Heuristic Approach for Cognitive Digital Twin Technology – A Technical Report

Atiq ur Rehman, Mobyen Uddin Ahmed, Shahina Begum

Artificial Intelligence and Intelligent Systems Research Group,
School of Innovation, Design and Engineering, Mälardalen University
{atiq.ur.rehman,mobyen.uddin.ahmed,shahina.begum}@mdu.se

Abstract. Essential for the next generation of production are intricate systems that integrate virtualization and simulation platforms with real-time data from industrial processes. Among these systems, digital twins stand out as they offer numerous advantages, particularly in the realm of manufacturing where they can enhance productivity across the entire production life cycle. By leveraging cognitive digital twins, enterprises gain the ability to extract valuable implicit insights from ongoing production operations in a creative, efficient, and effective manner. Over time, the advancement of various technologies has greatly enhanced the capabilities and sophistication of the digital twin concept. In this study, we propose a heuristic approach for advancing cognitive digital twin technology, representing the next stride in digital twin development crucial for realizing the objectives of Industry 4.0. To infuse cognitive functionalities, we advocate for the adoption of a heuristic approach in the creation of cognitive digital twins. Specifically, we introduce heuristic optimization as a feature selection tool, aimed at augmenting the cognitive capabilities of a digital twin throughout the product design phase of production. The efficacy of this proposed approach is demonstrated through a practical application in Power Transfer Unit (PTU) production. This validation resulted in a noteworthy 8.83% enhancement in classification accuracy for identifying faulty PTUs on the assembly line. This translates to a considerable improvement in throughput for the PTU assembly line, while also conserving resources that would have otherwise been expended on faulty units.

Acknowledgments. This work was supported in part by the project DIGICOGS project which is financed by Vinnova (Vinnovas Diarienr: 2019-0532) and the innovation program Process Industrial IT and Automation (PiiA) at Mälardalen University. The authors would like to thank Michael Osbakk, Mikael Eriksson, Jonathan Widén, Jimmy Vesa, and 'GKN Drive line' for all the help and support during this study.

1. Introduction

This project delves into the transformative potential of Industry 4.0, often referred to as the fourth industrial revolution [1]. Industry 4.0 encompasses a spectrum of cutting-edge technologies, including cloud computing, cybersecurity, IoT, advanced robotics, and machine learning, all converging to revolutionize manufacturing processes [2]. Among the key technologies driving this wave of innovation are Digital Twins (DTs), a fusion of digital simulations and real-world production systems. DTs facilitate seamless communication between the digital and physical realms, offering a paradigm shift in manufacturing operations [3], [4].

Building upon the foundation of DTs, this project takes a significant step forward by introducing Heuristic Cognitive Digital Twins (HCDTs). By integrating cognitive processes such as perception, reasoning, and problem-solving, CDTs elevate the capabilities of DTs [5]. They empower real-time tracking, anomaly detection, root cause analysis, and informed decision-making for manufacturing components. Further, these CDT are integrated with the heuristic optimization and this integration of cognitive functionalities with optimization marks a pivotal advancement in the realm of digital twin technology.

The project's primary focus is on the Power Transfer Unit (PTU) manufacturing industry. The PTU, a critical component in automobiles, plays a pivotal role in distributing power to all four wheels. However, the production of faulty PTUs can result in substantial economic losses. To address this challenge, a heuristic optimization approach is introduced within the digital twin framework. Specifically, the project leverages the Jumping Particle Swarm Optimization (PSO) method [6], [7] to refine the fault detection process in PTU manufacturing. The Jumping PSO is an extension of the classical PSO algorithm [8], [9] which enhances its ability to skip the local optimum solution with the jumping search strategy. The results of this optimization effort are striking, showcasing a remarkable enhancement in the accuracy of fault detection.

In summation, this project stands at the forefront of digital twin technology, demonstrating the profound impact of integrating cognitive capabilities. Focused on the PTU manufacturing industry, this endeavor not only refines fault detection processes but also serves as a testament to the potential of cognitive digital twins in enhancing manufacturing operations. Through this innovative approach, the project propels us further into the era of Industry 4.0, promising more efficient, accurate, and resilient manufacturing practices.

