Journal of Integrated Engineering and Technology (JIET), 1(1), 2024, 55-68

Published by: ***Higher Institute of Engineering and Technology, Kafrelsheikh*** **(KFS-HIET)**

Journal homepage: https://jiet.journals.ekb.eg/

Print ISSN 3009-7207 **Online ISSN** 3009-7568

An Intelligent Optimized Digital Twins Framework Based on Fault Diagnosis in Complex Control Systems

Samar M. Zayed

Electrical Engineering Department, Higher Institute of Engineering and Technology, Kafrelsheikh, Egypt

Received: 2 March 2024; Revised: 7 June 2024; Accepted: 10 June 2024

ABSTRACT: Digital Twins (DT) is considered as the backbone of several industrial systems in manufacturing category. The DT strategy has a vital role for dataset generation especially in fault prediction and diagnosis aspects. Recently, these approaches are considered the tending in research by utilizing the support of Artificial Intelligence (AI) techniques for critical industrial applications. The virtual assets of DT can produce a performance that is close to the real counterpart, which is an opportunity for fault diagnosis and prediction under different fault conditions. Therefore, this study proposes an intelligent AI-based framework that is based on Genetic Algorithm (GA) and machine learning classifiers (MLCs) such as Logistic Regression (LR), Linear Discriminant Analysis (LDA), Naïve Bayes (NB), and K-nearest neighbors (KNN) for industrial digital twins systems namely Transmission System (TS) model. The proposed hybrid GA–ML framework is validated using a simulated dataset which is generated from TS model. The proposed framework achieves superior results for MLCs such as LR, LDA, NB, and KNN with accuracy equal to 96.5%, 98.3%, 97.4%, and 97.4 % compared with the ordinary MLCs with 87.3%, 87.3%, 82.5, and 85.7% respectively. Also, it is considered as a superior compared with the existing model's performance for diagnosing the complex future faults. So, the proposed framework will efficiently help for diagnosing and detecting faults in several manufacturing inspects.

KEYWORDS: Digital Twins (DT), Genetic Algorithm (GA), Machine Learning (ML), Fault Diagnosis, Industrial Control Systems.

1. INTRODUCTION

Recently, DT is considered the base of several infrastructures several categories [1-3]. DT is based on data exchange from remotely devices and learning methods. These categories indicate the future events that can be occurred. Moreover, US researchers have great efforts for providing extensive knowledge to data-driven research [4]. The DT is based on three components: the real asset, its counterpart, and the data exchange between the two models as indicated in Fig 1. The communication between the real system and its avatar is important for preserving the vitality of digital twins. In general, the virtual model can produce a simulated data close to its real counterpart. In other words, the DT provides new strategies for simulated data production under different instances. So, recently, DT is keeping eye catching in the entire aspects of manufacturing category [5].

Recently, the concept of DT took a different direction and was listed as one of the hot topics and top trends in several categories, specially manufacturing which is the main scope of this work [6-8]. Moreover, DT provides several strategies such as AI, big data, and IoT to make a seamless connection between physical and virtual environments to simulate "what-if" scenarios [9]. The main issue appeared during raising the complexity of manufacturing operations. As a result, several failures have been occurred which caused irretrievable losses. Therefore, fault diagnosis has always been an important aspect of manufacturing category [10].

The DT has a vital role with ML models for managing and controlling systems. Paving the way, intelligent systems become a witness to systems development [11]. By ML development, knowledge acquisition is not fully dependent on experience, but it uses other approaches based on data [12-14]. Data based on ML methods can produce highly effective data with high information with reflects the essence of data without missing [5]. Unfortunately, several ML models are the key for diagnosing system failures such as support vector machines (SVMs) [15], Decision Trees [16], Bayesian Networks (BN) [17], Artificial Neural Networks (ANNs) [18-21].

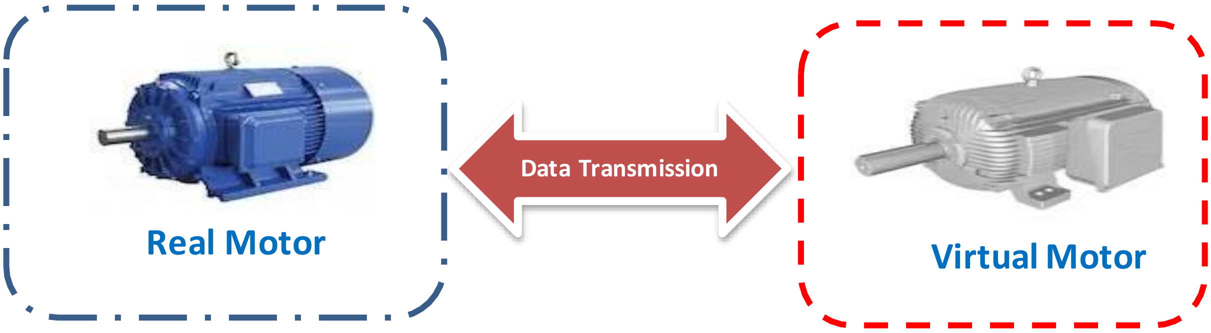


Fig. 1. DT Industrial model.

There are a lot of approaches have been introduced for classifying and detecting system faults such as IFD method [22]. The IFD is based on three steps: data acquisition, extracting features, and classifying failures [23, 24]. Feature extraction is based on extracting features that are produced from data according to several aspects such as time domain, frequency domain, and time-frequency domain [25]. Finally, AI strategies can be added for predicting system failures such as DL models and ML models [26, 27].

Therefore, this study proposes a new DT-assisted machine learning strategy based on optimization algorithm for diagnosing failures. The proposed framework depends on a simulated data for detecting future system events. The hybrid approach is based on MLCs and GA to increase the DT model performance for predictive maintenance. The hybrid framework is evaluated through a case study of TS model for diagnosing and predicting future failures that can be occurred. This study contributes by several points as follows:

- An efficient framework is proposed based on DT model, GA, and MLCs, to produce system failures close to those of the real one. The GA-ML model is based on GA and MLCs such as LR, LDA, NB, and KNN.
- The proposed GA-ML model is validated using a simulated dataset which are produced from TS model. The mentioned framework is based on two strategies: a SMOTE method for imbalanced dataset and GA for binary fault classification.
- According to results, the ordinary LR and LDA models achieve an accuracy of 87.3 %, while the NB and KNN gives an accuracy of 82.5 %, and 85.7%, respectively. After applying the genetic algorithm, hybrid GA-LDA model gives the highest accuracies of 98.3 %. while the GA-NB, GA-LR, and GA-KNN give accuracies of 97.4%,96.5, and 97.4%, respectively.

The rest of this paper is organized as follows. Section 2 discusses the previous studies while Section 3 introduces the hybrid GA-ML model. Section 4 discussed the simulation results. Furthermore, Sections 5 and 6 present the discussion, limitations, and future work, respectively. Finally, the conclusion of this study introduced in Section 7.

2. PREVIOUS STUDIES

Recently, digital twins systems had a vital attraction, especially for predictive maintenance in manufacturing category. Table 1 summarized earlier studies in this field and the employment of AI approaches for fault diagnosis category. The authors in [28], presented a DT model which is based on predictive maintenance by linking several components for raising system information and functions. Each component in this model is considered a distinct sub-model to form the final system. Furthermore, In [29], a new intelligent approach is depicted for predicting "what – if" scenarios by using the root mean square which outperformed to catch system failures before occurrence. While In [30], the study is based on building a DT model which is based on combining IoT strategies for data transmission in real-time and predictive maintenance by using several ML approaches for system investigation. Similarly, the authors in [31] focused on the previous concept for improving the wind turbine system efficiency by using a hybri d framework which is based on IoT aspects and intelligent approaches. Moreover, the authors in [32] proposed a framework which is based on DL strategies for developing the current methods for detecting future events in the complex systems.

