



Computational RSOM-MapReduce Intelligence for Real Time Faults Detection in Wind Energy Generator

Mohamed Salah Salhi¹, Mounir Bouzguenda², Nejib Khalfaoui¹,
Ezzeddine Touti^{3,*}

¹National Engineering School of Tunis. Research Laboratory of Signal Image and Information Technology LR-SITI, University of Tunis El Manar-Tunis Tunisia

(Mohamed s. s.) medsalah.salhi@enit.utm.tn, (Nejib K.) nejib.khalifaoui@isetj.rnu.tn

² Department of Electrical Engineering, College of Engineering, King Faisal University, Al Ahsa, 31982, Saudi Arabia, mbuzganda@kfu.edu.sa.

³ Center for Scientific Research and Entrepreneurship, Northern Border University, Arar 73213, Saudi Arabia, esseddine.touti@nbu.edu.sa

*Correspondence: esseddine.touti@nbu.edu.sa

Abstract:- This paper investigates a computational intelligence approach using the Recurrent Self-Organizing Map (RSOM) and its MapReduce framework for anomaly detection in wind energy generation. Given that wind power represents approximately 25% of the global installed renewable energy capacity, its reliability is crucial. The proposed methodology leverages an intelligent dynamic unsupervised deep learning algorithm within a distributed MapReduce processing paradigm to diagnose and isolate faults in real time across various wind energy sources. The applied computational approach differs from existing methods by simultaneously analyzing, in adverse environments, multiple signals acquired in real time for the purpose of intelligent diagnosis of wind turbine systems located in remote mountainous regions. Signal acquisition and processing were carried out in an experimental setup. The findings provide insights into its potential benefits, limitations, and economic viability.

Keywords: Wind energy generation; computational MapReduce intelligence; unsupervised deep learning; distributed processing; anomaly detection

1. INTRODUCTION

Wind energy is one of the fastest growing and most reliable sources of renewable energy, offering a clean and sustainable alternative to fossil fuels. By harnessing the kinetic energy of the wind through turbines, wind energy converts natural air movement into electricity without producing harmful emissions or depleting finite resources. As the world transitions toward more sustainable energy systems, wind energy plays a critical role in reducing greenhouse gas emissions, combating climate change, and promoting energy independence. One of the key advantages of wind energy is its abundance and availability in many regions around the world. From large onshore wind farms to innovative offshore installations, wind energy systems can be adapted to diverse environments, making it a versatile and scalable energy solution. Additionally, advancements in wind turbine technology have significantly improved



efficiency, reduced costs, and increased the capacity of wind energy, making it more competitive with traditional energy sources [1], [2].

Beyond its environmental benefits, wind energy contributes to economic growth by creating jobs in manufacturing, installation, and maintenance, while also reducing reliance on imported fossil fuels. As a renewable energy source, wind energy is not only an essential component of a sustainable energy future but also a key driver for achieving global energy security and resilience.

Due to the need for maximum wind exposure, these systems are often installed in remote mountainous areas, making maintenance and fault diagnosis challenging [3], [4]. The dynamic and highly variable nature of the signals captured further complicates anomaly detection. Over the past decade, various approaches have been developed to improve fault recognition. Statistical Process Control (SPC) has been widely used, relying on control charts and statistical thresholds to identify deviations in process variables [5], [6]. Supervised machine learning techniques, such as decision trees, support vector machines (SVMs), and neural networks trained on labeled data, have been employed to classify anomalies accurately. In parallel, unsupervised machine learning methods have emerged, using clustering algorithms like k-means and DBSCAN, as well as density estimation techniques such as Gaussian Mixture Models (GMM) and One-Class SVM, to detect anomalies without requiring labeled data [7]. Deep learning has also made significant contributions by leveraging architectures like autoencoders, Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs) for complex pattern recognition tasks [8], [9]. Additionally, time-series analysis methods, including ARIMA, exponential smoothing, and state-space models, have been applied to detect anomalies that evolve over time [10]. More recently, ensemble methods have gained attention by combining multiple approaches to enhance detection accuracy and robustness [11]. While effective, these models depend heavily on high-quality training data, are sensitive to outliers and class imbalances, and may struggle to adapt to environmental variations. To address these limitations, this paper proposes an intelligent RSOM-MapReduce approach, integrating Recurrent Self-Organizing Maps (RSOM) convolved with a distributed processing paradigm. This hybrid method enhances dynamic anomaly detection by incorporating temporal dependencies and parallel data processing.

The proposed system processes three distinct wind generator signals in two stages: first, applying RSOM for adaptive pattern recognition and second, leveraging the MapReduce framework for parallel anomaly isolation. This paper demonstrates the suitability of a recurrent deep neural network model, investigates its application in wind energy diagnostics, and evaluates the use of the MapReduce paradigm for distributed fault detection.

Following this introduction, Section 2 provides a comparative review of relevant models reported in the literature. Section 3 details the proposed methodology developed to tackle the key challenges in this domain. Section 4 presents the experimental implementation and the results obtained, while Section 5 offers a thorough analysis and discussion of these findings.

2. REVIEW ON THE INTEGRATED RECURSIVE SOM MODEL

The Recursive SOM (RSOM) is derived from the unsupervised Kohonen SOM algorithm and was introduced by Thomas Voegtlin, drawing inspiration from Elman's SRN (simple recurrent network). Elman's SRN modifies the perceptron network by incorporating a hidden layer, using



a delayed version of its activities as an additional input. Its primary function is to establish connections between input and preceding output sequences [12].