2. Proposed Approach

The comprehensive concept of the proposed Cognitive Digital Twin (CDT) for Power Transfer Unit (PTU) manufacturing is illustrated in Figure 1. Data sourced from both the lapping and assembly lines serve as the foundation for predicting faulty PTUs in future scenarios. The abundance of sensor data from these lines encompasses various manufacturing parameters, each defined by distinct sensor values. While human experts affirm that most of these parameters bear no relevance to the prediction of faulty PTUs, a subset is deemed influential in this regard.

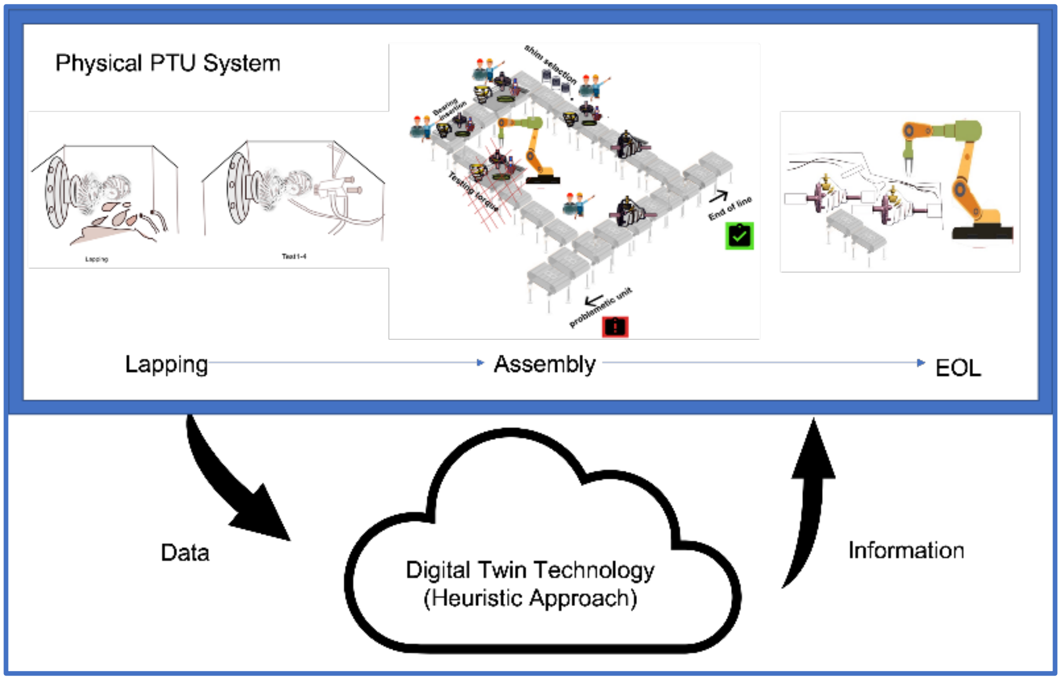


Fig. 1. Overall concept for the proposed CDT for PTU manufacturing.

To systematically analyze the full spectrum of manufacturing parameters, the Jumping Particle Swarm Optimization (PSO) method is integrated with machine learning algorithms. This amalgamation facilitates the optimization of the feature space, ultimately leading to the selection of the most effective combination of parameters for accurate faulty PTU predictions. The PSO process encompasses three fundamental optimization steps:

1. Assessing particles based on fitness function.
2. Maintaining a record of the best position and associated fitness values.
3. Updating particle velocity and position.

This optimization process operates iteratively, and over successive iterations, the swarm of particles gradually converges. Once convergence is achieved, either by reaching a maximum number of successive iterations or a specific fitness value, the global best position of the algorithm identifies the optimal combination of features (manufacturing parameters) for prediction.

