be separated linearly, other types of kernels are used. Kernels are functions used in the SVM algorithm for nonlinear classification, such as the Gaussian kernel or the RBF kernel.

Multi-class SVM is also done in three ways:
• Compare each class with all classes
• Compare each class with another class
• DAG method

In these methods, the same binary algorithm, i.e., two classes, is used several times or in parallel and is implemented in MATLAB program.

This book describes how faults such as motor vibration, motor over current, overload and multi-class SVM are detected. First, the learning algorithm is performed with a small amount of training data and then new input information is called to the system to distinguish healthy data from faulty ones. Important articles on rotary motor fault diagnosis by SVM method from 2001 to 2018 have been collected and introduced in this book.

II. IMPLEMENTATION AND SIMULATION OF SVM ALGORITHM IN DCS

According to the increase in production in industries and the desire for more automation, the need for automatic troubleshooting has also increased. Fault detection methods are either model-based or data-based. In the model-based method we need a very accurate and comprehensive model of the system, while in the data-based method we need a history of the data and the existing pattern.

A. Proposed method

In the discussion of control, when the number of variables are very small, it is better to use the model-based method, but in complicated systems or the so-called Large Scale, the model-based methods can no longer be used to troubleshoot the system, because building a mathematical model will be very complicated and time consuming or basically impossible.

Consider a system in which three drums with almost identical characteristics are operating. Each section has multiple control valves, pressure sensors, temperature, level

and flow. In this system, with all the variables, an event or a miraculous event occurs. Which model can detect the fault? How to model a variable? It is best to store the data that is normally available in industrial control systems, and analyze it. If it follows a certain pattern, there is no fault, but if it deviates from the pattern, we are looking for, there is probably a fault. Occurs and is detected by the troubleshooting system. Therefore, although the base model troubleshooting methods are more attractive, but in large systems with a large number of variables and sensors and operators and complex functions, the use of these methods is not cost-effective and efficient.

Mathematical modeling for systems that affect more than a few variables is really complex and time consuming. That's why data analysis methods came into being. But the problem that may occur in data-based fault detection is that new conditions may occur in the system that were not previously planned or predicted in the template, in which case the template needs to be updated accordingly. So here, with a simple software update, this problem can also be solved, while the classic model-based fault diagnosis cannot be easily changed and requires hardware changes.

In model-based classical fault detection, we define any difference between the mathematical model and reality as a residual signal. If this signal exceeds a threshold, it indicates a fault in the system. Now, if the model is not accurate or there are changes in the real system that the model no longer adapts to, how can we correct the fault diagnosis? These are the questions that arise in this case. This can be solved in the case of data-based diagnostics with a simple software update.

In data-based systems, we can define the fault ourselves. We can define any fault in the signal output or in the system control logic, and these faults are given to the computer by a template or program. The actual system transmits the data to the computer, the data analysis software analyzes the data, and reports a specific fault (Fig 6.).
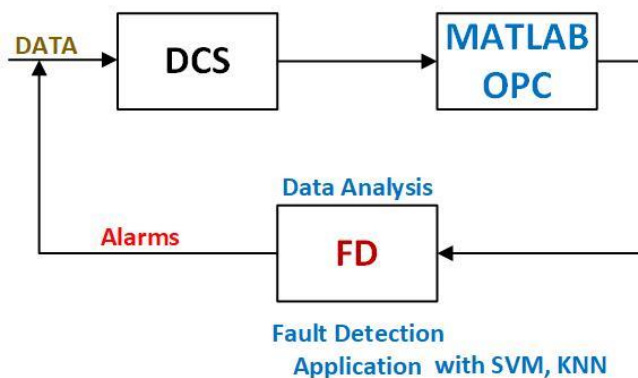

Fig 6. FDA (Fault Detection Automation): Connect DCS system to MATLAB OPC with FD application

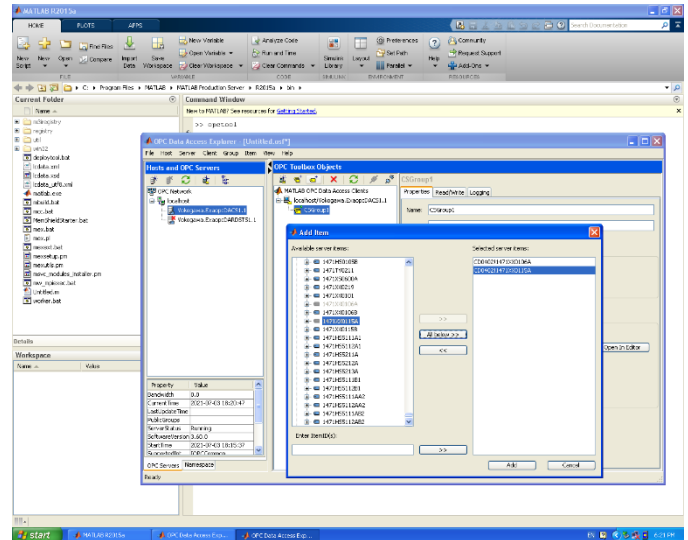Therefore, we can define any fault that is possible, provide

Figure 7. Connection between MATLAB and ExaOPC

its model to the data analysis application program, for example, MATLAB, and by connecting the data from the real system to the MATLAB program and checking them, the defined fault in the case recognizes the occurrence. "Fig. 7" shows the connection between MATLAB and ExaOPC related to Yokogawa Centum DCS.