In this framework, each neuron is trained to encode a specific pair comprising input and context. Consequently, the learning process involves iteratively acquiring representations of longer sequences, building upon the already acquired representations of shorter sequences. The quantization error can be expressed by the following mathematical model [13]:

$$E_i = \alpha \|x(t) - w_i^x\|^2 + \beta \|y(t-1) - w_i^y\|^2 \quad (1)$$

In this equation, $x(t)$ and $y(t-1)$ represent, respectively, the input sequence and that of the context (previous output). w_i is the 'codebook' reference, or the weight vector associated with unit i ; α and β are stability coefficients.

The network learns by associating the current input with previous states of activity. Therefore, each neuron becomes responsive to an input sequence, and the BMU is given by [14]:

$$V = \arg \min \{E_i\}; i \in N \quad (2)$$

The best matching unit vector V is minimizing the Euclidian distance between the input vector and the weight vector. The learning rules used to update the feed-forward and recurrent weights are given by [15], [16]:

$$\Delta w_i^x = \gamma \cdot h_{iv} \cdot (x(t) - w_i^x) \quad (3)$$

and

$$\Delta w_i^y = \gamma \cdot h_{iv} \cdot (y(t-1) - w_i^y) \quad (4)$$

Given that h_{iv} represents the neighborhood function applied to the RSOM map, the dynamic nature of recurrent links can lead to unstable representations. To mitigate this instability, one approach is to apply a transfer function F of the exponential type, empirically chosen to ensure continuity and constrain its values between 0 and 1.

$$y_i(t) = F(E_i) = \exp(-\alpha \|x(t) - w_i^x\|^2 - \beta \|y(t-1) - w_i^y\|^2) \quad (5)$$

The concept of employing a transfer function to stabilize the learning process of recursive SOM originates from its modeling using a single-neuron network. The output state, denoted as Y , is determined by the transfer function of the sum of input weights [17].

Thomas Voegtlin validated the self-referential algorithm using a two-dimensional recursive SOM comprising 20x20 neurons. This SOM was trained on "Aldous Huxley's Brave New World", an English text where each letter is represented in 5-bit encoding and presented to the network as individual inputs. Punctuation symbols were excluded from the text, and the neural activity was reset between consecutive words [18]. The learning rate was constant: $\gamma = 0.01$.



The neighborhood function has a constant size: $\sigma = 0.5$. Other parameters are chosen: $\alpha = 3$ and $\beta = 0.7$.

When the leaky integrators of the unit's outputs are relocated to the inputs, a modified TKM, known as the Recurrent Self-Organizing Map (RSOM), is obtained. This adjustment results in a leaky time-difference vector, defined by the following relation [19], [20]:

$$y_i(t) = (1 - \alpha) \cdot y_i(t - 1) + \alpha \cdot (x(t) - w_i(t)) \quad (6)$$

In the equation α , a constant between 0 and 1 designates the leakage coefficient which replaces d in the temporal Kohonen map (TKM). The d is a time constant between 0 and 1, which denotes a forgetting term. An RSOM unit is therefore schematized as in Figure 1 below:

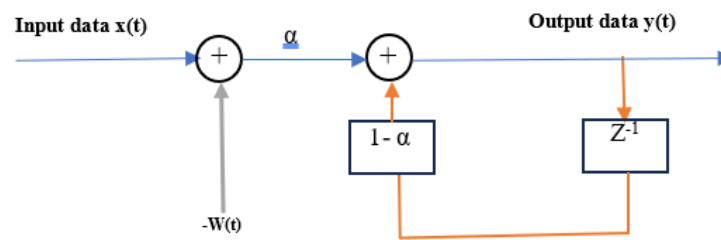


Figure1. Representation of an RSOM unit that acts as a recurrent filter

This suggests that a high α value signifies shorter memorization, while a low α value indicates longer memorization and slower activation decay. When α equals 1, RSOM functions as a regular SOM, as seen in the equation of the output $y_i(t)$ [21]. The incorporation of past algorithm executions in a recursive manner has been implemented in the Recurrent SOM (RSOM).

The Kohonen learning algorithm serves two primary purposes. First, it identifies optimal prototypes that represent the dataset, a process known as vector quantization. Second, it organizes these prototypes on a map such that nearby prototypes in the data space correspond to neighboring neurons on the map. This proximity is typically defined either by an Euclidean metric or by the topological arrangement of neurons that react similarly to input data. This algorithm is considered efficient: when an individual is exposed to the network, neurons engage in competition until only one emerges as the winner. The winning neuron possesses the prototype with the smallest Euclidean distance from the exposed individual. Competitive learning entails reinforcing the winning neuron, making it more responsive to subsequent exposures of the same individual. Meanwhile, connection weights are adjusted based on inputs. Ultimately, neurons in a competitive learning network function as detectors, each aiming to identify a unique characteristic present in the input data [22]. The victorious neuron from the competition phase determines the center of a map area termed the "neighborhood," whose size varies over time. The update or adaptation phase relocates the prototypes to align them with the input presentations to the network. This algorithm is characterized by several key



parameters. It uses a 10×10 neuron grid, providing a map with 100 processing units to facilitate straightforward percentage-based analysis. The chosen topology is hexagonal, offering a well-structured and coherent visualization. Training is performed in an unsupervised manner, making it suitable for handling large volumes of data. Additionally, the algorithm is set to run for 200 iterations to enable real-time processing.

Hierarchical methods are commonly applied in scenarios requiring the breakdown of a complex task into several simpler sub-tasks across multiple levels, aiming for more precise identification outcomes. Each hierarchical level comprises one or multiple RSOM maps, often delineated by various time scales. The differentiation among current methodologies lies in how RSOM outcomes from one level are encoded for transition to another. Variations are noted in the quantity of levels used, the quantity of RSOM units per level, and the interconnections between these diverse levels. Using of a specific hyperbolic lattice structure significantly accelerates the search process for larger maps [23], [24]. The growing hierarchical RSOM (GHRM) represents an enhancement to the RSOM map's capabilities in two key aspects. Firstly, it incorporates an increasingly sophisticated version of RSOM. Secondly, it tailors the RSOM map to accommodate tree-structured data [25], [26]. Figure 2 illustrates the hierarchical RSOM architecture in Anomaly detection.