Moreover, in [33] the authors discussed integrating AI approaches with CPSs for improving the performance of manufacturing productions. Unfortunately, Utilizing DL approaches achieved a superior result by emerging with DT categories for the detection failure processes which is presented in [34]. In addition, in [35] the authors introduced a new vision of smart production by integrating DT concepts with IIoT applications which increase the efficiency of predicting and detecting the future strange events. While in [36], a new model is presented for rotating machine which is constructed according to analyzing specific parameters to produce adaptive framework with highly performance. Moreover, the authors in [37] introduced a review study for evaluating the performance of predictive maintenance according to building life cycle. Zhang et al. depicted a new framework that depended on DT concept and AI approaches for tackling insufficient data issues for detecting the faults that are generated from complex systems such as rolling bearings [38].

Recently, Hassan et al. proposed a DT method detecting faults in autonomous ships. The data is collected sensor array in real-time [39]. In addition, the authors in [40] who proposed a study by developing a DT model which depended on several categories such as IoT approaches, remotely sensors, and digital twins strategies to improve PV systems performance. On the other side, several research are going to utilize AI approaches for diagnosing and predicting several anomalies. Most of research ensured that intelligent techniques achieved suitable performance compared with the previous strategies. Deeply, Zhang at al. introduced an approach by using intelligent methods for imbalanced data and faulty samples restriction for detecting accurate failures [41]. Cheng et al. presented a framework which is based on combining deep learning methods and continuous wavelet transform in rotating machine system [42].Moreover, Gao et al. introduced a study which reviewed fault diagnosis approaches and applications with respect to model and signal perspectives [43].Similarly, the authors in [44] discussed a review study of linear discrete time-varying systems from the perspective of diagnosing and detecting the system failures. Turned to Wang et al. who depicted an intelligent framework by utilizing DNN model with a new strategy for samples classifications in PHM systems [45].

Unfortunately, Samanta et al. reviewed a study which is based on MLCs for detecting multiple failures that can occur in lithium-ion battery systems [46]. In addition, Xing et al. proposed a framework by considering several techniques such as data-driven techniques and ANN model for failures that are generated from fuel cell system [47]. Moreover, Taqvi et al. developed a new a strategy that is based on deep learning models for complex fuel cells. The study achieved satisfactory results with respect to other models for early detection and prediction faulty instances [48]. So, authors in [49] presented a study for tackling data issues that produced from multiple faulty instances in PV systems. The study is constructed by utilizing CNN model and some intelligent techniques for detecting and diagnosing system failures. The authors in [50] applied a study for CRWU dataset which is widely utilized in research of the fault diagnosis approach. The study is considered as a hybrid approach that combined selecting features techniques with optimization and ML models. Moreover, Li et al. introduced a study that summarized the previous studies that is based on using intelligent approaches for fault detection in AF systems [51]. The authors in [52], developed an intelligent approach for faulty conditions that are generated from air compressor system in real-time. The study consists of several stages such as extracting features, selecting features, and classification. In addition, in [53] proposed a framework that constructed according to two strategies such as DRL models and A2C algorithm for two rolling bearing datasets. The results of the framework ensured its effectiveness for detecting several faulty instances. Consequently, this paper presents a new AI-based approach by utilizing GA and MLCs for complex TS model. The hybrid GA-ML framework is a combination of GA, and MLCs such as LR, LDA, NB, and KNN for early detection and prediction of the anomalies events.

Table 1: DT Applications for Fault Diagnosis approach.

Work	Application	Explanation	Challenges
[28]	A combination of several information and functions for producing AAS model.	The model depends on linking several components for raising system information and functions to form the final structure.	Improving the model performance by utilizing ML approaches
[29]	Wind Turbines with OPC-UA protocol for data streaming.	Presenting and evaluating data streams for predicting the future events.	Predictive system failures before occurring by intelligent techniques
[30]	Creating a new model which is (IoTwins)	Emerging two categorical phases such as IoT and ML models for building an efficient digital twins' system.	The effect of the cloud computing and ML approaches for real-time applications.
[31]	Wind Turbine Plants	The model has a sensor for tracking data using IoT approaches.	Improving system performance at fog nodes, and sensor data preprocessing before sending to cloud
[32]	DTDL-CPS for Smart Manufacturing	Integrating AI approaches with CPSs for intellectual development.	Combining deep learning approaches with CPSs for immediately system failure detection.

Table 1: Continue.

Work	Application	Explanation	Challenges
[33]	Prognostics and Health Monitoring (PHM)	Utilizing DL approaches for evaluating and monitoring the data status.	The implementation of detecting anomalous or faulty behaviors.
[34]	Electrical Machines	The model is based on Intelligent categories for machine diagnostics.	Using physical data from the real asset, by low-cost microcontrollers in real-time.
[35]	Steam Turbine	A Combination of the concept of digital twins and IIoT applications for detecting the anomalies events.	Improving the performance of the smart productions.
[36]	Rotating Machinery	The model is constructed according to parameters operations and its sensitivity.	Enhancing the system performance for achieving adaptive analysis and diagnosis.

3. Proposed Hybrid GA-ML Framework

This paper introduced a new hybrid intelligent framework for fault classification. The framework depends on three stages. Firstly, imbalanced data were produced from the simulated DT model of TS model. Secondly, using SMOTE technique for imbalanced dataset. Thirdly, the hybrid GA-ML model is applied for binary fault classification. The hybrid framework is detailed as shown in Fig. 2 the following:

- **Stage 1:** Dataset generation from the simulated model of a TS model under binary fault conditions.
- **Stage 2:** Using SMOTE technique for unbalanced data.
- **Stage 3:** Training and testing data splitting.
- **Stage 4:** utilizing a combination of GA and Machine Learning models such as LR, LDA, NB, and KNN.
- **Stage 5:** Evaluate the performance according to the corresponding MLCs for binary fault classification.

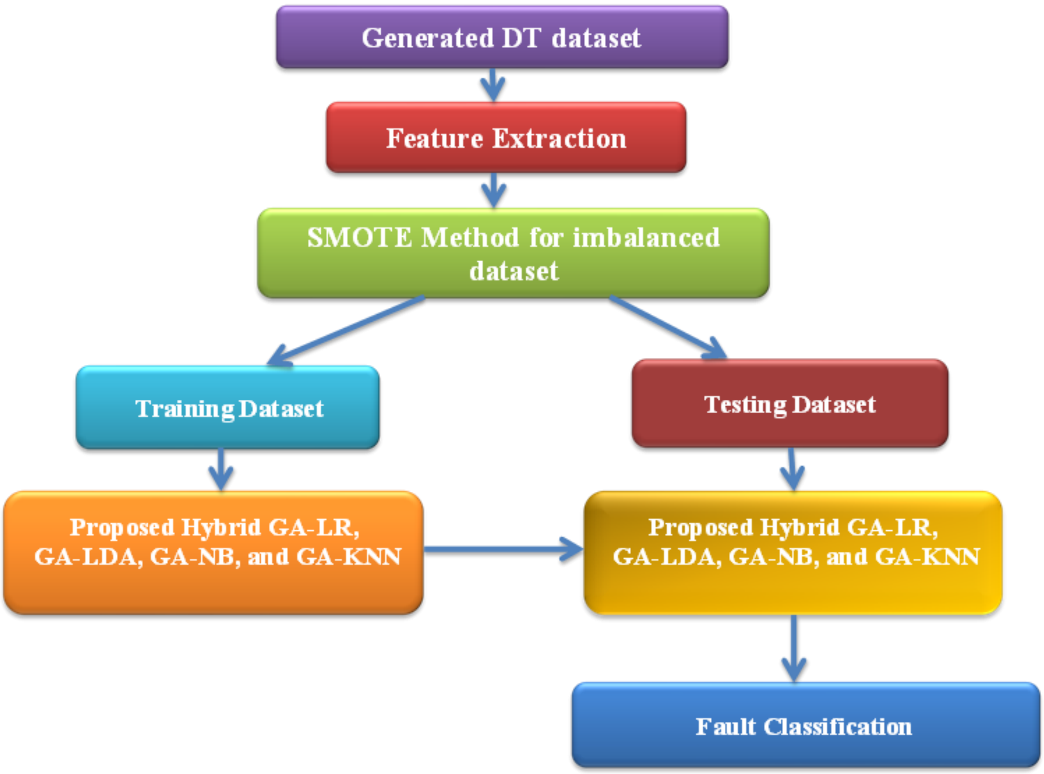


Fig. 2. The Proposed Hybrid GA-ML Framework.