This iterative process is visually explained in Figure 2, where binary digits within a row signify a particle's position. '0' denotes unselected features, while '1' indicates selected features. The collective arrangement of particles (rows) forms a swarm, with each particle subjected to performance evaluation against various machine learning algorithms. The personal best position (p(t)) of a particle corresponds to the combination of features (represented by '1's) at which the machine learning algorithm attained the highest classification accuracy at the particle level. The global best position (g(t)) encompasses the combination of features across the entire swarm, achieving the highest classification accuracy. The rest of the details are provided in [6], [7], [10]

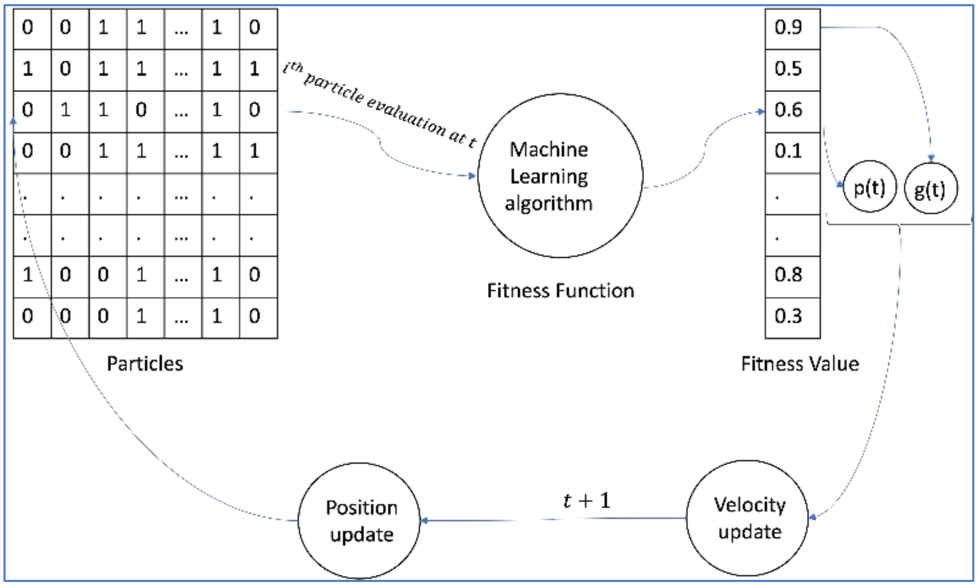


Fig. 2. Proposed process of selecting best manufacturing parameters for the prediction of faulty PTUs.

3. Validation and Results

To assess the efficacy of the proposed heuristic digital twin architecture, a range of standard machine learning algorithms are evaluated within a wrapper framework. The evaluations are conducted under two distinct scenarios: (i) employing important variables identified by human experts, and (ii) utilizing an optimized feature space as per the proposed architecture. All results are obtained through 5-fold cross-validation.

Table 1 presents the performance of various machine learning algorithms using the parameters deemed significant by human experts. Notably, the results indicate a rather limited performance in predicting faulty PTUs when employing all the features designated by human experts. The highest performance recorded is 66.17% using the K-Nearest Neighbors (KNN) model with Euclidean distance and 3 nearest neighbors. These results in Table 1 represent the most optimized outcomes following hyperparameter tuning across all machine learning models.

In contrast, Table 2 showcases the results achieved through the proposed heuristic digital twin model. It is evident from these findings that the proposed framework exhibits notably enhanced performance. Achieving the highest accuracy of 75.00% with an optimized subset of the feature space, it demonstrates a significant improvement compared to the accuracy obtained using all features without optimization. This underscores the crucial importance of optimizing the expansive feature space and discerning the vital manufacturing parameters for preemptively identifying faulty PTUs.

Table 1. Performance of different algorithms with all the parameters.

Model	Accuracy (%)
SVM (RBF)	57.35
Random Forest	55.88
KNN	66.17
Nåive Bayes	50.00
Discriminant Analysis	53.63
Generalized additive model	58.08
Decision Tree	55.14

Table 2. Performance of different algorithms with optimized feature space.

Model	#Features	Features	Accuracy (%)
SVM (RBF)	12	3,4,5,18,26,28,29,30,33,39,43,46	75.00
Random Forest	10	4,7,11,16,20,22,27,45,46,47	72.79
KNN (k=1, cosine)	11	3,4,9,11,18,29,34,38,39,41,44	75.00
Nåive Bayes	6	4,5,9,21,27,45	66.18
Discriminant Analysis	8	4,11,15,18,26,28,36,37,38	68.38
Generalized additive model	12	5,9,11,16,19,20,30,33,40,41,42,43	69.85
Decision Tree	12	3,5,8,11,13,14,16,19,21,29,42,47	73.53

The proposed framework brings forth two substantial advantages: (i) the capacity to elevate prediction accuracy, and (ii) the proficiency to pinpoint pivotal prediction parameters. Table 2 further presents different subsets of features, representing the most effective combinations for prediction using the specified model.