Another advantage of data-driven fault detection of control systems is that, since these systems are usually data-driven and information and data are widely available, we do not need additional hardware in systems. In the model-based, special hardware is needed to form the evaluation system, and the hardware itself may be damaged or need to be repaired, and all this is costly, but in data-based systems, only a small laptop can be used for example is sufficient to support MATLAB, so it can easily be used in aircraft, spacecraft and other large scale or complicated systems.

In summary, the advantages of data-based fault diagnostics are as follows:

- Ability to constantly update
- Define new faults in the system for diagnosis
- Flexibility with changes to the actual system model
- Ability to implement on very complex systems
- Fault detection in systems with very large variables that seem impossible in model-based mode
- Implementation on large scale systems

B. Fault detection in DCS using MATLAB fault classification

The following figure Fig. 8, shows a part of the shutdown diagram of the boiler, which is of complex logic due to the fact that in the sequence of dryers of unit 114 Fig. 9, of a gas refinery (the unit that separates water and CO2 from gas molecules) Used and usually many problems occur in this unit, troubleshooting this part is time consuming and difficult.

In this part, it is always a challenge to quickly find the cause of the fault and fix it in a short time, because the sequence cannot be kept in one state for a long time.
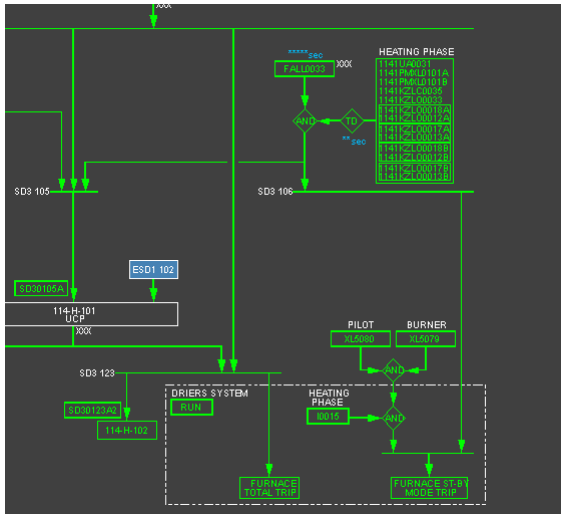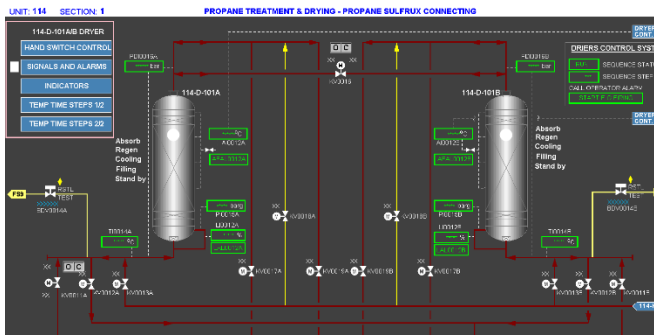
Fig 8. logic bar sequence dryers and furnaces

Fig 9. Dryer drums unit 114 that are in the absorption and resuscitation service during a sequence

The idea of troubleshooting in DCS is that to distinguish any type of fault in complex logics, several features can be specified as input and fault as the output of a classification syste.

For example Table II, lists the conditions for the occurrence of a type of fault. This table is called feature extraction and Table III, is training data.

The data is selected from the DCS system to a data collection system, and in this system the software is also installed. The stored data is provided to the processor. The definition of several faults and training data for learning the faults are done by the system, respectively.