3.1. Simulated DT Model

Firstly, the data will be produced from industrial control system which is TS model. The actual system is simulated using sims cape in MATLAB [54], as shown in Fig. 3. The simulation model was implemented according to the real one to generate the simulated data which is considered the container of all failure events.

3.2. Feature extraction

Feature extraction is the key for reducing the corresponding samples. Several cases suffer from many variables which requires many resources. So, the proposed framework aims to select and convert variables into the original features of the dataset. But data is not enough for detecting and diagnosing system failures which required large amount of data. Thus, extracting features is considered a crucial stage for final decision-making. In this work, the feature extraction method used is Principal Component Analysis (PCA). PCA plays a significant role in selecting effective variables in large datasets, especially in classification and regression approaches. Moreover, PCA is the most popular method utilized in datasets that contain anomalies or faults because it is obtained as a noise in the system response curve [55].

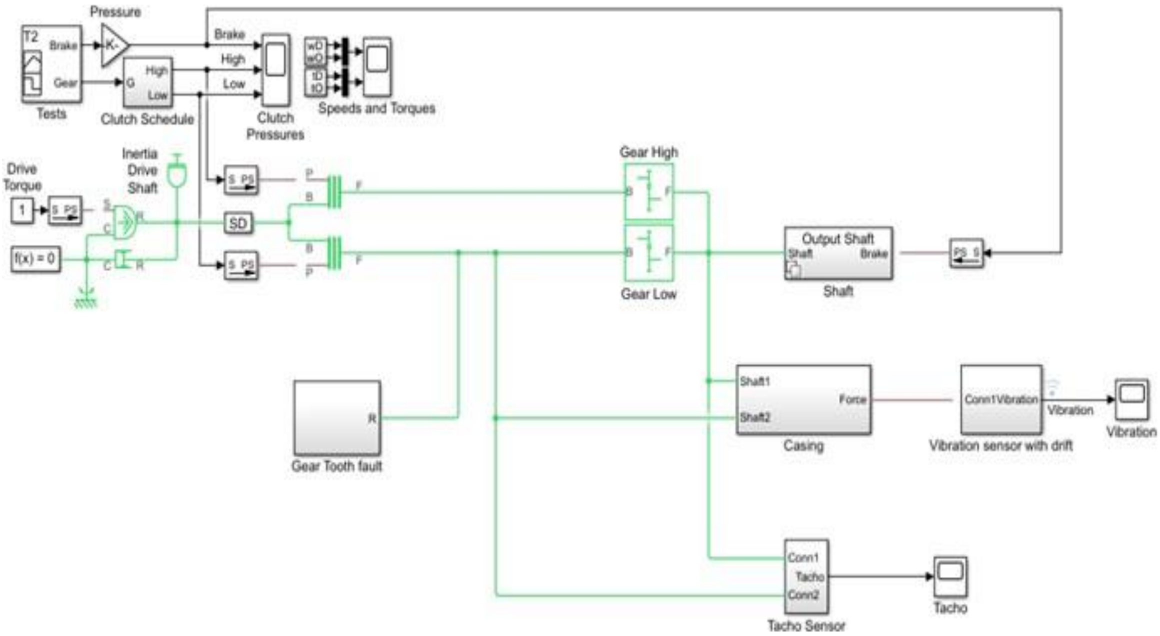


Fig. 3. The simulated DT-TS model.

3.3. Synthetic Minority Oversampling Technique (SMOTE)

In this study, imbalanced data is observed after dataset generation while one class is higher than the other. Therefore, SMOTE technique is utilized to keep balanced data for both classes. SMOTE is an oversampling technique which is dealing with the synthetic samples that are generated for the imbalanced data. It is useful for the overfitting issues caused by random oversampling. In addition, It focuses on the feature space to produce new balanced instances between healthy and faulty conditions [56]. In this study, SMOTE technique is applied to the corresponding dataset for achieving the balance between the two instances. Then, the model performance is evaluated before and after adding GA. The basic steps of SMOTE method are indicated in Fig. 4.

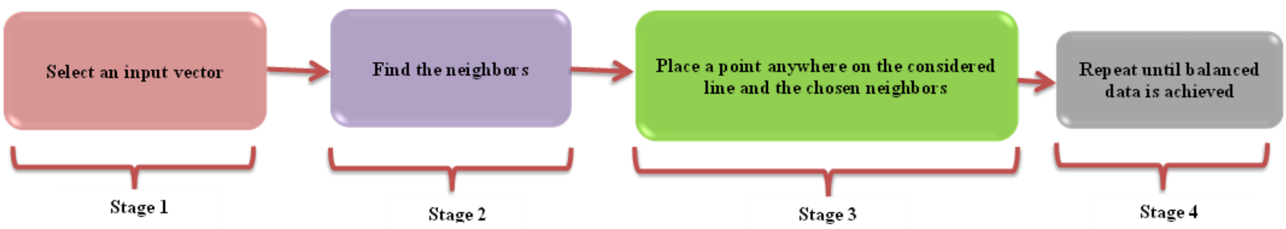


Fig. 4. The basic steps of SMOTE algorithm.

3.4. Machine Learning Classifiers (MLCs)

MLCs are utilized for learning machines to make its decisions according to specific dataset. This section explained the proposed classifiers and its parameters that are used in the hybrid GA-ML framework during the implementation process [57].

1) Logistic Regression (LR): is the process that based on the discrete outcome probability according to the input variable. LR strategy is useful in the problems that have two instances. it depends on several parameters such as solver, penalty, and C parameter which controls the penalty strength. The parameter values that are used in the proposed GA-ML model are solver = 'liblinear', penalty = '12', and C = 50.

2) Linear discriminant analysis (LDA): The main role of LDA is decreasing the dimensionality of the classification problem. Moreover, it is considered the main key of maximizing the separation between classes. Experimentally, LDA is the backbone of maximizing and minimizing two main parameters such as distance and variation, respectively by creating a new axis which received data with low dimensions.

3) Gaussian Naive Bayes (NB): This classifier is based on features probability evaluation that are normally distributed. NB appears its efficiency by assuming that data is distributed normally and build its assumptions according to this formula. It has the ability of predicting the probability of variables which is useful in several classification tasks.

4) K-nearest neighbors (KNN): is the most popular classifier used for classification and regression problems. the prediction process is based on the distances between instances in the testing side. This study depended on Euclidean distance in the proposed framework between the corresponding conditions. The key parameter in KNN is the number of k which indicates the number of neighbors that algorithm selected and classifying according to the majority of this selection. Consequently, in this study KNN parameters that are implemented are k =3, weights = 'uniform', and leaf size = 20.

3.5. Genetic Algorithm (GA)

It was inspired in 1975 by John Holland. It makes use of biological theory concepts like mutation, crossover, and selection to provide excellent results for optimization and search issues [58]. Chromosomes, which are made up of genes, are used to represent GA solutions. Population collection is the term used to describe chromosomes. These genes represent a preferred response to the issue that has to be resolved by GA. starting with the population, which was randomly started depending on the data. Now let's talk about the three fundamental processes that are employed to alter this population: reproduction, crossover, and mutation. In this study, there are several parameter values that are used such as Mutation rate = 0.01, Crossover rate = 0.7, Population Size = 100, Generation Size = 100, and the selection method is Roulette-Wheel method. Unfortunately, the harmony between parameters in the proposed framework that are produced from the combination of GA and each model in MLCs has its vital role in preventing several issues such as stagnation that can occur by using individual algorithms with its single stages. The following lists the GA stages in detail:

3.5.1. Evaluation

The fitness of the potential solutions are assessed while the initial population or the new population is produced. Each chromosome's fitness value is computed to determine the fitness solution. A possible solution to the issue that needs GA optimization is defined by a chromosome.

3.5.2. Selection

In this stage, the best chromosomes will be selected for creating a new population. The technique that is used is known as Roulette- Wheel selection [50]. In this method, chromosomes are selected according to their fitness concerning other chromosomes as follows:

A. Calculate the fitness for everyone, Ai.

B. Compute Fi, which is known as the probability of each sample in the population.

$$F_i = \frac{A_i}{\sum_{n=1}^{k} A_n} \tag{1}$$

where k is the size of the population.