To assess the impact of swarm size, the experiment replicated with varying swarm sizes of Jumping PSO. Table 3 details the results for two of the best-performing classifiers. Noticeably, differing results for the same model highlight the influence of swarm size.

Table 3. Influence of Jumping PSO's swarm size.

Model	swarm size	#Features	Features	Accuracy(%)
SVM	150	12	3,4,5,18,26,28,29,30,33,39,43,46	75.00
SVM	100	12	4,5,6,12,15,18,21,23,39,43,44,45	71.32
SVM	50	8	5,10,18,22,29,30,39,47	69.85
SVM	150	9	3,5,11,18,23,29,30,41,47	72.05
SVM	100	8	10,11,14,22,23,31,34,39	71.32
KNN	150	11	3,4,9,11,18,29,34,38,39,41,44	75.00
KNN	100	14	5,10,11,14,18,19,21,27,29,30,31,32,39,41	69.85
KNN	50	2	35,47	66.17
KNN	150	13	3,8,13,18,19,22,27,36,38,39,41,43,44	71.32
KNN	100	13	2,3,4,5,7,18,24,29,30,35,36,39,44	68.38

4. User Interface Design

The User Interface (UI) designed to test the data from industry in quite simple, yet it contains all the options to test the new data. The options available for simulation include data balancing, feature selection using PSO, option to select the classifier, re-optimizing the feature space using PSO, and display options for data and results.

The UI starts with a 'RUN' button, and depending on the selected settings it displays the results. The users can define the path/name of the data to load in the system. Once the data path/name is provided, the simulation starts with the selected options. The options provided in the user interface are shown in Fig. 3. Some of the results of designed UI are also provided here in Fig 4. In this figure, the results depict the running condition with selected options of the UI designed.