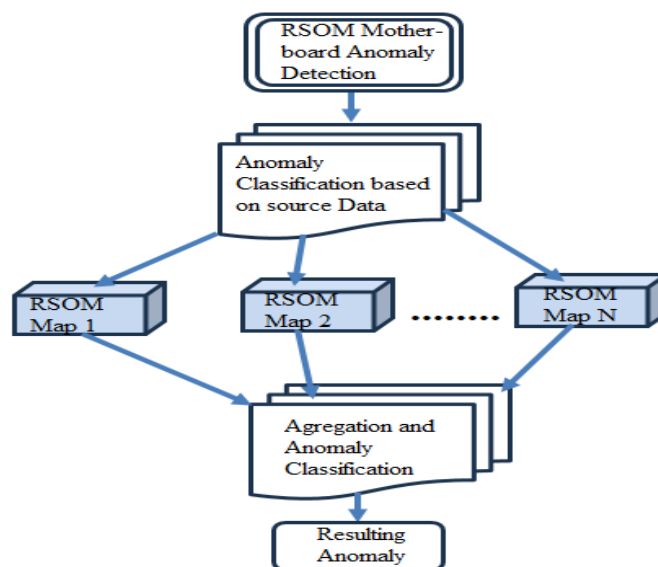


Figure 2. Diagram of a hierarchical RSOM variant in anomaly detection

This model comprises two tiers. The initial tier consists of a singular motherboard, RSOM, responsible for classifying the four anomaly categories: electrical, mechanical, thermal, and environmental. The subsequent tier encompasses multiple cards dedicated to classifying anomalies within each category. Notably, the first tier of the hierarchy undergoes comprehensive learning using the entire dataset of signal segments labeled with macro class identifiers.



The RSOM maps in the second tier are trained on segmented datasets, with each card serving as an independent module. However, a drawback of this hierarchical RSOM structure is that incorrect classification at the initial level can impact the results in the subsequent level. To ensure ease in fault recognition rates, the size of the RSOM map at the second level for recognition purposes is set at 10×10 . This model is specifically employed for anomaly classification at the source of its origin. The fusion in a convolutional way of RSOM and Map Reduce is implemented within the terminal board of the Wind Energy Generator System presented by Figure 3. This system is a typical wind energy generation setup designed to convert wind energy into electrical power and integrate it into the electrical grid. The wind turbine captures the kinetic energy of the wind and converts it into rotational mechanical energy through its blades, which are connected to a shaft that drives the generator. The induction generator (IG) then converts this mechanical energy into electrical energy. Operating asynchronously, the IG is particularly suitable for variable-speed wind turbines, enabling it to adapt to fluctuating wind conditions efficiently.

The power converter ensures the generated power. First, the rectifier converts the variable AC power from the generator into DC, and then the inverter converts the DC power back into a controlled AC output with stable voltage and frequency. This consistency is crucial for seamless integration with the grid. The transformer steps up the voltage of the conditioned power to meet the high-voltage requirements of the grid, ensuring efficient long-distance power transmission. Finally, the generated electricity is fed into the grid, where synchronization of voltage, frequency, and phase is achieved through the combined efforts of the power converter and transformer. This comprehensive system ensures reliable and efficient generation and delivery of wind energy to consumers.

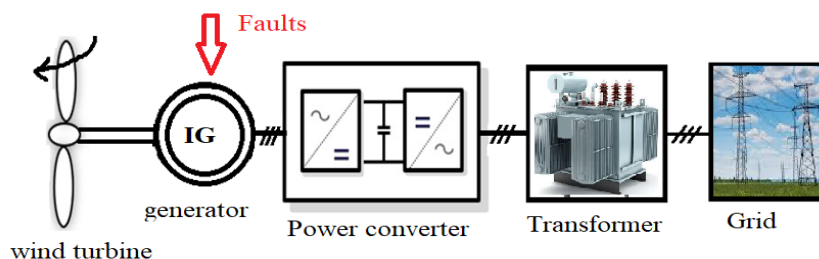


Figure 3. Power system wind energy generation

To ensure the reliable operation of the wind energy generation system, it is critical to diagnose faults that may occur within its components. Faults in these systems can arise in various parts, such as the wind turbine, induction generator, power converter, transformer, or sensors used for monitoring. These faults, if left undetected, can lead to reduced efficiency, increased downtime, and even catastrophic failures, compromising the reliability of energy production. Diagnosing these faults in wind generators is particularly challenging due to the dynamic and variable nature of wind energy systems, as well as the complexity of signals generated by the components. This study focuses on developing an advanced fault diagnostic approach for the wind generator in the presented system, leveraging a hybrid Convolutional RSOM MapReduce



framework. By integrating efficient data processing and intelligent anomaly detection techniques, the proposed method aims to identify, classify, and isolate faults accurately and in real time, ensuring the system's continuous and optimal operation.