C. Assess the probability, Xi for each sample:

$$X_i = \sum_{n=1}^{i} F_i \tag{2}$$

D. Create a positive random number, z ∈ [0, 1].

E. Check for z<X_1, then select the first chromosome, else select other the individual where X_(i-1)<z<X_i.

F. To create the k candidates for the mating pool, repeat two steps D, E k-times.

3.5.3. Crossover

Combining the parents results in a new set of chromosomes that are then added to the new generation. In this stage, two parents are chosen, and a random point between two genes is chosen. Consequently, chromosomes are split into two halves. The first offspring will be produced by combining the first and second parents. Similar rules apply to the first and second halves of the first and second parents, respectively. These descendants will produce the subsequent population.

3.5.4. Mutation

The goal of the mutation stage is to change a chromosome's gene values in the new generation. The population's variety increased in this stage. Diversity plays a crucial part in the search algorithm by looking for information in places where it hasn't been found before. The following generation will include the total population. The stages will be repeated as shown in Fig. 5 until the ideal answer is found.

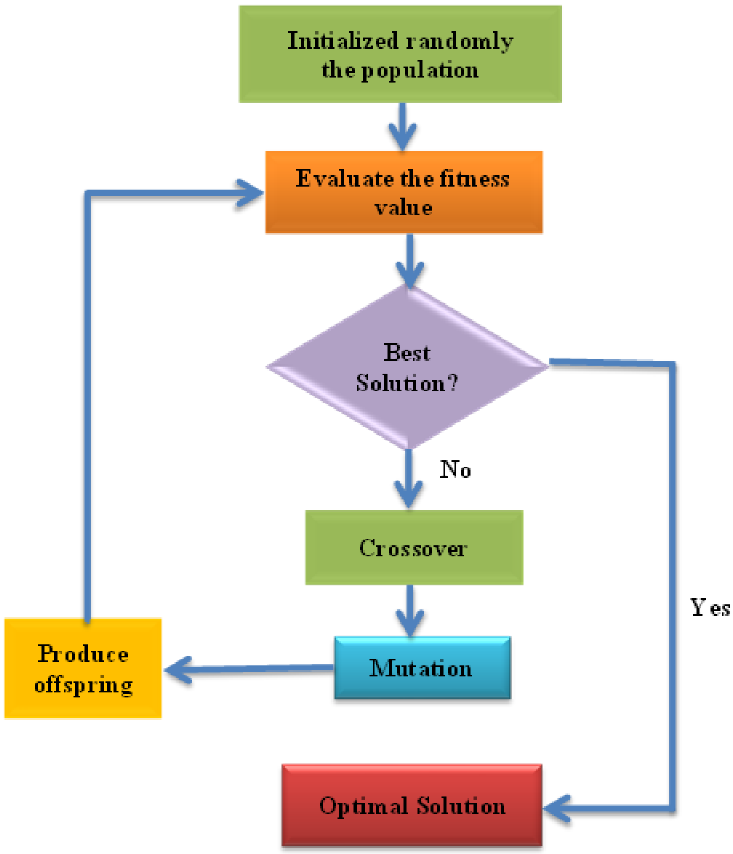


Fig. 5. The flowchart of GA.

4. Simulation Analysis

4.1. Dataset Generation

In this study, the dataset produced from the simulated model of the TS model as indicated in Fig. 3. It consists of 208 samples and 18 features with two cases which are healthy condition while no faults occurred and faulty condition while faults occurrence. The performance is tested by hybrid GA-ML models to improve the performance of diagnosing and classifying system failures.

4.2. Evaluation metrics

The hybrid framework uses two strategies. Firstly, we used SMOTE technique for imbalanced dataset. Then, we combined ML models with GA for a binary classification to improve the accuracy of dataset which is generated from the corresponding DT model. The comparison is performed the classical MLCs before and after applying the hybrid GA-ML model. As shown in Table 2, a comparison between real and predicted values is applied to evaluate the system performance.

Table 2: Comparison between real and predicted values.

		Real	
		Positive	**Negative**
Predicted	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Once the parameters are calculated, the performance metrics can be computed as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

4.3. Simulation Results

Four machine learning models have been applied for classifying and detecting healthy and faulty conditions. These classifiers are LR, LDA, NB, and KNN for fault detecting and diagnosing. The performance of the ordinary classifiers such as LR, LDA, NB, and KNN is evaluated as indicated Figs. 6 and 7 for healthy and faulty data, respectively. The results of all classifiers ensured the need of an enhanced framework for improving the system performance.

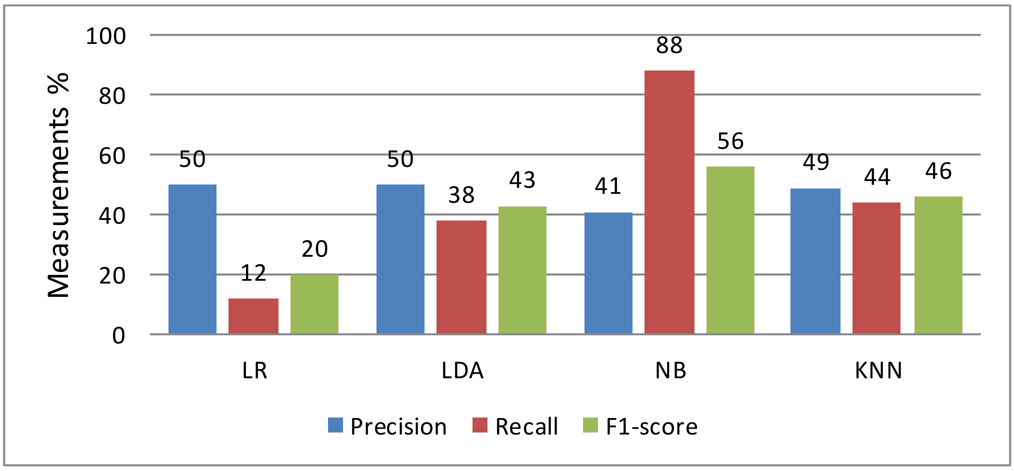


Fig. 6. The assessment of healthy data for the ordinary classifiers.

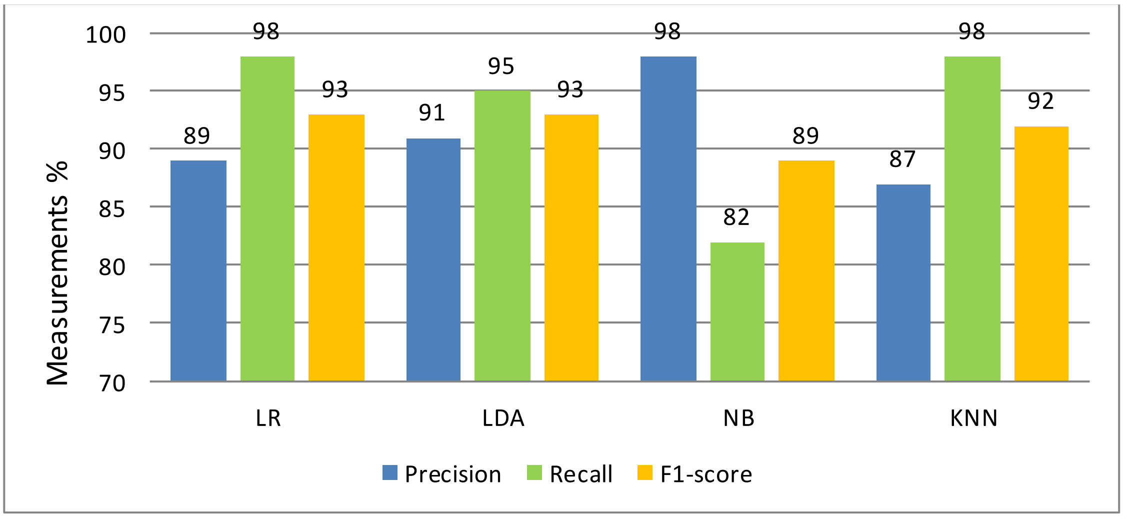
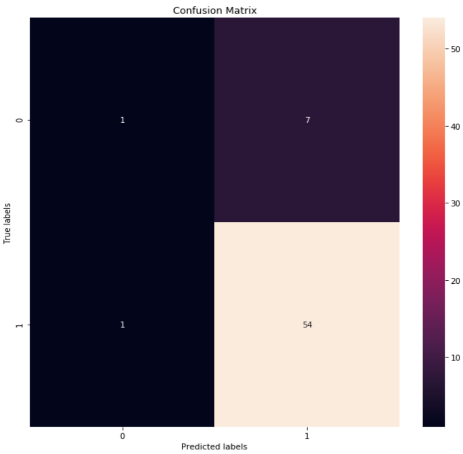
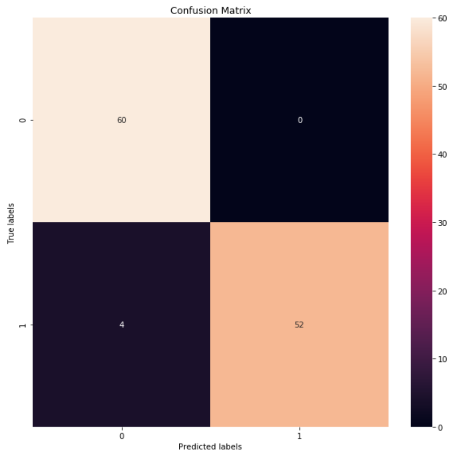


