Vol. 4(2), pp. 1-7, 2025
eISSN: 2976-4009

ASEAN Journal of Process Control

Research Article

Attention-Based Gated Recurrent Unit Approach for Fault Diagnosis in the Tennessee Eastman Process

Remigius Nnadozie Ewuzie[1], **Norazwan Md Nor**[1*], **Zainal Ahmad**[1]

[1]School of Chemical Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, MALAYSIA
*Corresponding Author: chnorazwan@usm.my

Abstract: Fault diagnosis in chemical processing systems remains difficult because these operations are highly complex, nonlinear, and safety-critical; undetected abnormalities can escalate into major operational and economic losses. Traditional FDD techniques often have limited capability in handling high-dimensional sensor measurements and typically overlook time-dependent behavior, which reduces their diagnostic reliability. In this study, an attention-enhanced gated recurrent unit (Attention-GRU) model is introduced to improve process fault identification. The architecture couples GRU layers with an attention module that highlights the most relevant temporal segments, strengthening the learned feature representations. The derived context vector is then passed to a fully connected SoftMax classifier. Using the Tennessee Eastman process (TEP) as the evaluation benchmark, the method achieved average accuracy, precision, and F1-score values of about 90%, surpassing the standard GRU baseline, which recorded roughly 86%, 85%, and 85%, respectively. The findings indicate better diagnostic performance, adaptability, and earlier fault recognition, providing a more robust and reliable solution for industrial chemical process monitoring.

Keywords: Tennessee Eastman process, fault diagnosis, gated recurrent unit, Attention mechanism, classification accuracy.

1. Introduction

The increasing level of automation and increasing operational complexity in today's chemical plants have improved efficiency but also amplified the likelihood of process faults, which can trigger safety incidents, economic losses, and unexpected shutdowns [1, 2]. Chemical processes are typically nonlinear, uncertain, and highly dynamic, often operating under conditions that pose significant risks [3]. As a result, accurate monitoring, rapid fault detection, and reliable diagnosis are essential for ensuring safe and stable operation [4]. Conventional monitoring approaches, model-based, knowledge-based, and data-driven, each present limitations. Model-based methods struggle with nonlinearities and complexity, while knowledge-based systems rely heavily on expert rules and lack adaptability [5]. Data-driven approaches utilize historical process data and have produced multivariate statistical techniques such as PCA, PLS, ICA, and FDA [6].

However, these methods often depend on fixed thresholds, leading to false alarms and reduced reliability in dynamic environments. Although deep learning models such as convolutional neural networks (CNNs) and recurrent structures like RNNs or GRUs have improved detection accuracy, they still face limitations. CNNs primarily focus on spatial feature extraction and lack the ability to model long-term temporal dependencies, while basic RNNs and GRUs treat all time steps equally, disregarding the varying importance of temporal features. Consequently, these models may overlook critical information relevant to distinguishing subtle fault patterns or early fault evolution. To address these limitations, this study integrates an attention mechanism into a gated recurrent unit (GRU) framework to enhance the model's capability for temporal feature weighting and fault pattern recognition. The attention mechanism dynamically assigns higher importance to informative time steps while suppressing irrelevant ones, allowing the model to focus on significant temporal dependencies that contribute most to fault discrimination. By combining the sequential learning strength of GRU with the interpretability and selective focus of attention, the proposed model aims to achieve more accurate, robust, and explainable fault detection in the TEP.

2. Attention-Based Gated Recurrent Unit Approach

The proposed method integrates a gated recurrent unit (GRU) network with an attention mechanism to enhance fault detection in the Tennessee Eastman process). Multivariate sensor data are preprocessed and passed through stacked GRU layers to capture temporal dependencies, after which the attention module assigns adaptive weights to important time steps. The resulting weighted representations form a context vector that highlights key temporal features, which is then fed into a fully connected output layer for fault classification.

2.1. Attention mechanism

The attention mechanism, modeled after human selective focus, allows deep learning models to prioritize important inputs while downplaying less relevant information. By dynamically assigning weights to features, it improves both interpretability and robustness, making it particularly effective for modeling temporal dependencies in process data. In neural architectures, attention acts as a dynamic weighting layer that measures the similarity between query and key vectors to generate normalized weights, which reweight the corresponding values. This allows the model to focus on the most significant time steps or features for accurate fault detection. An overview of this concept is shown in Figure 1.