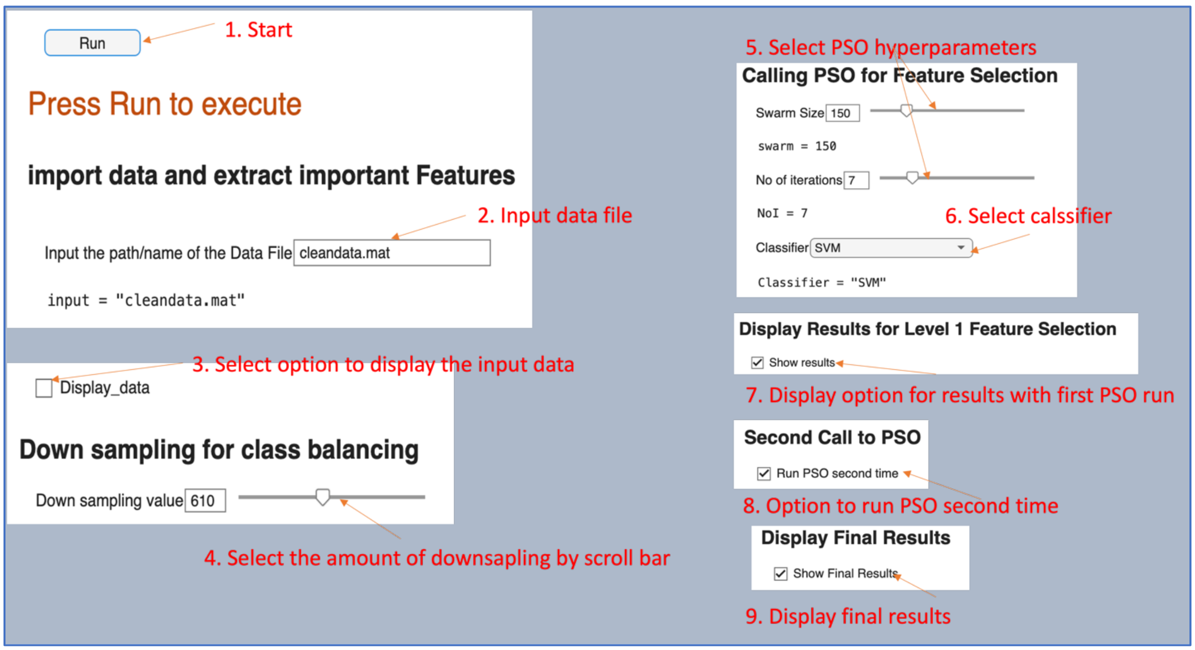
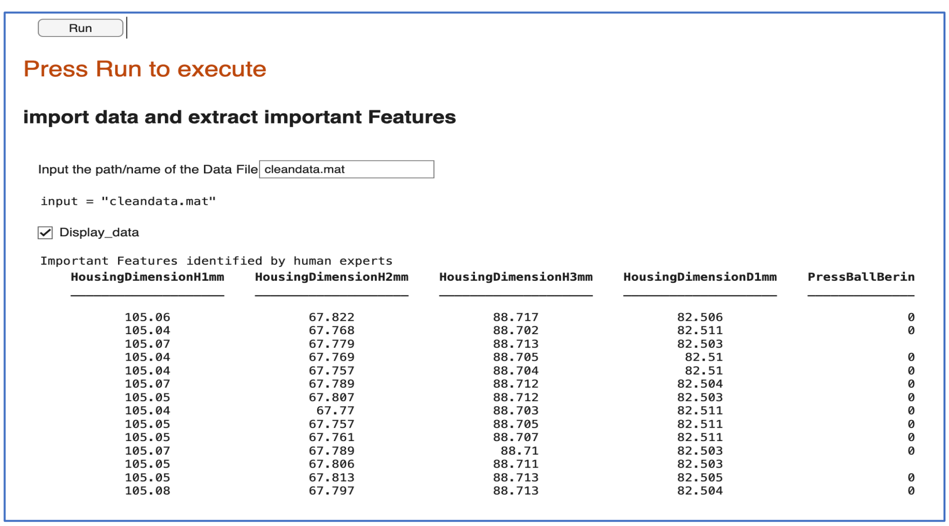


Fig 3: User Interface Options.



HousingDimensionH1mm	HousingDimensionH2mm	HousingDimensionH3mm	HousingDimensionD1mm	PressBallBerin
105.06	67.822	88.717	82.506	0
105.04	67.768	88.702	82.511	0
105.07	67.779	88.713	82.503	
105.04	67.769	88.705	82.51	0
105.04	67.757	88.704	82.51	0
105.07	67.789	88.712	82.504	0
105.05	67.807	88.712	82.503	0
105.04	67.77	88.703	82.511	0
105.05	67.757	88.705	82.511	0
105.05	67.761	88.707	82.511	0
105.07	67.789	88.71	82.503	0
105.05	67.806	88.711	82.503	
105.05	67.813	88.713	82.505	0
105.08	67.797	88.713	82.504	0

Fig 4: User Interface to import data.