Fig. 7. The assessment of faulty data for the ordinary classifiers.

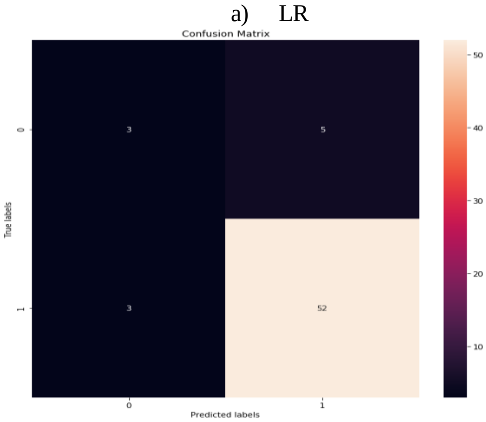
For all four ML models, the measurements between real values and their predicted counterpart are shown in Fig. 8. Furthermore, the ROC-Curve of the proposed GA-ML model is illustrated in Fig. 9. The measurements of the hybrid GA-ML model are indicated in Figs. 10 and 11. As observed, the combination between GA and ML models give better results than its classical counterpart. Deeply, GA-LDA model gives highly performance compared to the others. As observed from Fig. 12, GA-ML accuracy increases for all classifiers compared with the classical ones such as GA-LDA, GA-NB, and GA-KNN which show 98.3%, and 97.4%, respectively. Finally, Table 3 presents a comparison between the measurements of the classical MLCs and the hybrid GA-ML framework for fault diagnosis of a DT model as healthy or faulty status as shown in Fig.13.



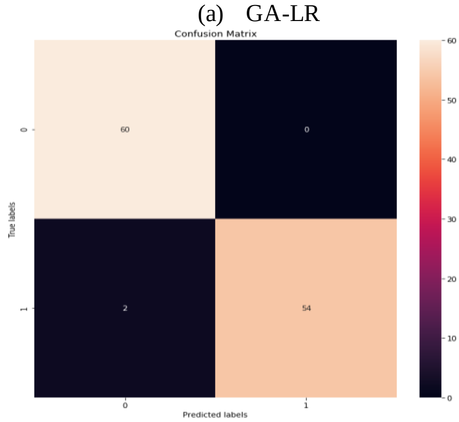
a) LR



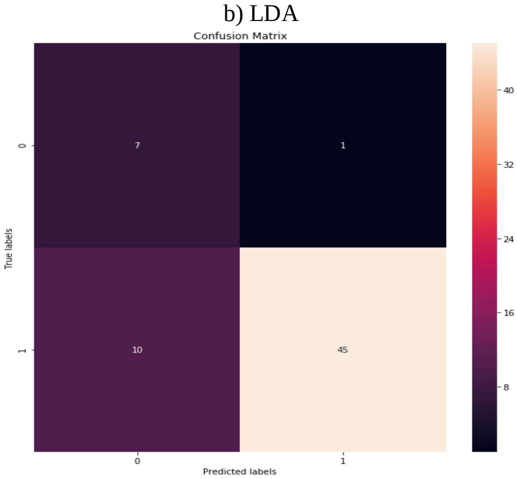
(a) GA-LR



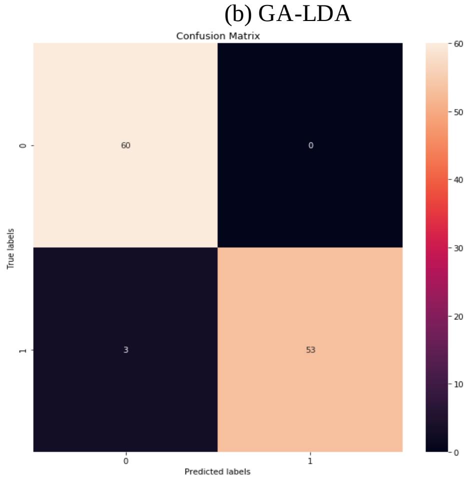
b) LDA



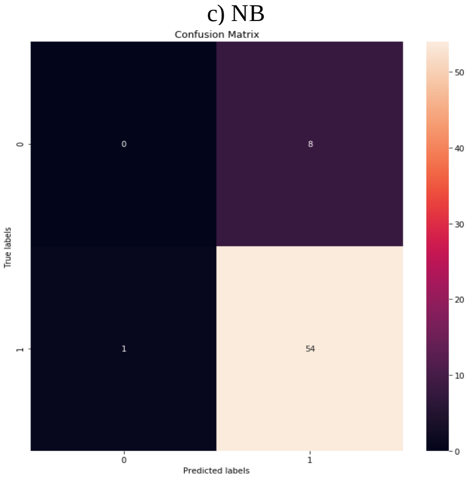
(b) GA-LDA



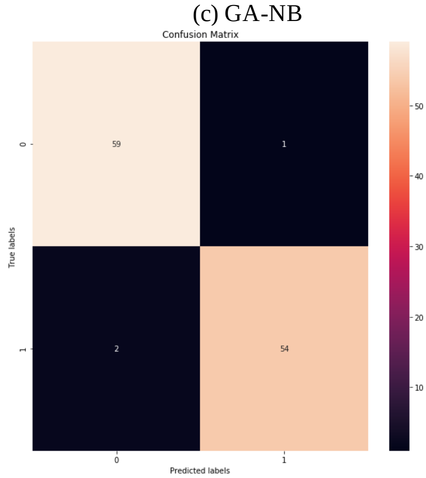
c) NB



(c) GA-NB

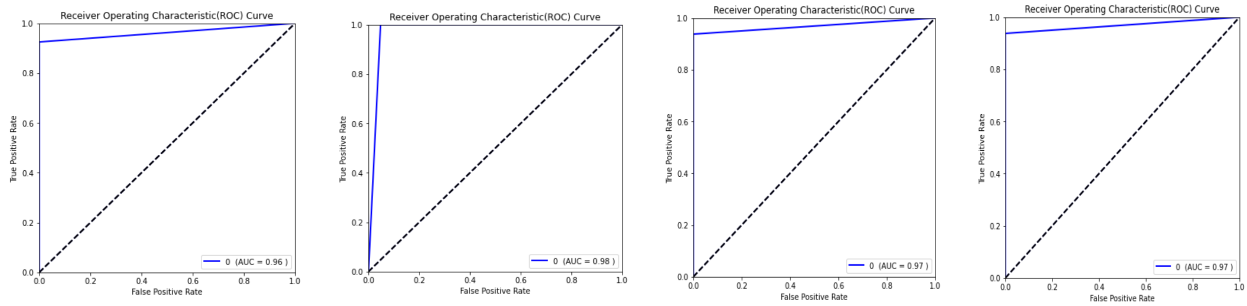


d) KNN



(d) GA-KNN

Fig. 8. Classical MLCs vs. Proposed GA-ML Framework.