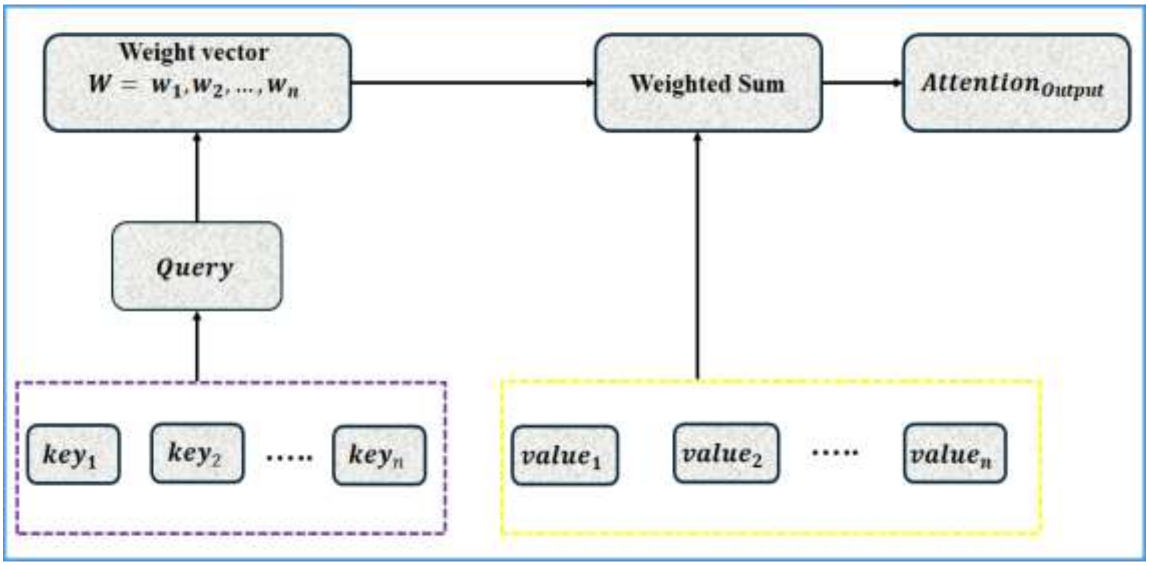

Figure 1. Schematic diagram illustrating the structure and operation of the attention mechanism.

2.2. Gated recurrent unit

The gated recurrent unit (GRU) is a variant of the recurrent neural network (RNN) designed to effectively capture temporal dependencies in sequential data. It employs two main gates, the update gate and the reset gate, to regulate the flow of information through the network. The update gate determines how much of the past information should be carried forward, while the reset gate controls how much of the previous state should be forgotten. By simplifying the gating structure compared to LSTM, GRU achieves efficient training and strong performance on time-series data. Its architecture helps mitigate the vanishing gradient problem, allowing it to learn long-term dependencies and complex dynamic behaviors commonly found in chemical process data.

2.3. Model training

An attention-based GRU architecture was developed for fault detection and diagnosis (FDD) in the TEP. Normalized multivariate sensor data were segmented into supervised sequences to preserve temporal order and process dynamics. Stacked GRU layers captured sequential dependencies, while the attention mechanism adaptively emphasized fault-relevant timesteps and suppressed redundant inputs, improving interpretability and robustness. A dropout layer (0.4) prevented overfitting, and a SoftMax output layer classified normal and faulty conditions. The model was trained using the Adam optimizer (learning rate = 0.001) for 150 epochs with a batch size of 64, applying stratified sampling to maintain balanced fault representation.

3. Experimental Setup

3.1. Dataset description

The Tennessee Eastman process is a well-established benchmark for evaluating process monitoring and fault detection methods. It simulates a complex chemical production system consisting of five major units: reactor, condenser, compressor, separator, and stripper with 52 measured variables, including 41 sensors (XMEAS) and 11 manipulated (XMV) variables (Figure 2).