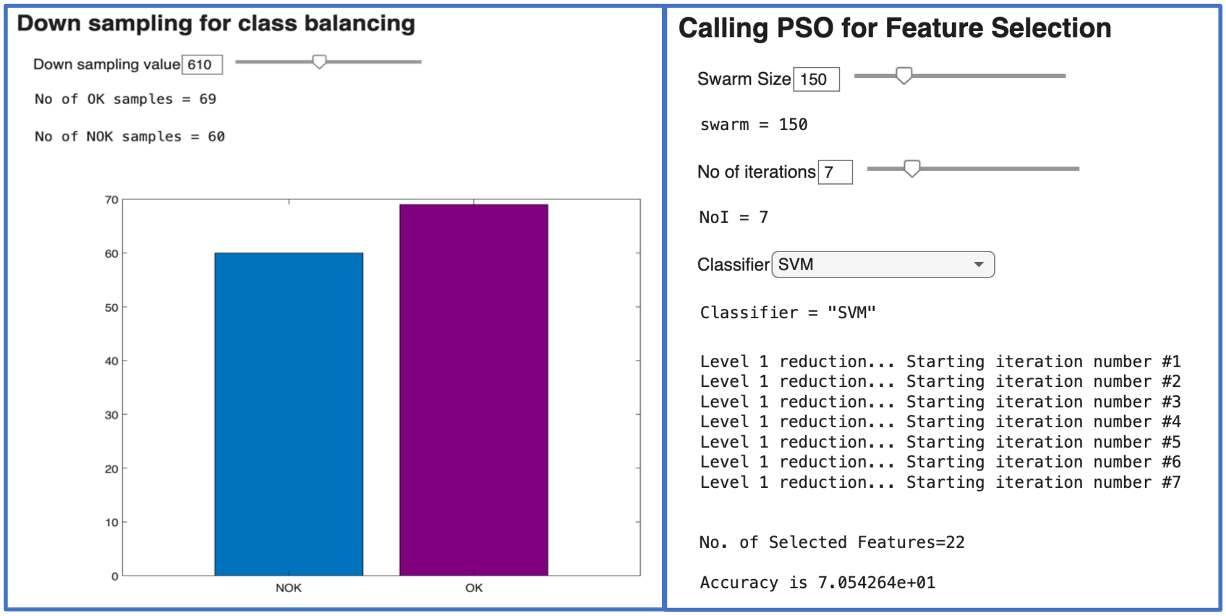


Fig 5: User Interface for class balancing (left) and feature selection (right).

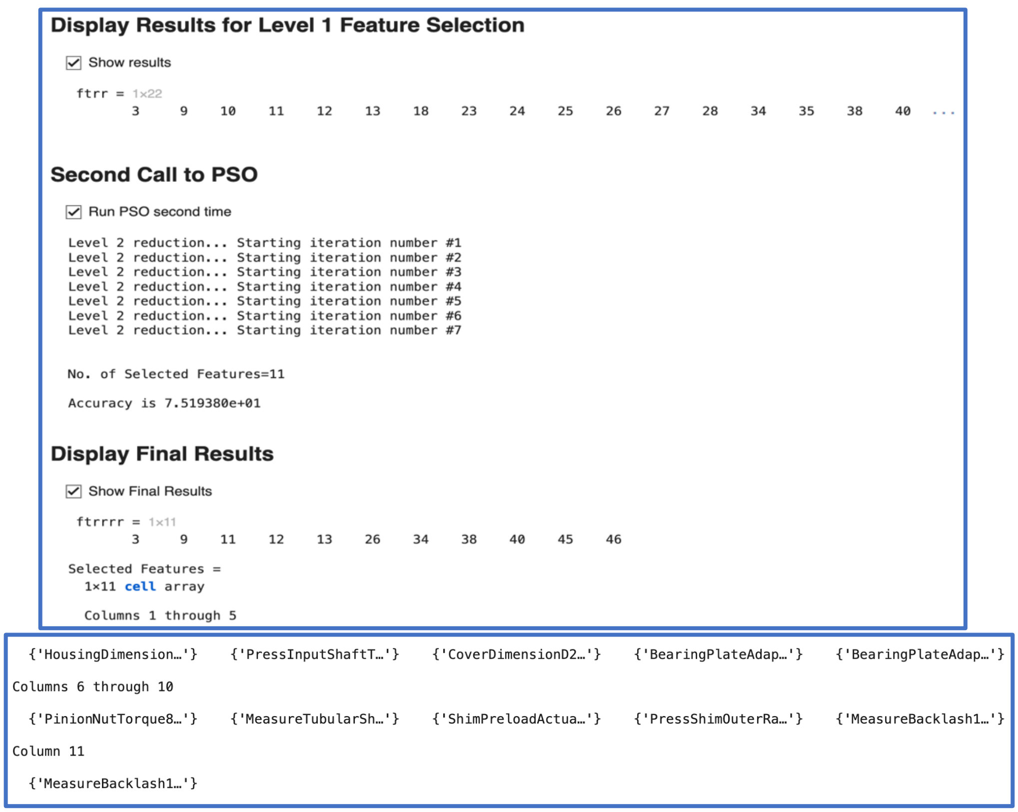


Fig. 6: UI simulation results with selected options.