TABLE II. FEATURES LIST AND FAULTS TYPE OUTPUT

Step no.	1141FT0033	1141I001 500	OUTPUT (FAULT)
14	<2	1	"STOP NO FAULT"
14	>2	1	"FT0033 FAULT"
14	<2	0	"VALVE POSITION FAULT"
14	>2	0	"CORRECT OR NOT DEFINED"
<> 14	DON'T CARE	0	"CORRECT OR NOT DEFINED"

TABLE III. TRAINING DATA

| step no.	1141FT0033	1141I0015	OUTPUT (FAULT)
14	1.1	1	"STOP NO FAULT"
14	1.2	1	"STOP NO FAULT"
14	1.5	1	"STOP NO FAULT"
14	0.5	1	"STOP NO FAULT"
14	0	1	"STOP NO FAULT"
14	-0.05	1	"STOP NO FAULT"
14	1.9	1	"STOP NO FAULT"
14	2	1	"STOP NO FAULT"
14	1.7	1	"STOP NO FAULT"
14	0.9	1	"STOP NO FAULT"
14	2.1	1	"FT0033 FAULT"
14	2.5	1	"FT0033 FAULT"
14	4	1	"FT0033 FAULT"
14	3.8	1	"FT0033 FAULT"
14	12	1	"FT0033 FAULT"
14	9	1	"FT0033 FAULT"
14	2.6	1	"FT0033 FAULT"
14	25	1	"FT0033 FAULT"
14	12.5	1	"FT0033 FAULT"
14	18	1	"FT0033 FAULT"
14	1.1	0	"VALVE POSITION FAULT"
14	1.2	0	"VALVE POSITION FAULT"
14	1.5	0	"VALVE POSITION FAULT"
14	0.5	0	"VALVE POSITION FAULT"
14	0	0	"VALVE POSITION FAULT"
14	-0.05	0	"VALVE POSITION FAULT"
14	1.9	0	"VALVE POSITION FAULT"
14	2	0	"VALVE POSITION FAULT"
14	1.7	0	"VALVE POSITION FAULT"
14	0.9	0	"VALVE POSITION FAULT"
1	2	0	"CORRECT OR NOT DEFINED"
2	3	0	"CORRECT OR NOT DEFINED"
3	4	0	"CORRECT OR NOT DEFINED"
4	0.5	0	"CORRECT OR NOT DEFINED"
5	0	0	"CORRECT OR NOT DEFINED"
6	-0.5	0	"CORRECT OR NOT DEFINED"
7	18	0	"CORRECT OR NOT DEFINED"
8	25	0	"CORRECT OR NOT DEFINED"
9	2	0	"CORRECT OR NOT DEFINED"
10	4	0	"CORRECT OR NOT DEFINED"
11	2.7	0	"CORRECT OR NOT DEFINED"
12	2.8	0	"CORRECT OR NOT DEFINED"
13	12	0	"CORRECT OR NOT DEFINED"
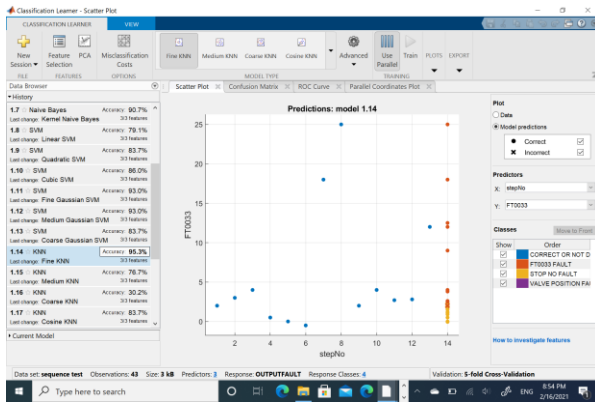
Figure 10. MATLAB classification application on training data

The data is stored as a feature and type of fault in an Excel file and entered into the MATLAB. With the MATLAB classification application, the faults are separated and the system is trained by all methods, including KNN and SVM.

There are two stages, the first is the training stage Fig. 10 and the second is the testing stage Fig. 11. In test stage the system is ready to diagnose the faults from the new data. Result of testing data is shown in Table IV. that compare real fault happen with output of fault detection application.