3. PROPOSED COMPUTATIONAL RSOM MAPREDUCE MODEL

The MapReduce model operates as a distributed processing paradigm, initially popularized by Google in its implementation that involved a file system distribution for data exchange. However, subsequent implementations have surfaced targeting diverse architectures and communication channels, including shared memory systems or distributed systems with varied communication protocols. Each MapReduce algorithm incorporates two pivotal functions: the map function and the reduced function. The map function transforms a dataset into another dataset, organizing individual elements into grouped pairs (key/value). Subsequently, the reduced function processes the output of the map function by consolidating these key-value pairs into a smaller set. The primary advantage of this model lies in its ease of deploying data processing across multiple computing nodes. Initially breaking down a data processing application into mappers and reducers might seem unconventional. However, once an application is built based on the MapReduce paradigm, scaling it to operate on hundreds, thousands, or even tens of thousands of machines within a cluster merely requires configuration adjustments. This straightforward scalability is a key factor that entices numerous programmers to use this model for distributed processing of vast volumes of data. The essence of the MapReduce paradigm lies in directing processing tasks to the location of the data itself. This program operates through three distinct steps: the map step, the shuffle step, and the reduce step (refer to Figure 4). The map step involves processing the input data, typically stored in the Hadoop File System (HDFS), presented in the form of a file or directory. The input file is parsed line-by-line and sent to the map function for processing, generating multiple smaller data items.

The Shuffle step encompasses copying, organizing, and merging the outputs from the mappers. This process generates datasets structured as key-value pairs, which subsequently undergo processing by the reducer. The Reduce step aggregates the results derived from the mappers. Once processed, it generates a fresh output set, which is then stored back into the HDFS. Most computations occur on nodes equipped with data stored on local disks, effectively minimizing network traffic. Upon completion of data processing tasks, the cluster consolidates and condenses the data to produce the desired outcome, sending it back to the Hadoop server. Refer to Figure 4 below for a visual representation.

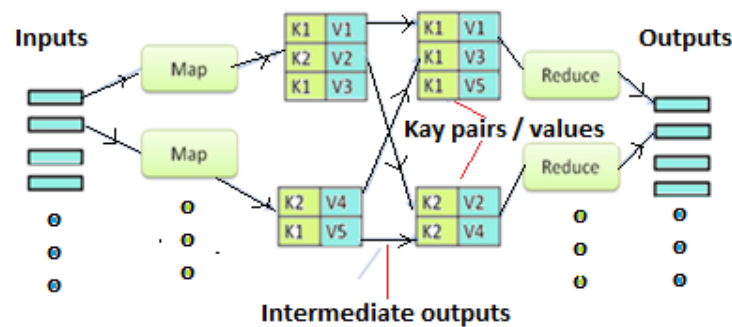


Figure 4. Schematic diagram of the MapReduce model

As example, we suppose that with four data elements: $D = \{d_1, d_2, d_3, d_4\}$, we use two shuffle keys: $\{k_1, k_2\}$. Having in key mapping: $d_1 \rightarrow k_1, d_2 \rightarrow k_2, d_3 \rightarrow k_1, d_4 \rightarrow k_2$.

We use a transposed vector of values: $v = [2, 5, 3, 7]^T$, and the following key-assignment

matrix $K = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$,

the aggregated result will be $R = K \cdot v = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 + 3 \\ 5 + 7 \end{bmatrix} = \begin{bmatrix} 5 \\ 12 \end{bmatrix}$ (7)

The MapReduce provides as result: total for k_1 equal 5 and total for k_2 equals 12.

The proposed method capitalizes on the advantages offered by the MapReduce algorithm in conjunction with the RSOM deep learning model integrated into a single board for the control of wind energy generators. This system serves as a parallel processing model for handling vast amounts of data, aiming to minimize the time necessary for anomaly detection and isolation. Notably, the time required by the RSOM map for anomaly detection is notably significant, particularly during the training phase, which in this scenario takes 34 minutes for processing the three signals; electrical, mechanical and thermal. To tackle this challenge, a distributed approach is introduced for processing electrical and mechanical signals across multiple RSOM maps, aiding in the detection and isolation of anomalies. This approach leverages pipelined information handling for massive data (Big Data), employing distributed processing techniques across various RSOM maps, consolidating their outputs into a reduced dimension to facilitate fault recognition decisions at the final layer. This solution harnesses the Hadoop MapReduce framework and the RSOM Deep Learning approach, significantly improving the detection, isolation, and classification of outsourced anomalies within a reasonable real timeframe, particularly during the supervision stage. Upon signal processing completion, a distribution of each signal matrix occurs to facilitate the analysis and defect detection based on their respective indicators. Initially, the submatrices undergo exploitation through mappers, applying the RSOM map for anomaly detection. Subsequently, the reducers aggregate this information and generate the conclusive catalog of classified and isolated defects. Finally, the outcome is showcased on an edge panel for visualization.



The adopted strategy in detecting and isolating anomalies from wind energy generation systems revolves around the following Algorithms:

Algorithm 1. Map Function

Require: signals matrix M_i

Ensure: key/value pairs of sources anomalies

```
1:  $RE_i \leftarrow RSOM(M_i)$ 
2:  $An_i = \emptyset$ 
3: FOREACH Frequency  $f$  in  $RE_i$  DO
4:   IF  $f \neq h$  THEN
5:     Add  $f$  to  $An_i$ 
6:   END IF
7: END FOR
8: Emit Intermediate ( $Source_i, An_i$ )
```

Algorithm 2. Reduce Function

Require : set of key/value pairs $\langle source, anomalies \rangle$

Ensure key/value pairs, $\langle source, anomalies \rangle$

```
1: FOREACH  $source_i$  DO
2:   Emit Intermediate ( $Source_i, anomalies$ )
3: END FOR
```

4. EXPERIMENTAL RESULTS

The experimental setup comprises several key components. First, the Wind Energy Generator consists of a 12V DC wind turbine used to drive a flow servo mockup. The Control System is based on a Siemens S7-1215 PLC, which regulates the system through Digital-to-Analog (DAC) and Analog-to-Digital (ADC) converters. This PLC is also equipped with an electromechanical flowmeter capable of measuring wind flow rates ranging from 0 to 1.2 knots per minute. The Siemens S7-1215 PLC features 14 digital inputs and 10 digital outputs, along with 2 analog input channels and 2 analog output channels, each with 12-bit resolution. Additionally, an Arduino MEGA board is integrated into the setup, providing 54 digital I/O ports, 16 analog inputs via its ADC, and operating with a 16 MHz oscillator. The Arduino is powered through either a USB 5V supply or a DC input ranging from 7V to 12V. The experimental bench is illustrated in Figure 5.



