Villeneuve d’Ascq, August, 23, 2023

*Paper title: “***Prior Knowledge-Infused Self-Supervised Learning and Explainable AI for Fault Detection and Isolation in PEM Electrolyzers”**

By: Balyogi Mohan Dash, Belkacem Ould Bouamama, Komi Midzodzi Pekpe and Mahdi Boukerdja

To Editor in Chief.

Neurocomputing

**Object**: First Revised version

Dear Editor in Chief,

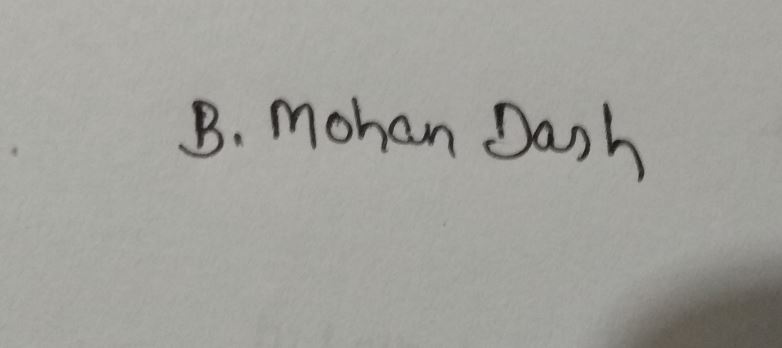
Please find enclosed the sources files of the **first revised version** of the paper NEUCOM-D-23-05968 entitled “Prior Knowledge-Infused Self-Supervised Learning and Explainable AI for Fault Detection and Isolation in PEM Electrolyzers”

Are included:

1. Revision notes
2. Manuscript Latex source file of the paper with **Bibliography file** and **Figures**
3. Revised Manuscript in PDF
4. Cover letter
5. Biographies and pictures of all authors

All comments have been taken into account.

Thank you for revising the paper. Please receive my best regards



Balyogi Mohan Dash

**Response to Reviewers’ Comments**

Thank you for considering our revised draft of the manuscript titled "**Prior Knowledge-Infused Self-Supervised Learning and Explainable AI for Fault Detection and Isolation in PEM Electrolyzers**" for publication in the Journal of Neurocomputing. We are grateful for the time and effort you and the reviewers dedicated to providing feedback on our work. The insightful comments and valuable improvements suggested by the reviewers have been immensely helpful. We have diligently incorporated all of their suggestions, and you will find these changes in ***red*** within the “**Manuscript with marked correction”**. Please refer to the point-by-point response to the reviewers' comments and concerns provided below. All page numbers mentioned correspond to the “**Manuscript with marked correction”**.

**Comments of The Editor**

***Comment 1***. The authors must improve the linguistic quality of their manuscript; upon screening the latest revision, there are still a few language issues. This is highly important to ensure a proper understanding of your manuscript.

**Answer:** Thank you for your valuable feedback and careful examination of our manuscript. We appreciate your constructive comments regarding the linguistic quality. We have diligently reviewed and revised the manuscript to address the language issues to the best of our ability.

We understand the significance of clear and effective communication in scientific writing, and we are committed to ensuring that our manuscript meets the required standards.

***Comment 2***. This document will benefit significantly if the authors share some demo code in a public repository or on the web to help readers adopt the proposed method.

**Answer:** Thank you for your insightful suggestion. Following your advice, we have established a public repository on GitHub, which includes the demo code corresponding to the proposed method presented in our manuscript. In addition, we have thoughtfully integrated the repository link into the abstract for convenient access and reproducibility. We believe that this addition will greatly enhance the utility of our work and facilitate a more thorough understanding for readers interested in adopting the proposed method. <https://github.com/mohan696matlab/SSL_based_Hybrid_FDI>

**Where the revisions are made:**

page-1, paragraph-Abstract

***Comment 3***. Please ensure that there is a section (CreDiT) outlining the specific contributions made by each author.

**Answer:** Thank you for your suggestion. We have duly incorporated a CreDiT section in the revised manuscript, outlining the specific contributions made by each author.

**Where the revisions are made:**

page-30, paragraph-3

**Comments of Reviewer 2**

This paper proposes Prior Knowledge-Infused Self-Supervised Learning and Explainable AI for Fault Detection and Isolation in PEM Electrolyzers. In general, this paper is well presented. The following issues can be further considered.

***Comment 1***. More background and motivation of this study can be added, in case the readers are not very familiar with the topic.