TABLE IV. OUTPUT TABLE FROM MATLAB

| step no.	1141FT00 33	1141I001 5	Real output expected	MATLAB model output
14	0.5	1	"STOP NO FAULT"	' STOP NO FAULT '
14	2.5	1	"FT0033 FAULT"	' FT0033 FAULT '
14	2.8	0	"CORRECT OR NOT DEFINED"	' VALVE POSITION FAULT '
14	8	1	"FT0033 FAULT"	' FT0033 FAULT '
14	0.7	1	"STOP NO FAULT"	' STOP NO FAULT '
14	1	1	"STOP NO FAULT"	' STOP NO FAULT
14	1	0	"VALVE POSITION FAULT"	' VALVE POSITION FAULT '
14	12	0	"VALVE POSITION FAULT"	' CORRECT OR NOT DEFINED '
12	1	1	"CORRECT OR NOT DEFINED"	' STOP NO FAULT '
11	0.5	0	"CORRECT OR NOT DEFINED"	' CORRECT OR NOT DEFINED '
10	2	0	"CORRECT OR NOT DEFINED"	' CORRECT OR NOT DEFINED '
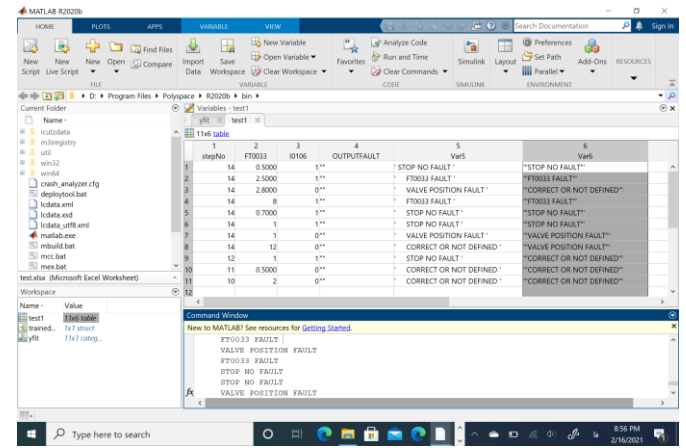
Figure 11. Test data

The percentage of fault diagnosis accuracy is usually higher than 85% to about 95%.

III. REFERENCES

[1] M. . Khosrowrjerdi, *Fault Detection and Compensation methods in Control systems*, First. Tabriz, Iran: Sahand University of Technology, 2016.

[2] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, "Introduction to Diagnosis and Fault-Tolerant Control," in *Diagnosis and Fault-Tolerant Control*, 2016, pp. 1–35.

[3] C. Cortes and V. Vapnik, "Support Vector Networks," *Mach. Learn.*, vol. 20, pp. 1–25, 1995.

[4] E. Azimifar, Zohreh. Ahmadi, *Statistical Identification of Pattern / Volume III*, First. shiraz: Navid, 2017.

[5] Y. Wang, Z. Tang, and B. Zhao, "Modeling of Boiler Steam Flow Based on Adaptive Least Squares Support Vector Machine," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2017, vol. 1, pp. 187–190, doi: 10.1109/IHMSC.2017.49.

[6] A. Widodo and B. S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007, doi: 10.1016/j.ymssp.2006.12.007.

[7] Z. Tong, W. Li, B. Zhang, F. Jiang, and G. Zhou, "Bearing Fault Diagnosis under Variable Working Conditions Based on Domain Adaptation Using Feature Transfer Learning," *IEEE Access*, vol. 6. pp. 76187–76197, 2018, doi: 10.1109/ACCESS.2018.2883078.

[8] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, 2016, doi: 10.1016/j.neucom.2015.09.081.

[9] N. Laouti, N. Sheibat-Othman, and S. Othman, "Support vector machines for fault detection in wind turbines," *IFAC Proc. Vol.*, vol. 44, no. 1 PART 1, pp. 7067–7072, 2011, doi: 10.3182/20110828-6-IT-1002.02560.

[10] F. P. G. De Sa, D. N. Brandao, E. Ogasawara, R. D. C. Coutinho, and R. F. Toso, "Wind Turbine Fault Detection: A Semi-Supervised Learning Approach with Automatic Evolutionary Feature Selection," *International Conference on Systems, Signals, and Image Processing*, vol. 2020-July. pp. 323–328, 2020, doi: 10.1109/IWSSIP48289.2020.9145244.

[11] S. U. Jan and I. S. Koo, "Sensor faults detection and classification using SVM with diverse features," *International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017*, vol. 2017-Decem. pp. 576–578, 2017, doi: 10.1109/ICTC.2017.8191044.

[12] N. Asoke and H. Ahmed, "Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machine," 2019.

IV. BIOGRAPHIES

Seyed Hossein Ahmadi was born in 1977 in Shiraz-Iran. He graduated in 1991. In 2002, he received a bachelor's degree in electronics and then in 2012, a master's degree in instrumentation and automation in gas and oil Industries from Shiraz University. He worked for Siemens (Shiraz) and SPGC (Assaluyeh). He has also taught courses at the Shiraz University of Applied Sciences. His research interests include designing electronic circuits such as power supplies, solar power systems, PLC, distributed control systems DCS, instrumentation and now fault detection in complex systems. He is currently, 2021 a PhD student in Control at the Faculty of Electrical Engineering, Sahand University of Technology, Tabriz-Iran.

Mohammad Javad Khosrowjerdi received the B.S., M.S., and Ph.D. degrees in electrical engineering from the K. N. Toosi University of Technology, Tehran, in 1993, 1996, and 2003, respectively. He is currently a Professor with the Electrical Engineering Department, Sahand University of Technology, Tabriz, Iran. His current research interests include fault detection and isolation, fault tolerant control, multi-objective control, and robust and nonlinear control.