Figure 5. Photography of the experimental bench



Signal measurements at the sensor level of the experimental bench are shown in Figures 7, 9, and 10, highlighting abnormalities under nominal operating conditions from various sources. Three types of checks are performed: first, a comprehensive examination of the entire wind energy generator system, with Figure 6 illustrating the referenced healthy signal.

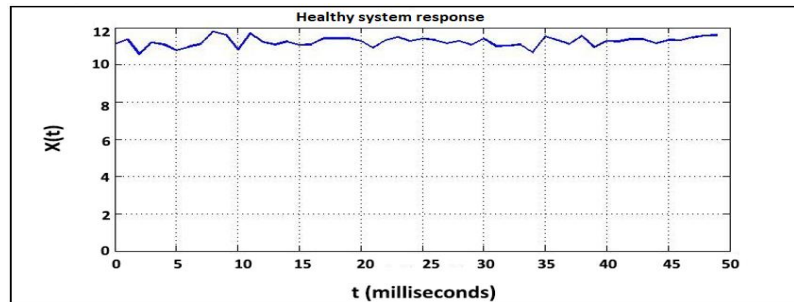


Figure 6. Recording the healthy signal from the wind energy generator system

Following a system fault, the acquired signal state alters as illustrated in figure 7.

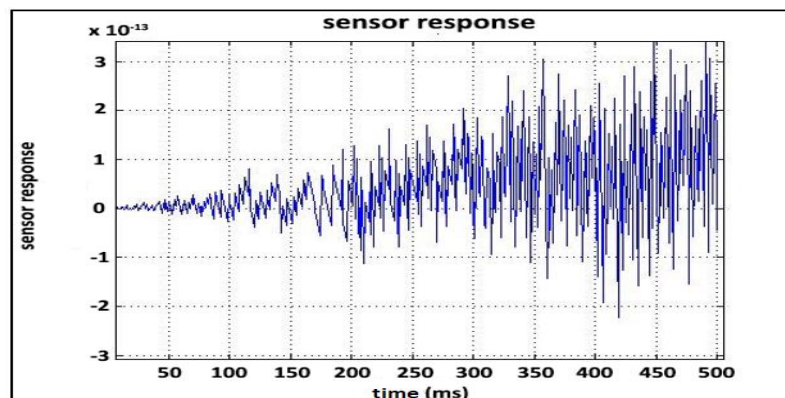


Figure 7. Recording signal 1 from the malfunctioning wind energy generator over time domain

The spectral analysis of the flawed signal 1 indicates that the fault manifested at a frequency of 12.3 Hz, depicted in figure 8.

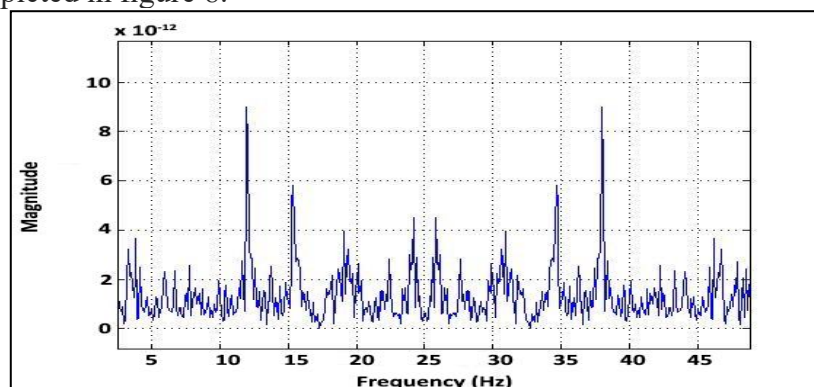


Figure 8. Spectrum of signal 1 from faulty wind energy generator ($f_p = 12.3$ Hz)



Another inspection targeted the control card to verify the wind energy generator system, with Figure 9 showing the status of the acquired flawed signal 2.

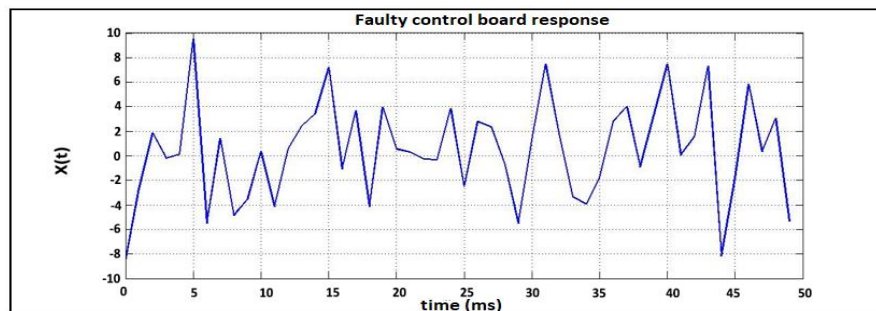


Figure 9. Recording signal 2 from the defective control card in the temporal domain

The spectral examination of the flawed signal 2 reveals the fault's occurrence at a frequency of 62 Hz, as depicted in figure 10.