**Answer:** We sincerely appreciate the reviewer for providing valuable insights and constructive comments. In response to this suggestion, we have incorporated additional background details that elucidate the concept of green hydrogen and its diverse applications. Recognizing the diverse background of our readership, we have also expanded upon the various types of fault diagnosis for those who may be new to this domain.

Furthermore, to underscore the motivation behind our study, we have introduced a dedicated paragraph in the introduction section. This addition aims to clearly articulate the driving factors and significance that motivate our research.

**Where the revisions are made:**

page-1, paragraph-1 & 2

page-3, paragraph-4

***Comment 2***. The descriptions of the well known knowledge can be properly reduced.

**Answer:** We have taken steps to streamline the descriptions of well-known knowledge in our manuscript. Specifically, we have reduced the content related to existing Fault Detection and Isolation methods for PEM electrolyzers, as well as sections discussing transfer learning. Additionally, we have made adjustments to the eXplainable Artificial Intelligence (XAI) portion to ensure conciseness while maintaining clarity.

**Where the revisions are made:**

page-1, paragraph-3

page-2, paragraph-2,8

***Comment 3***. Why introducing the interpretable AI and self supervised learning method for the problem? What is the major benefits compared with traditional methods?

**Answer:** The introduction of interpretable AI and self-supervised learning methods in this research addresses key challenges and offers several benefits compared to traditional methods in fault detection and isolation for PEM electrolyzers.

Interpretable AI (XAI):

Enhanced Trustworthiness: Traditional deep-learning models often function as "black boxes," making it challenging to trust their decisions, especially in safety-critical applications. By incorporating XAI techniques such as the BG-XAI method, the decisions of the deep-learning model become interpretable, providing insights into why a fault is detected. This enhances trust in the model's predictions, a critical factor in applications where reliability is paramount.

Operator Understanding: XAI methods, particularly local explanation techniques like BG-XAI, offer specific insights into the reasons behind fault detection. This aids operators in understanding the nuances of the fault and facilitates more informed decision-making in response to detected faults.

Self-Supervised Learning (SSL):

Utilization of Unlabeled Data: Traditional supervised learning heavily relies on labeled data, which can be labor-intensive and time-consuming to obtain, particularly for fault detection tasks in complex systems. SSL allows the model to learn from vast amounts of unlabeled data, enhancing its ability to generalize and adapt to diverse operating conditions.

Reduced Dependency on Labeled Data: The scarcity of labeled data is a common limitation in fault diagnosis. SSL reduces the dependency on labeled data by leveraging the system's prior knowledge during the pre-training phase. This makes the method more data-efficient and applicable in scenarios where obtaining labeled data is challenging.

Integration of System Knowledge: SSL, when combined with prior knowledge about the system, allows for meaningful representations to be learned from unlabeled data. This integration of system knowledge enhances fault isolation capabilities using minimal labeled data, a significant advantage over traditional methods.

**Where the revisions are made:**

page-2, paragraph-5

***Comment 4***. Some related works on this topic should be reviewed, such as "Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions", "Intelligent Machinery Fault Diagnosis With Event-Based Camera", etc.

**Answer:** Certainly, we appreciate the reviewer's suggestion to include relevant studies in our manuscript. Recognizing the significance of the suggested works, namely “Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions” and “Intelligent Machinery Fault Diagnosis With Event-Based Camera” we have now incorporated a review of these studies into our manuscript.

We believe that this addition enhances the context of our work and provides readers with a more comprehensive understanding of the existing literature in the field.

**Where the revisions are made:**

page-2, paragraph-1,2

***Comment 5***. A couple of ablation studies should be added to evaluate the effects of the key parameters of the proposed method on the performance.

**Answer:** In the initial version of the manuscript, we conducted four ablation studies (Section 7.2) to highlight the impact of key parameters on the performance of our proposed method. These studies encompassed the effects of utilizing self-supervised and pure supervised learning, the influence of raw sensor measurements versus residual signals as input features, the impact of window length variation, and the effects of employing various deep learning models as feature extractors.

Understanding the significance of ablation, we have expanded our evaluation in the revised version. Specifically, we have included two additional experiments: one to demonstrate the effect of hierarchical combination on overall performance and another to showcase the influence of the amount of unlabeled data available during pretraining on the overall performance.

We believe these supplementary ablation studies contribute further insights into the robustness and adaptability of our proposed method.