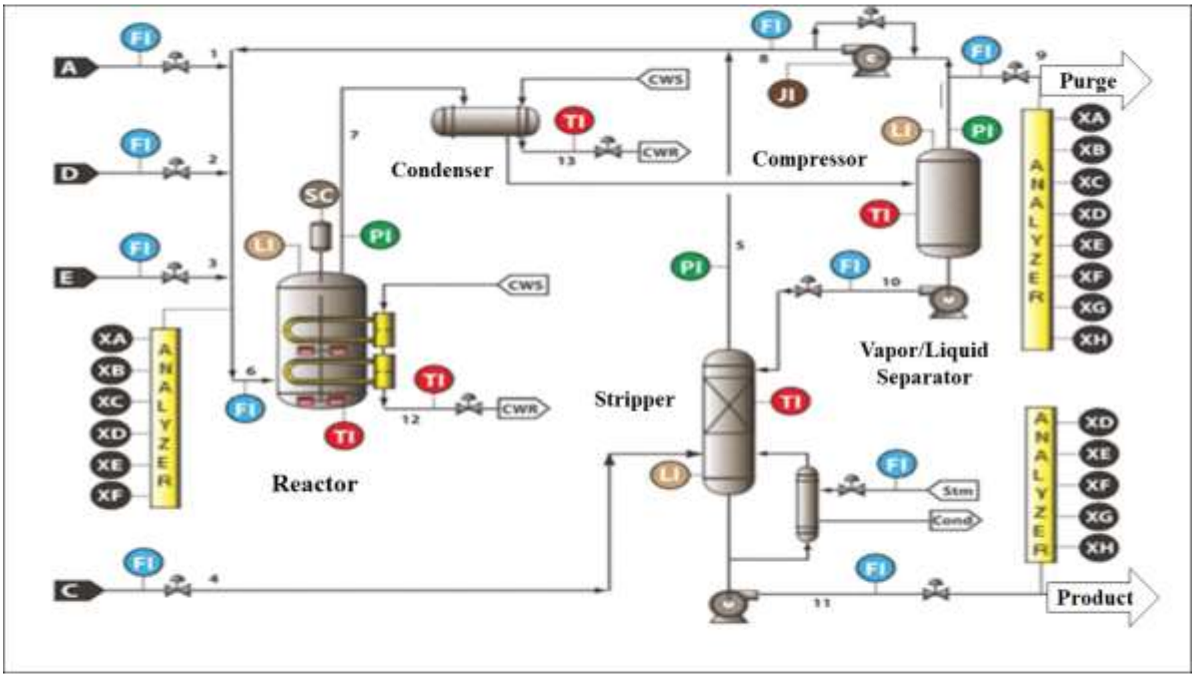

Figure 2. Schematic overview of the Tennessee Eastman process.

The process is capable of operating under normal and faulty modes, with 20 predefined fault conditions encompassing actuator, sensor, and process disturbances. In this study, both fault-free and faulty datasets were employed. The training data were collected over 24 simulation hours, while the testing data spanned 48 hours. Each fault condition comprised 480 training samples, and each corresponding test set contained 960 samples, with the fault introduced at the 160th sample. The multivariate process variables were first standardized using Z-score normalization to ensure uniform scaling and eliminate bias from variable magnitudes. A sliding window segmentation technique was then applied to transform the continuous process data into supervised learning sequences of fixed length, preserving the temporal order necessary for capturing process dynamics. The resulting samples were randomly partitioned into training (80%) and testing (20%) sets using stratified sampling to maintain balanced fault representation across all classes.

3.2. Evaluation metrics

The model's performance was comprehensively evaluated using multiple classification metrics. Accuracy quantified the overall proportion of correctly classified samples, and precision measured the reliability of positive predictions. The F1-score provided a balanced assessment, while a confusion matrix was employed to visualize class-wise prediction outcomes, facilitating detailed analysis of correctly detected and misclassified fault types.

4. Results and Discussion

As shown in Figure 3, the GRU and Attention-GRU models were evaluated on the Tennessee Eastman process across 20 fault classes. The baseline GRU achieved moderate performance, detecting several faults with high accuracy (>80%) while struggling with subtle or overlapping faults (e.g., Faults 3, 9, 15, 16, 19; <40%), often misclassifying them as similar classes. Integrating an attention mechanism substantially improved detection, enhancing diagonal dominance in the confusion matrix and increasing accuracy for challenging faults (e.g., Fault 3: 21% to 55%), yielding a 4–5% overall gain in accuracy, precision, and F1-score. These results demonstrate that attention effectively highlights fault-relevant temporal features, improving robustness and interpretability in multivariate chemical process fault diagnosis.