5. Discussion

The outcomes are subjected to a comprehensive analysis employing established evaluation metrics, including Sensitivity, Specificity, Precision, False Positive Rate, F1 score, Matthews Correlation Coefficient, Kappa, and Receiver Operating Characteristic (ROC) curve. The findings across these evaluation metrics are presented in Table 4 and depicted in Fig. 5. Table 4 notably highlights a relatively low sensitivity in the models, indicating instances where faulty PTU samples are misclassified as standard PTU samples. Additionally, the machine learning models exhibit a maximum accuracy of 75% in classifying faulty PTUs.

These results may be attributed to the inherent complexity and challenging nature of the data, which possesses limited discriminatory power. Moreover, the scarcity of available data samples for faulty PTUs in real-world scenarios may constrain the adequacy of training samples for effective model training. To ensure optimal performance, it is imperative to provide an ample supply of training samples for faulty PTUs. Given that the current study is based on a limited dataset, enhanced model performance could be achieved with an expanded pool of training samples.

The chosen subset of optimized features is further scrutinized to ascertain the frequency of selection for each feature across various experiments. Features that exhibit a higher frequency of selection hold greater significance, as they are consistently prioritized in different experiments. Notably, *'Cover Dimension D2 mm'* emerges as the most frequently selected feature, followed by *'Press Ball Bering To Housing Force kN'* and *'Press Shim Head Outer Race To Housing Force kN.'* Several other parameters also exhibit a notable frequency of selection, underscoring their importance in constructing a Cognitive Digital Twin (CDT) model for the PTU manufacturing process.

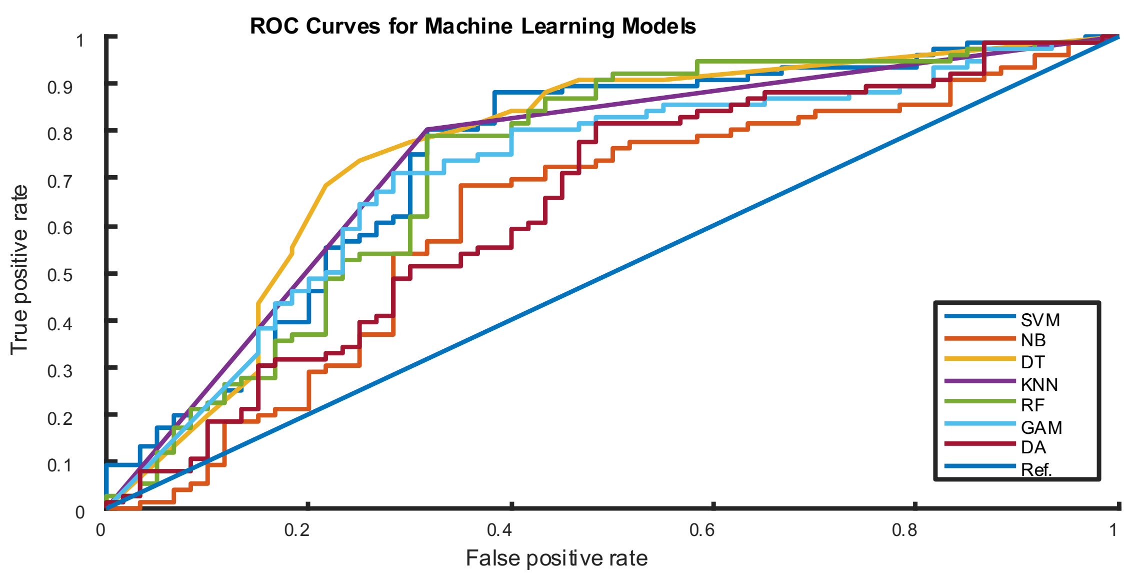


Fig.7. ROC curves for Machine Learning Models.

Table 4. Performance evaluation on different metrics.