(a) GA-LR (b) GA-LDA (c) GA-NB (d) GA-KNN

Fig. 9. ROC Curve of GA-ML model.

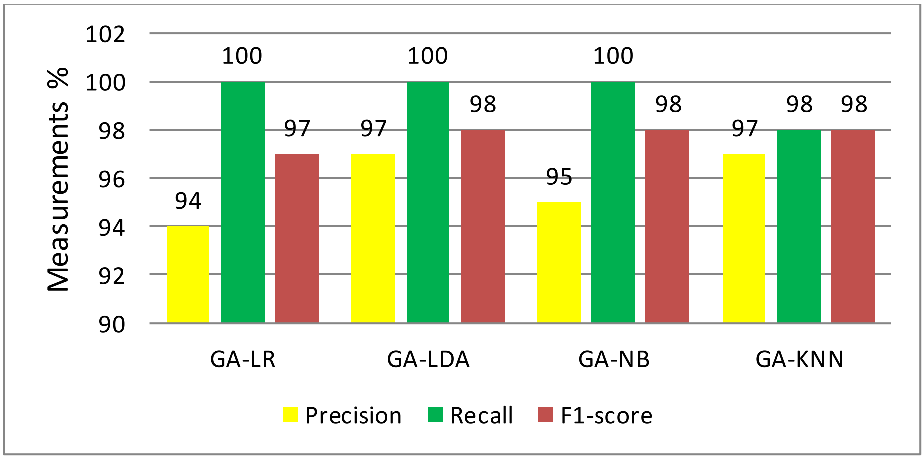


Fig. 10. Hybrid GA-ML performance of healthy data for the corresponding classifiers.

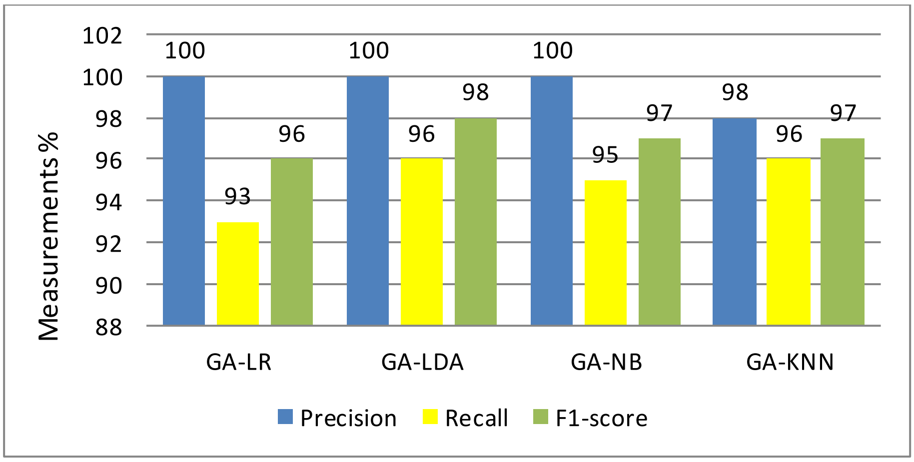


Fig. 11. Hybrid GA-ML performance of faulty data for the corresponding classifiers.

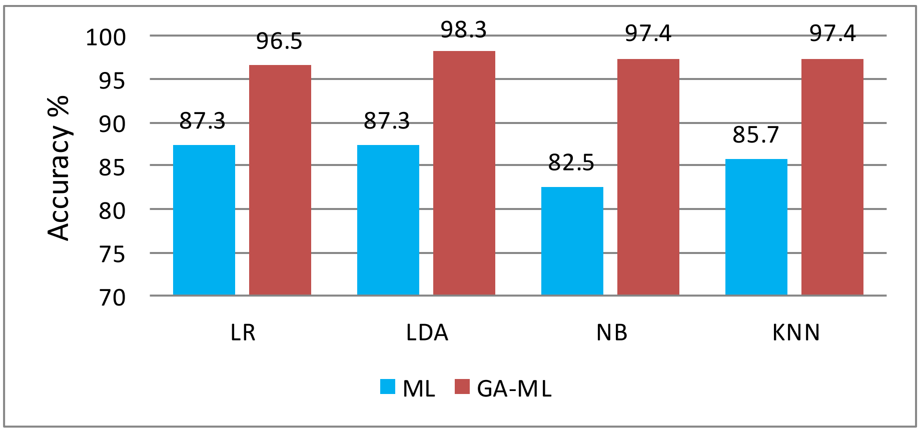


Fig. 12. The Comparison between hybrid GA-ML Framework and ML classifiers.

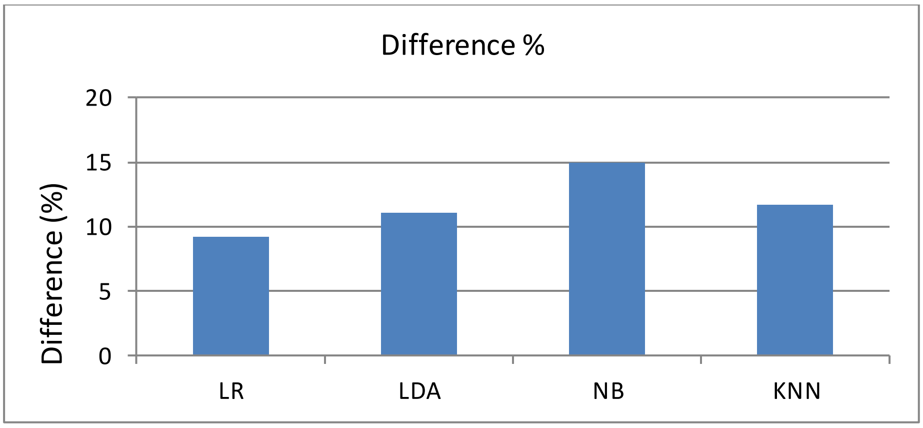


Fig. 13. The improvement percentage of GA-ML framework with respect to the ordinary classifiers.

Table 3: Comparison between real and predicted values.

Metrics	Data status	LR/GA-LR	LDA/GA-LDA	NB/GA-NB	KNN/GA-KNN
Precision	Healthy	0.50/0.94	0.50/1.00	0.41/0.97	0.49/0.97
	Faulty	0.89/1.00	0.91/0.93	0.98/0.96	0.87/0.98
Recall	Healthy	0.12/0.97	0.38/1.00	0.88/0.98	0.44/0.98
	Faulty	0.98/1.00	0.95/0.96	0.82/0.98	0.98/0.98
F1-score	Healthy	0.20/0.95	0.43/1.00	0.56/0.98	0.46/0.98
	Faulty	0.93/1.00	0.93/0.95	0.89/0.97	0.92/0.97

5. Discussion

The digital twins and ML approaches have been emerged to construct the suggested hybrid GA-ML framework for problem diagnosis in complex control systems. The hybrid framework effectively combines optimization technique with MLCs. The suggested framework intends to enable the industrial operator's prediction and diagnosis of failures. This may be achieved by combining digital twins and ML strategies together to observe industrial processes. The suggested plan focuses on several crucial phases for achieve the system goals for intelligent fault detection. Each stage aligns with the others by presenting a certain job and operation. The framework with the key phases is as the following:

1. A simulated data is generated from TS model.
2. Data is passed through preprocessing stages such as extracting features using PCA method.
3. Using SMOTE technique for unbalanced dataset.
4. Data is processed for prediction and classification inspections. This is done using the proposed hybrid GA-ML approach.
5. Model evaluation according to classification metrics for diagnosing and predicting the future events.
6. According to results, hybrid GA-LDA model gives the highest accuracies of 98.3 %. while the GA-NB, GA-LR, and GA-KNN give accuracies of 97.4%,96.5, and 97.4%, respectively which ensured the outperformance of the proposed framework.