**Where the revisions are made:**

page- 27

**Comments of Reviewer 3**

In this paper, a hybrid fault detection and isolation method is proposed. By using advanced sensor devices, the residual input replaces the traditional data input and is trained through a neural network to extract features, using automatic supervised learning is used to complete the division of data labels, and other operations related to deep learning are further improved. However, there are some issues in the process of intelligent fault detection that are not well understood. The comments are given as follows:

***Comment 1***. First of all, since the function $k\_\phi$ can be ignored after the construction is completed and fine-tuned by directly labeled real data (according to the performance test evaluation in Figure 7, it may also be adjusted during the test process), then why use a large number of Pseudo-labeled data to train the feature extraction function instead of using real data directly? Is this a common form of processing for SSL?

**Answer:** Thank you for your thoughtful observation regarding the function and the use of pseudo-labeled data in our Self-Supervised Learning approach. The function indeed plays a crucial role in the pre-training phase, where it acts as a classifier added on top of the feature extractor . This pre-training step is essential to leverage the abundant unlabeled data and generate pseudo-labels for a pre-text task.

Although the function is discarded post pre-training, its role in utilizing pseudo-labeled data is pivotal for knowledge transfer to the subsequent fine-tuning stage. Even after the removal of , the feature extractor retains the acquired knowledge from the pre-training phase. During fine-tuning, when labeled data is limited, becomes instrumental in transferring learned knowledge to the target task (fault classification) through the incorporation of a new fully connected layer .

The use of a large number of pseudo-labeled data during pre-training is a common strategy in SSL. It allows us to exploit the wealth of unlabeled data available, addressing the challenge of limited labeled data. Furthermore, the comprehensive pre-training facilitates a more effective fine-tuning process, contributing to the overall performance of the proposed method.

**Where the revisions are made:**

page-5, paragraph-3

***Comment 2***. Secondly, in Figure 6, is the threshold $h\_{i,t}$ assigned separately for each matrix element in the sliding window through LFT-BG? How is its production quality (reliability) measured?

**Answer:** Thank you for your query regarding Figure 6. In our approach, the threshold is indeed assigned separately for each matrix element in the sliding window through LFT-BG. The production quality, or reliability, of this threshold is assessed through a dynamic process based on sensor measurements and system inputs.

As described in Equation 4, the residual signals and adaptive thresholds are obtained by feeding sensor measurements and other inputs to the system at time into the LFT-BG model. This dynamic calculation ensures that the thresholds are tailored to the specific conditions of each time step within the sliding window.

The reliability of the thresholds is implicitly addressed through the adaptability of the LFT-BG model, which incorporates uncertainties associated with system parameters during its development. This adaptability is crucial in accounting for potential changes in system dynamics or estimation errors. Here are some examples who used LFT-BG to generate reliable and robust thresholds for fault diagnosis:Djeziri, M.A., Ould Bouamama, B., Dauphin-Tanguy, G. and Merzouki, R., 2011. LFT bond graph model-based robust fault detection and isolation. Bond graph modelling of engineering systems: theory, applications and software support, pp.105-133.

Djeziri, M.A., Merzouki, R., Ould Bouamama, B. and Dauphin-Tanguy, G., 2007. Robust fault diagnosis by using bond graph approach. IEEE/ASME Transactions on Mechatronics, 12(6), pp.599-611.

**Comments of Reviewer 4**

In this paper, a novel fault detection and isolation (FDI) method for Proton Exchange Membrane (PEM) electrolyzers is presented. The paper is well written. The results is interesting. Here are some specific suggestions:

***Comment 1***. Compared with existing results, the main contribution of the paper is ambiguous at the present stage. It seems the method is just combined several existing methods.

**Answer:** We appreciate the reviewer's feedback and acknowledge the importance of clarifying the main contribution of our paper. The ambiguity in the presentation might have obscured the novelty of our approach. While it is true that we leverage existing techniques such as LFT-BG and self-supervised learning, our innovation lies in the seamless integration of these methods to address the specific challenges posed by FDI in complex systems.

Our key contributions can be summarized as follows:

* Integration of LFT-BG for Pseudo-Label Generation: We propose a method to automatically generate pseudo-labels using a bond graph-based model of the system. This is a novel application of LFT-BG for fault detection, leveraging system dynamics to create informative labels for self-supervised learning.
* Self-Supervised Learning with Residual Signals: We introduce the use of residual signals obtained from the LFT-BG as input features for self-supervised learning. This choice is motivated by the sensitivity of residuals to faults, enhancing the performance of deep learning models in fault detection.
* Hierarchical Combination for Online FDI: We present a hierarchical combination strategy, integrating both the LFT-BG and the fine-tuned deep learning model, to minimize false alarms and enhance the detection of novel faults in real-time.
* Explanation of Fault Class Predictions with BG-XAI: Our work extends beyond fault detection to include an explanation mechanism (BG-XAI) that provides interpretable insights into the decision-making process of the deep learning model, enhancing the transparency and trustworthiness of the overall system.