Evaluation	SVM	RF	KNN	NB	DA	GAM	DT
Accuracy	0.75	0.72	0.75	0.66	0.68	0.69	0.73
Error	0.25	0.27	0.25	0.33	0.31	0.30	0.26
Sensitivity	0.58	0.65	0.68	0.63	0.51	0.66	0.60
Specificity	0.88	0.78	0.80	0.68	0.81	0.72	0.84
Precision	0.79	0.70	0.73	0.61	0.68	0.65	0.75
False Positive Rate	0.11	0.21	0.19	0.31	0.18	0.27	0.15
F1 score	0.67	0.67	0.70	0.62	0.59	0.66	0.66
Matthews Correlation Coef.	0.49	0.44	0.49	0.31	0.35	0.38	0.45
Kappa	0.47	0.44	0.48	0.31	0.34	0.38	0.45

6. Conclusion

The significance of incorporating a heuristic approach in crafting a Cognitive Digital Twin (CDT) for manufacturing processes is assessed and endorsed in this project. A practical case study involving Power Transfer Unit (PTU) manufacturing serves as the testing ground for the proposed heuristic CDT model. The outcomes clearly underscore the crucial role of heuristic optimization in refining the feature space for CDTs. This optimization not only amplifies the efficacy of machine learning models but also furnishes valuable insights into critical parameters.

In addition to augmenting decision-making capabilities and control autonomy, the proposed model exhibits the potential to elevate enterprise performance on a substantial scale. Hence, it is strongly advised to employ such frameworks in the development of CDT models within the manufacturing industry. Future endeavors will focus on augmenting the pool of available faulty PTU samples, thereby further refining the classification accuracy of the system.

Furthermore, a user-friendly interface has been meticulously designed in MATLAB, tailored to cater to industrial personnel. This interface is poised for future testing on real-world data, positioning it as an asset for practical implementation in industrial settings.

8. References

[1] H. Rahman, R. S. D'Cruze, M. U. Ahmed, R. Sohlberg, T. Sakao, and P. Funk, "Artificial Intelligence-Based Life Cycle Engineering in Industrial Production: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 133001–133015, 2022, doi: 10.1109/ACCESS.2022.3230637.

[2] S. Teerasoponpong and P. Sugunnasil, "Review on Artificial Intelligence Applications in Manufacturing Industrial Supply Chain – Industry 4.0's Perspective," in *2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, IEEE, Jan. 2022, pp. 406–411. doi: 10.1109/ECTIDAMTNCON53731.2022.9720417.

[3] L. Li, B. Lei, and C. Mao, "Digital twin in smart manufacturing," *J Ind Inf Integr*, vol. 26, p. 100289, Mar. 2022, doi: 10.1016/j.jii.2021.100289.

[4] J. Friederich, D. P. Francis, S. Lazarova-Molnar, and N. Mohamed, "A framework for data-driven digital twins of smart manufacturing systems," *Comput Ind*, vol. 136, p. 103586, Apr. 2022, doi: 10.1016/j.compind.2021.103586.

[5] S. S. Sheuly, M. U. Ahmed, and S. Begum, "Machine-Learning-Based Digital Twin in Manufacturing: A Bibliometric Analysis and Evolutionary Overview," *Applied Sciences*, vol. 12, no. 13, p. 6512, Jun. 2022, doi: 10.3390/app12136512.

[6] A. Ur Rehman, A. Islam, N. Azizi, and S. B. Belhaouari, "Jumping Particle Swarm Optimization," 2022, pp. 743–753. doi: 10.1007/978-981-16-2380-6_65.

[7] A. U. Rehman, A. Islam, and S. B. Belhaouari, "Multi-cluster jumping particle swarm optimization for fast convergence," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3031003.

[8] A. U. Rehman and A. Bermak, "Swarm Intelligence and Similarity Measures for Memory Efficient Electronic Nose System," *IEEE Sens J*, vol. 18, no. 6, 2018, doi: 10.1109/JSEN.2018.2799611.

[9] A. U. Rehman and A. Bermak, "Recursive DBPSO for computationally efficient electronic nose system," *IEEE Sens J*, vol. 18, no. 1, 2018, doi: 10.1109/JSEN.2017.2771388.

[10] A. ur Rehman, M. U. Ahmed, and S. Begum, "Cognitive Digital Twin in Manufacturing: A Heuristic Optimization Approach," 2023, pp. 441–453. doi: 10.1007/978-3-031-34107-6_35.