6. Limitations and Future Work

In this study, the author only selected four MLCs which are LR, LDA, NB, and KNN. These classifiers can be increased according for providing effectiveness and investigation of the proposed framework. Moreover, this hybrid approach can be implemented on several digital twins applications for more comprehensiveness of the model. On the other side, the performance of GA-ML framework requires vital enhancement under study. This improvement will be the first step of the future procedure. Unfortunately, this study can be expanded by combining GA and other optimization technique with various ML models for multiple digital twins applications. Furthermore, emerging and evaluating additional ML classifiers with GA can led to better results. Consequently, future work aims to develop digital twins systems by adding several directions that can be outlines as follows:

- Utilizing of multiple CNNs for fault classification with different digital twins models.
- Development of digital twins fault detection techniques by utilizing different IoT-based approaches.
- Use different smart Industrial-IoT techniques for digital twins diagnosis and prediction.
- Using the help of Blockchain in building secure and accurate digital twins-based IoT models for real-time detection of criminals and attackers especially for Critical Industrial Control Systems (ICS) such as electrical and power networks.
- Build a cloud-based digital twins processing environment for handling enormous data using big data platforms such as Apache Spark.

- Explore Blockchain, big data analytics, and digital twins relationships and how big data analytics can easily and quickly handle of massive generated data from digital twins-based IoT applications in a secure manner.

By the way, future directions can significantly raise the effectiveness of the hybrid GA-ML framework in several applications and real-world industries.

7. CONCLUSION

This work aims to improve the effectiveness of system performance, by integrating two essential approach such as optimization methods by GA and ML classifiers such as LR, LDA, NB, and KNN. The dataset originates from a simulated system which is TS model. The hybrid GA-ML model is then applied to the corresponding classifiers with satisfied results according to the ordinary ones. According to results, the classical LR and LDA classifiers give an accuracy of 87.3 % while the NB and KNN give an accuracy of 82.5%, and 85.7 % respectively. Moreover, after applying the optimization algorithm, hybrid GA-LDA model gives the highest accuracy as 98.3% while both GA-NB and GA-KNN give an accuracy of 97.4%. In addition to GA-LR which gives 96.5%. The hybrid framework is verified by comparing it with classical ML models. The performance of the LDA based on the GA are better than the classical MLCs. Finally, the framework effectiveness is verified through comparisons with the traditional techniques. The GA-ML framework presented higher efficiency and accuracy compared to standard ML models. Ultimately, the hybrid framework not only aids for detecting future events but also its vitality can be utilized in several digital twins-IoT applications.

REFERENCES

[1] He, Bin, Bai, Kai Jian, 2021. Digital twin-based sustainable intelligent manufacturing: a review. Adv. Manuf. 9 (1) https://doi.org/10.1007/s40436-020-00302-5. ISSN 21953597.

[2] Liu, Ying, Zhang, Lin, Yuan, Yang, Zhou, Longfei, Ren, Lei, Wang, Fei, Liu, Rong, Pang, Zhibo, Jamal Deen, M., 2019. A novel cloud- based framework for elderly healthcare Services using a digital twin. 10.1109/ACCESS.2019.2909828. ISSN 21693536.

[3] Caputo, F., Greco, A., Fera, M., Macchiaroli, R., 2019. Digital twins to enhance the integration of ergonomics in workplace design. Int. J. Ind. Ergon. 71, 20–31. https://doi.org/10.1016/j.ergon.2019.02.001. ISSN 18728219.

[4] Gil, Yolanda, et al 2019. Intelligent systems for geosciences: an essential research agenda. Commun. ACM 62 (1), 76–84. https://doi.org/10.1145/3192335. ISSN 15577317.

[5] Guo, K., Wan, X., Liu, L., Gao, Z., & Yang, M. (2021). Fault diagnosis of intelligent production line based on digital twin and improved random forest. Applied Sciences (Switzerland), 11(16). https://doi.org/10.3390/app11167733

[6] Psarommatis, F., & May, G. (2023). A literature review and design methodology for digital twins in the era of zero defect manufacturing. International Journal of Production Research, 61(16), 5723–5743. https://doi.org/10.1080/00207543.2022.2101960

[7] O'Connell, E., O'Brien, W., Bhattacharya, M., Moore, D., & Penica, M. (2023). Digital Twins: Enabling Interoperability in Smart Manufacturing Networks. Telecom, 4(2), 265–278. https://doi.org/10.3390/telecom4020016.

[8] Feng, X., & Wan, J. (2024). Digital Twins for Discrete Manufacturing Lines: A Review. Big Data and Cognitive Computing, 8(5), 45. https://doi.org/10.3390/bdcc8050045.

[9] Sayed, S., Nassef, M., Badr, A., & Farag, I. (2019). A Nested Genetic Algorithm for feature selection in high-dimensional cancer Microarray datasets. Expert Systems with Applications, 121, 233–243. https://doi.org/10.1016/j.eswa.2018.12.022.

[10] Bauer, Peter, Stevens, Bjorn, Hazeleger, Wilco, 2021. A digital twin of Earth for the green transition. Nat. Clim. Change 11 (2), 80–83. https://doi.org/10.1038/s41558-021- 00986-y. ISSN 1758–6798.

[11] Pourpanah, F., Shi, Y., Lim, C. P., Hao, Q., & Tan, C. J. (2019). Feature selection based on brain storm optimization for data classification.Applied Soft Computing Journal, 80, 761–775. https://doi.org/10.1016/j.asoc.2019.04.037.

[12] Van, M., Hoang, D. T., & Kang, H. J. (2020). Bearing fault diagnosis using a particle swarm optimization-least squares wavelet support vector machine classifier. Sensors (Switzerland), 20(12), 1–19. https://doi.org/10.3390/s20123422.

[13] Toma, R. N., Prosvirin, A. E., & Kim, J. M. (2020). Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers. Sensors (Switzerland), 20(7). https://doi.org/10.3390/s20071884

[14] Sahebi, G., Movahedi, P., Ebrahimi, M., Pahikkala, T., Plosila, J., & Tenhunen, H. (2020). GeFeS: A generalized wrapper feature selection approach for optimizing classification performance. Computers in Biology and Medicine, 125(August), 103974. https://doi.org/10.1016/j.compbiomed.2020.103974.

[15] Islam, M.M.; Kim, J.M. Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector Machines. Reliab. Eng. Syst. Saf. 2019, 184, 55–66.

[16] Han, S.; Choi, H.J.; Choi, S.K.; Oh, J.S. Fault diagnosis of planetary gear carrier packs: A class imbalance and multiclass classification problem. Int. J. Precis. Eng. 2019, 20, 167–179.

[17] Bennacer, L.; Amirat, Y.; Chibani, A.; Mellouk, A.; Ciavaglia, L. Self-diagnosis technique for virtual private network combining Bayesian networks and case-based reasoning. IEEE Trans. Autom. Sci. Eng. 2015, 12, 354–366.

[18] Palacios, R.H.C.; Goedtel, A.; Godoy,W.F.; Fabri, J.A. Fault identification in the stator winding of induction motors