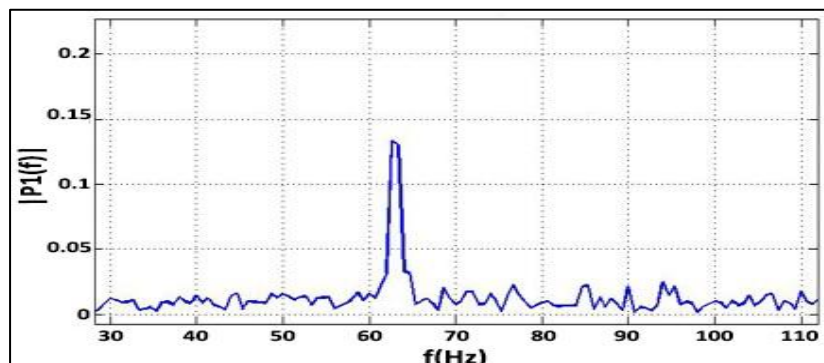


Figure 10. Spectrum of signal 2 from faulty control card ($f_c=62$ Hz)

A final examination was performed on the sensor component. Figure 11 illustrates details concerning the status of the acquired defective signal 3.

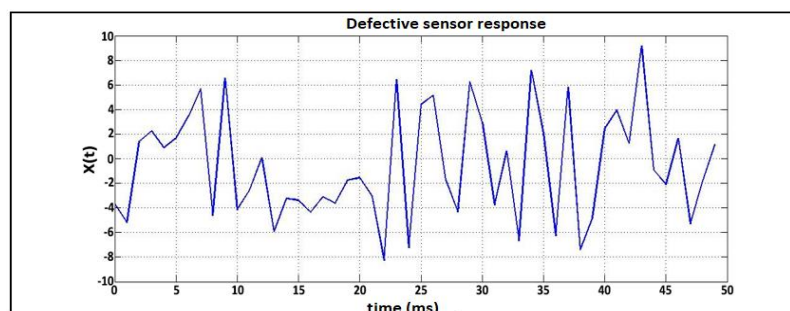


Figure 11. Recording signal 3 from the defective sensor in the temporal domain

The spectral examination of the flawed signal 3 indicates the fault's occurrence at a frequency of 75.6 Hz, as depicted in Figure 12.

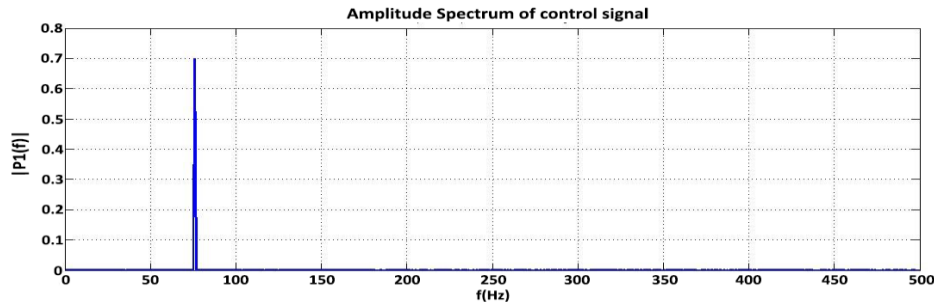


Figure 12. Spectrum of signal 3 from faulty sensor ($f_c=75.6$ Hz)

While spectral analysis of faults in the wind energy generator system enables the identification of fault frequencies, it remains subjective in determining the nature of these faults at this stage. It does not necessarily provide insight into the specific type of fault occurring. An objective solution consists of involving the RSOM model to target with precision the frequency and nature of the analyzed defect. Moreover, its defects are localized by characteristic frequencies, thus, defined by: fp characterizes the power generator fault, fe characterizes the control card fault and fc characterizes the sensor fault. An unbiased analysis of these defects using the RSOM deep learning model provides a genuine representation seen through the topology depicted in Figure 13.

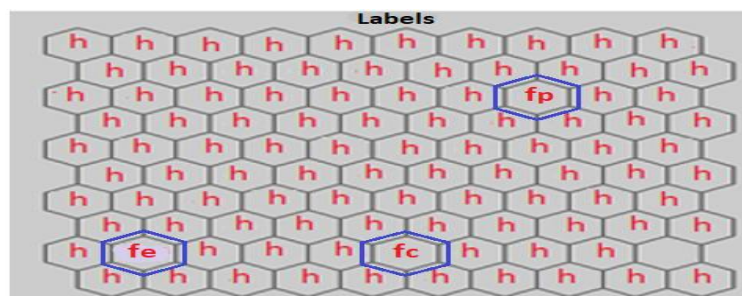


Figure 13. Visualization of occurring faults over the RSOM topolog

The various stages of processing scenario involving the three signals are also outlined in figure 14 below, demonstrating the fusion of the RSOM model with MapReduce.

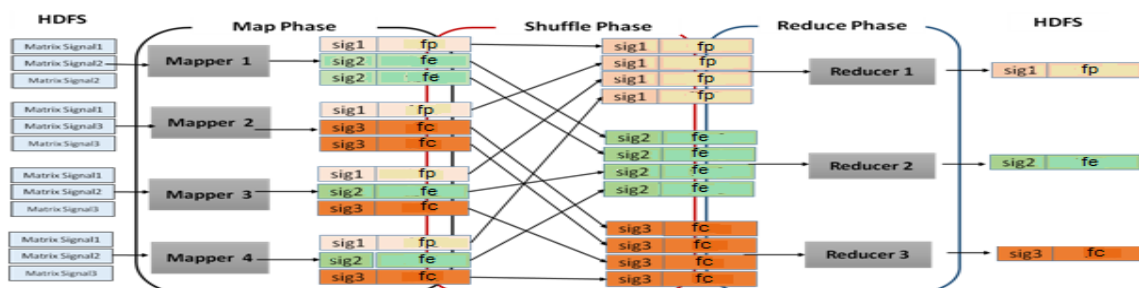


Figure 14. Implementation of the MapReduce methodology



As depicted in Algorithm 2 above in section 3, the reducer initiates the process by computing the respective anomalies for each source. Given the possibility of multiple pairs having identical values, the reducer necessitates sorting them based on increasing frequency and the count of anomalies. Ultimately, the reduced function delivers key/value pairs in the format of Source, anomalies, where 'Source' denotes a processed signal and 'anomalies' are defined by their representative frequency. The subsequent pair serves as an illustration of the output yielded by the Collapse function:

$\langle \text{signal1} ; \{fp\} \rangle$

Here, the key 1 signal is captured from the process while the value $\{fp\}$ represents the frequency indicating abnormalities. After anomaly isolation and classification, the result will be displayed by the dashboard as shown below:

$\langle \text{signal}_1; fp \rangle$ assigned to a fault in wind generator.

$\langle \text{signal}_2; fe \rangle$ assigned to a fault in control card.

$\langle \text{signal}_3; fc \rangle$ assigned to a fault in sensor.

5. RESULTS DISCUSSION

The strategy we've employed leverages the collective benefits associated with hybrid delta-models. In our Wind Energy Generator System, three signals are relayed via sensors for monitoring purposes. The initial signal indicates the operational status of the generator, the second signal monitors the condition of the generator's control card, and the third signal gauges the integrity and reliability of the sensor itself. These signals undergo an initial filtration process via map reduce aggregation. Subsequently, the RSOM model, functioning as a classifier, executes a secondary selective filtration by analyzing and making decisions regarding the identification of potential defects. Figures 6 to 12 present visual representations that necessitate subjective assessment relying on human observation, which can potentially lead to inaccuracies due to the fallibility of the human eye. Conversely, the proposed method relies on rigorously scientific algorithms and mathematical models. This approach enables an objective evaluation based on scores and recognition rates facilitated by an intelligent neural model, complemented by a clearly defined visualization of outcomes. As depicted in Figure 13, three faults, labeled as Fp, Fc, and Fe, are discernible by their respective frequencies. Additionally, among the **100** neurons on the RSOM map, only one neuron remains unassigned, signifying a momentary confusion in decision-making. The remaining neurons are denoted by 'h', indicating their healthy status. This observation leads to the inference that this model achieves a recognition rate of **99%**, as there is only one neuron out of 100 that exhibits confusion, affirming the model's efficiency compared to other indicated models. As an example, empirical findings from statistical models, particularly the HMM model, demonstrate a trap failure recognition rate ranging between 80% and 88%, contingent upon the type and dataset used. These models are primarily focused on identifying systematic and repetitive



defects. Moreover, the employed RSOM Map reduce approach excels in handling substantial data volumes for detection through unsupervised learning, as opposed to supervised machine learning models. Similarly, the chosen approach is driven by its rapid reaction and response time, facilitated by its parallel data processing structure, which operates at **50 μ s**. This stands in contrast to the response time of **17 ms** observed when employing Deep Learning models with multiple layers of hierarchical neurons. These allocated time results are established based on the implementation of this system in real-time applications, particularly in megawatt-scale wind turbines. This outcome validates the compatibility and consistency of the chosen delta-hybridization method, showcasing its robustness against varying environmental conditions. This advantage positions it favorably in contrast to simpler existing techniques. Nevertheless, it poses a challenge due to the considerable duration required during the learning phase. Once the diagnostic model undergoes its learning iterations and is fully adopted, it becomes adept at identifying familiar fault types. This capability enables the model to promptly react in real-time during the testing phase when integrated with the system to be monitored.

6. CONCLUSION

This paper presented an intelligent computational RSOM-MapReduce approach for anomaly detection and fault diagnosis in wind energy generation systems. The integration of Recurrent Self-Organizing Maps (RSOM) with the MapReduce distributed processing paradigm provides a robust and scalable framework capable of handling large volumes of data in real-time. The proposed model achieves a fault recognition rate of **99%**, significantly outperforming traditional statistical and deep learning models in terms of both accuracy and efficiency. Its ability to adapt to fluctuating environmental and operational conditions ensures reliable performance in the dynamic and complex domain of wind energy systems.

The experimental results validate the system's effectiveness in detecting and isolating faults related to the generator, control card, and sensor components, with a response time of **50 μ s**, which is well-suited for real-time applications in wind turbines. As perspective, the techno-economic analysis could further highlight the model's feasibility, demonstrating its cost-effectiveness, while achieving a **50%** reduction in downtime and improving energy production by more than **12%**, ensuring a high system availability of **99%**. The integration of the RSOM model with the MapReduce framework not only enhances fault detection and isolation but also establishes a foundation for deploying scalable and efficient monitoring systems in large-scale wind farms. The system's compatibility with existing hardware and its potential for parallel and distributed processing make it a promising solution for modern wind energy management and environmental sustainability.

Acknowledgements

The authors extend their appreciation to the Northern Border University, Saudi Arabia for supporting this work through project number "NBU-CRP-2025-2448".