We believe that these contributions collectively advance the state-of-the-art in FDI by offering a holistic and innovative approach. We will revise the manuscript to highlight these points more explicitly and improve the clarity of our main contributions.

**Where the revisions are made:**

page-3, paragraph-7

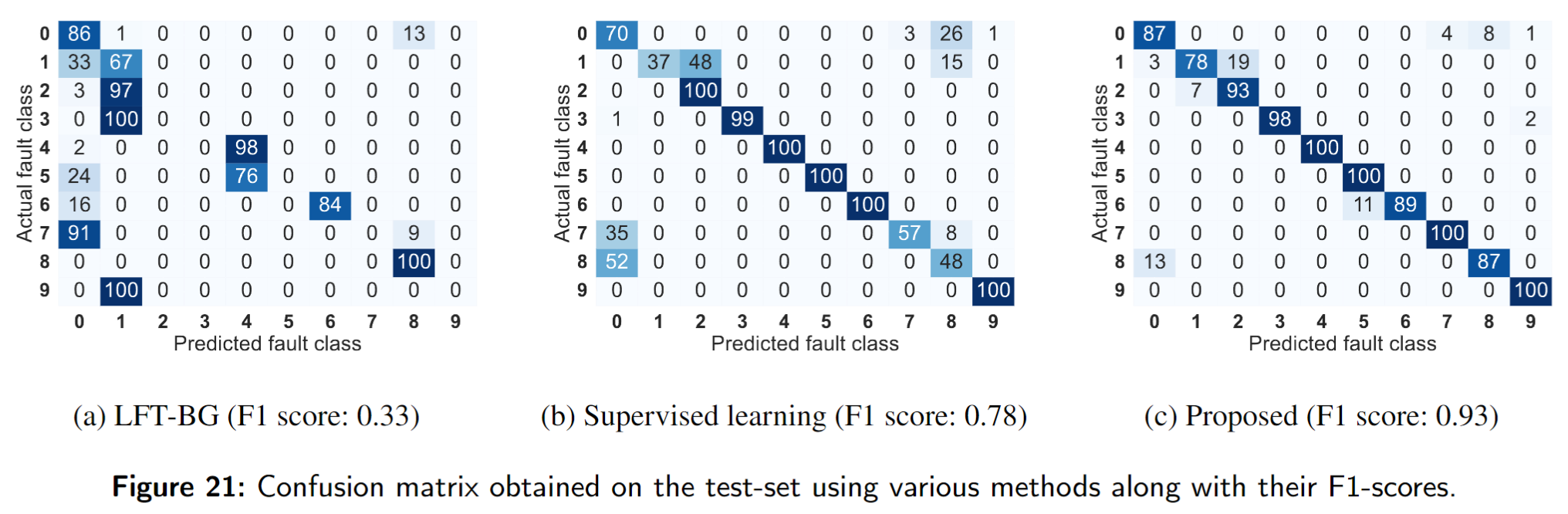
***Comment 2***. More illustrations and comparisons should be given to illustrate the advantages of the method in the paper.

**Answer:** We appreciate your suggestion regarding the need for more illustrations and comparisons in our manuscript. In response to your comment, we have included a comprehensive set of illustrations that directly compare the advantages of our proposed method with traditional approaches.

In Fig. 21, we present a concise comparison of three confusion matrices obtained on the test-set data. Each fault class in the test data has 100 samples, and the axes denote predicted fault class and actual fault label. Fig. 21a shows outcomes of the pure model-based approach with suboptimal fault isolation due to shared fault signatures. Fig. 21b displays a confusion matrix for a supervised method, significantly improving isolation, but with potential misclassifications. Fig. 21c illustrates our proposed SSL method, fine-tuned with 8 samples per fault class, showcasing a 20% higher F1-score than the pure supervised method, emphasizing enhanced performance.

We have also included two additional experiments (Section 7.2.5, 7.2.6): one to demonstrate the effect of hierarchical combination on overall performance and another to showcase the influence of the amount of unlabeled data available during pretraining on the overall performance.

We believe that these additional illustrations provide a clearer and more detailed comparison, addressing the concerns raised by the reviewer.

**

**Where the revisions are made:**

page-22, paragraph-4,5

page-23, paragraph-1

page-27