using PCA with artificial neural networks. J. Control Autom. Electr. Syst. 2016, 27, 406–418.
[19] Chen, Y.Q.; Fink, O.; Sansavini, G. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction. IEEE Trans. Ind. Electron. 2018, 65, 561–569.
[20] Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. IEEE Trans. Syst. Man Cybern. 1991, 21, 660–674.
[21] Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement 2016, 93, 490–502.
[22] Sun W, Zhao R, Yan R, Shao S, Chen X. Convolutional Discriminative Feature Learning for Induction Motor Fault Diagnosis. IEEE Trans Ind Informatics 2017;13 (3):1350–9. https://doi.org/10.1109/TII.2017.2672988.
[23] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mech Syst Signal Process 2020; 138:106587. https://doi.org/10.1016/j.ymssp.2019.106587.
[24] Glowacz A. Fault diagnosis of single-phase induction motor based on acoustic signals. Mech Syst Signal Process 2019;117:65–80. https://doi.org/10.1016/j. ymssp.2018.07.044.
[25] Shao S, McAleer S, Yan R, Baldi P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. IEEE Trans Ind Informatics 2019;15(4):2446–55. https:// doi.org/10.1109/TII.2018.2864759.
[26] Shen C, Qi Y, Wang J, Cai G, Zhu Z. An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder. Eng Appl Artif Intell 2018;76:170–84. https://doi.org/10.1016/j. engappai.2018.09.010.
[27] Kolar, D., Lisjak, D., Pajak?, M., & Gudlin, M. (2021). Intelligent fault diagnosis of rotary machinery by convolutional neural network with automatic hyper-parameters tuning using bayesian optimization. Sensors, 21(7). https://doi.org/10.3390/s21072411.
[28] Cavalieri, S.; Salafia, M.G. A Model for Predictive Maintenance Based on Asset Administration Shell. Sensors 2020, 20, 6028.
[29] Haghshenas, A.; Hasan, A.; Osen, O.; Mikalsen, E.T. Predictive digital twin for ofshore wind farms. Energy Inform. 2023, 6, 1.
[30] Costantini, A.; Di Modica, G.; Ahouangonou, J.C.; Duma, D.C.; Martelli, B.; Galletti, M.; Antonacci, M.; Nehls, D.; Bellavista, P.; Delamarre, C.; et al. IoTwins: Toward Implementation of Distributed Digital Twins in Industry 4.0 Settings. Computers 2022, 11,67.
[31] Abdullahi, I.; Perinpanayagam, S.; Hamidu, I. A Fog Computing Based Approach towards Improving Asset Management and Performance of Wind Turbine Plants Using Digital Twins. In Proceedings of the 2022 27th International Conference on Automation and Computing (ICAC), Bristol, UK, 1–3 September 2022.
[32] J. Lee, M. Azamfar, J. Singh and S. Siahpour, "Integration of digital twin and deep learning in cyberphysical systems: Towards smart manufacturing," IET Collaborative Intelligent Manufacturing, vol. 2, no.1, pp. 34–36, 2020.
[33] W. Booyse, D. Wilke and S. Heyns, "Deep digital twins for detection, diagnostics and prognostics,"Mechanical Systems and Signal Processing, vol. 14, no. 10, pp. 106–121, 2020.
[34] E. Hinchy, N. Dowd and C. McCarthy, "Using open-source microcontrollers to enable digital twin communication for smart manufacturing," Procedia Manufacturing, vol. 3, no. 8, pp. 1213–1219, 2019.
[35] J. Cheng, H. Zhang, F. Tao and C. Juang, "DT-II: Digital twin enhanced industrial internet reference framework towards smart manufacturing," Robotics and Computer-Integrated Manufacturing, vol. 6, no.2, pp. 101–108, 2020.
[36] J.Wang, L. Ye, R. Gao, X. Li and L. Zhang, "Digital twin for rotating machinery fault diagnosis in smart manufacturing," International Journal of Production Research, vol. 57, no. 12, pp. 3920–3934, 2019.
[37] Hodavand, F., Ramaji, I. J., & Sadeghi, N. (2023). Digital Twin for Fault Detection and Diagnosis of Building Operations: A Systematic Review. Buildings, 13(6). https://doi.org/10.3390/buildings13061426.
[38] Zhang, Y., Ji, J. C., Ren, Z., Ni, Q., Gu, F., Feng, K., Yu, K., Ge, J., Lei, Z., & Liu, Z. (2023). Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing. Reliability Engineering and System Safety, 234, 1–27. https://doi.org/10.1016/j.ress.2023.109186.
[39] Hasan, A., Asfihani, T., Osen, O., & Bye, R. T. (2024). Leveraging digital twins for fault diagnosis in autonomous ships. Ocean Engineering, 292(November 2023), 116546.https://doi.org/10.1016/j.oceaneng.2023.116546.
[40] Idrissi Kaitouni, S., Ait Abdelmoula, I., Es-sakali, N., Mghazli, M. O., Er-retby, H., Zoubir, Z., El Mansouri, F., Ahachad, M., & Brigui, J. (2024). Implementing a Digital Twin-based fault detection and diagnosis approach for optimal operation and maintenance of urban distributed solar photovoltaics. Renewable Energy Focus , 48(November 2023), 100530. https://doi.org/10.1016/j.ref.2023.100530
[41] T. Zhang, J. Chen, F. Li, K. Zhang and E. Xu, "Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions," ISA Transactions, vol. 11, no. 9, pp.152–171, 2022.
[42] Y. Cheng, M. Lin, J.Wu, H. Zhu and X. Shao, "Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network," Knowledge-Based Systems, vol.21, no. 6, pp. 132–145, 2021.
[43] Z. Gao, C. Cecati and S. Ding, "A survey of fault diagnosis and fault-tolerant techniques-part I: Fault diagnosis with model-based and signal-based approaches," IEEE Transactions on Industrial Electronics,vol. 62, no. 6, pp. 3757–3767, 2015.

[44] M. Zhong, T. Xue and S. Ding, “A survey on model-based fault diagnosis for linear discrete time-varying systems,” Neurocomputing, vol. 30, no. 6, pp. 51–60, 2018.
[45] C.Wang and Z. Xu, “An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis,” Neurocomputing, vol. 45, no. 6, pp. 550–562, 2021.
[46] Samanta, S. Chowdhuri and S. Williamson, “Machine learning-based data-driven fault detection/diagnosis of lithium-ion battery: A critical review,” Electronics, vol. 10, no. 11, pp. 130–139, 2021.
[47] Y. Xing, B.Wang, Z. Gong, Z.Hou andK. Jiao, “Data-driven fault diagnosis for pem fuel cell system using sensor pre-selectionmethod and artificial neural network model,”IEEE Transactions on Energy Conversion, vol. 3, no. 5, pp. 1–19, 2022.
[48] S. Taqvi, H. Zabiri, F. Uddin,M. Naqvi and A.Maulud, “Simultaneous fault diagnosis based on multiple kernel support vector machine in nonlinear dynamic distillation column,” Energy Science & Engineering, vol. 5, no. 3, pp. 1–17, 2022.
[49] X. Wang, B. Yang, Z. Wang and X. Guan, “A compressed sensing and CNN-based method for fault diagnosis of photovoltaic inverters in edge computing scenarios,” IET Renewable Power Generation, vol.2, no. 7, pp. 1–12, 2022.
[50] Dhingra, L. (2023). Bearing Fault Diagnosis With Envelope Analysis and Machine Learning Approaches Using CWRU Dataset. 2023 International Conference on Data Science and Network Security, ICDSNS 2023, 11(June). https://doi.org/10.1109/ICDSNS58469.2023.10244972.
[51] Li, J., King, S., & Jennions, I. (2023). Intelligent Fault Diagnosis of an Aircraft Fuel System Using Machine Learning—A Literature Review. Machines, 11(4). https://doi.org/10.3390/machines11040481.
[52] Afia, A., Gougam, F., Rahmoune, C., Touzout, W., Ouelmokhtar, H., & Benazzouz, D. (2024). Intelligent fault classification of air compressors using Harris hawks optimization and machine learning algorithms. Transactions of the Institute of Measurement and Control, 46(2), 359–378. https://doi.org/10.1177/01423312231174939.
[53] Li, Y., Wang, Y., Zhao, X. & Chen, Z. A deep reinforcement learning-based intelligent fault diagnosis framework for rolling bearings under imbalanced datasets. Control. Eng. Pract. 145, 105845, DOI: https://doi.org/10.1016/j.conengprac.2024.105845 (2024).
[54] https://www.mathworks.com/help/predmaint/ug/Use-Simulink-to-Generate-Fault-Data.html.
[55] Howley, T., Madden, M. G., O'Connell, M. L., & Ryder, A. G. (2006). The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data. Knowledge-Based Systems, 19(5), 363–370. https://doi.org/10.1016/j.knosys.2005.11.014.
[56] Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Journal of Artificial Intelligence Research, 61, 863–905. https://doi.org/10.1613/jair.1.11192.
[57] Randhawa K, Loo C, Seera M, Lim C, Nandi A (2018) Credit card fraud detection using AdaBoost and majority voting. IEEE Access 6:14277–14284.
[58] Toma, R. N., Prosvirin, A. E., & Kim, J. M. (2020). Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers. Sensors (Switzerland), 20(7). https://doi.org/10.3390/s20071884.