References

- [1] Raj, V. Ravi, Prabakaran, M. and Selvakumar, P: 2025, "Energy Efficiency and Renewable Energy Adoption". Driving Business Success Through Eco-Friendly Strategies. IGI Global Scientific Publishing, p. 131-148.
- [2] Savio F. Max, Joshua, S. Vinson and Usha K: 2025, "Design of a Solar-Wind Hybrid Renewable Energy System for Power Quality Enhancement: A Case Study of 2.5 MW Real Time Domestic Grid". Engineering Reports, vol. 7, no 1, p. e13101.
- [3] Luo Xiaoli, Ning JinYE and Wu Min: 2025, "Design and analysis of wind turbine fault diagnosis system based on convolutional neural network".Diagnostyka, 2025, vol. 26, no 1.
- [4] Priya G. Suriya, Geethanjali M. and Devi M. Meenakshi: 2025, "Active Protection Strategy for Transmission Networks with Wind Distributed Generation Systems Based on SVM". Journal of Electrical Engineering & Technology, p. 1-10.
- [5] Nesma Saleh and MA Mahmoud: 2024, "A re-evaluation of repetitive sampling techniques in statistical process monitoring", Taylor & Francis-Quality Technology & Quantitative Management, vol. 21, no. 5, pp. 786-804.
- [6] Vysakh Raveendran and PT Anjana: 2023, "Moving towards process-based radiotherapy quality assurance using statistical process control", Elsevier-Physica Medica, volume 112.
- [7] Anila Sebastian and Hira Naseem: 2024, "Unsupervised Machine Learning Approaches for Test Suite Reduction", Taylor & Francis-Applied Artificial Intelligence An International Journal, Volume 38, Issue 1.
- [8] R Noor, A Wahid, SU Bazai, and A Khan: 2024, " Undersampled mri reconstruction using deep learning based generative adversarial network", Elsevier- Biomedical Signal Processing and Control. Volume 93.
- [9] S Mihandoost and S Rezvantalab: 2024, "A Generative Adversarial Network Approach to Predict Nanoparticle Size in Microfluidics", ACS Publications-Modeling and Informatics Tools, vol. 11, issue 1. pp. 268–279.
- [10] Jiu Oha, Byeongchan Seong: 2024,"Forecasting with a combined model of ETS and ARIMA",Communications for Statistical Applications and Methods,Vol. 31,No.1,143–154.
- [11] Farhan Ullah, Irfan Ullah and Rehan Ullah Khan: 2024, " Conventional to Deep Ensemble Methods for Hyperspectral Image Classification: A Comprehensive Survey ". IEEE Xplore- Journal of selected topics. Volume 17.
- [12] Mohamed Salah Salhi, Manel Salhi and Ezzeddine Touti: 2024, "On the Use of Wireless Sensor Nodes for Agricultural Smart Fault Detection", Springer Nature- Wireless Personal Communications, Volume 134, pages 95–117.
- [13] Shan Huang, Hailong He and Robby Zachariah Tom: 2024, "Non-invasive optoacoustic imaging of dermal microcirculatory revascularization in diet-induced obese mice undergoing exercise intervention", Elsevier- Photoacoustics, Volume 38.
- [14] Roshan Kumar and Mala De:2024, "Power system resilience quantification and enhancement strategy for real-time operation", Springer Nature- Electrical Engineering, Volume 106, pages 6227–6250.
- [15] Raunak Raj, Avinash Awasthi and Komal Arora: 2024 "Ultra-Wideband Technology for Healthcare Monitoring: A Machine Learning Approach", IEEE Xplore- International Conference on Cybernation and Computation (CYBERCOM).



- [16] Manyu Xiao, Jun Ma, Xinran Gao and Piotr Breitkopf: 2024, “Primal–dual on-the-fly reduced-order modeling for large-scale transient dynamic topology optimization”, Elsevier- Computer Methods in Applied Mechanics and Engineering, Volume 428.
- [17] Xiuting Li, Yik Weng Yew, and Keertana Vinod Ram: 2024 “Structural and functional imaging of psoriasis for severity assessment and quantitative monitoring of treatment response using high-resolution optoacoustic imaging”, Elsevier- Photoacoustics, Volume 38.
- [18] Lanjun Liu and Dechuan Wang: 2024, “A Hierarchical Heuristic Architecture for Unmanned Aerial Vehicle Coverage Search with Optical Camera in Curve-Shape Area”, MDPI- Remote Sensing, Vol. 16, Issue 15.
- [19] Hesdras O. Viana and Aluizio F. R. Araújo: 2024, “ Self-organizing speech recognition that processes acoustic and articulatory features”, Springer Nature- Multimedia Tools and Applications, Volume 83, pages 39169–39195.
- [20] Yongqi Xu, Yujian Lee and Rong Zou: 2024, " LSROM: Learning Self-Refined Organizing Map for Fast Imbalanced Streaming Data Clustering ", Machine Learning- arXiv.
- [21] Shashikant Verma, A.D. Prasad and Mani Kant Verma: 2024, “Optimal operation of the multi-reservoir system: a comparative study of robust metaheuristic algorithms”. Inderscience- International Journal of Hydrology Science and Technology, Vol. 17, No. 3, pp 239-266.
- [22] A. Bautista, J.G. Pajuelo and J.A. González: 2024, “Reproductive traits of the deep-sea shrimp *Plesionika williamsi* (decapoda: Pandalidae) from the eastern-central atlantic”, Elsevier- Deep Sea Research Part I: Oceanographic, Volume 208.
- [23] Nandanwar, Siddharth, Desai, Aditya and Esfidani:2025, “Determining the optical and polaritonic properties of isotopically pure hBN using cryogenic FTIR micro-spectroscopy”. Applied Physics Letters, vol. 126, no 1.
- [24] Patino, Nicholas H., Lomazzi, Luca and De Beni 2024. “Elastic hyperbolic strip lattices”. arXiv preprint arXiv:2410.18420, 2024.
- [25] Kshitij Tripathi: 2024 “The novel hierarchical clustering approach using self-organizing map with optimum dimension selection”, Wiley-Health care science, Volume 3, Issue 2.
- [26] Wu Jinrong, Nguyen Su and Kempitiya: 2024 “A Hierarchical Machine Learning Method for Detection and Visualization of Network Intrusions from Big Data”, Technologies, Vol 12, Issue 10.