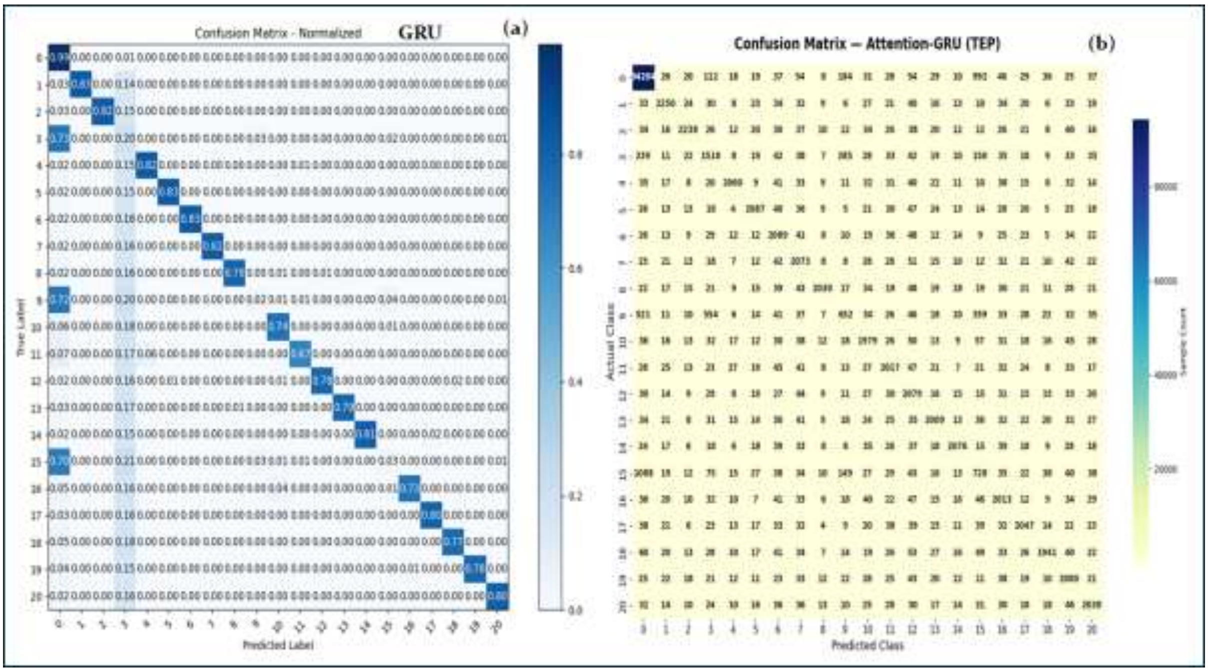

Figure 3. Confusion matrix comparison of GRU and Attention-GRU models.

The GRU model demonstrated strong overall performance in detecting faults in TEP as presented in Table 1, achieving an average accuracy, precision, and F1-score of approximately 86%, 85%, and 85%, respectively. High detection performance was observed for faults with distinct temporal signatures, such as Faults 1, 6, 7, 12–14, and 17–18, where accuracy exceeded 75% and F1-scores ranged from 82% to 89%. Conversely, the model struggled with subtle or overlapping faults, including Faults 3, 9, 15, and 16, which exhibited very low accuracy (as low as 1%) and F1-scores (<35%), indicating difficulty in capturing their patterns. Moderate performance was observed for faults such as 5, 8, 10, 11, 19, and 20, suggesting partial detectability with occasional misclassification. These results highlight that while the GRU effectively captures temporal correlations for many fault types, additional mechanisms, such as attention layers, may be required to enhance the detection of complex or low-amplitude faults.

Table 1. Performance metrics of the GRU model.

Fault ID	Accuracy (%)	Precision	F1-Score
Fault 1	83	87	86
Fault 2	83	80	81
Fault 3	21	19	20
Fault 4	82	73	77
Fault 5	85	68	76
Fault 6	83	97	89
Fault 7	82	92	87
Fault 8	75	85	79
Fault 9	15	18	46
Fault 10	52	55	53
Fault 11	54	63	58
Fault 12	76	92	83
Fault 13	75	91	82
Fault 14	80	90	85
Fault 15	1	19	2
Fault 16	37	32	34
Fault 17	79	79	79
Fault 18	76	77	77
Fault 19	47	48	47
Fault 20	66	57	61
Average	86	85	85

As shown in Table 2, the Attention-GRU model outperformed the baseline GRU, attaining average accuracy, precision, and F1-score of 90% across all metrics.

Table 2. Attention-based GRU model performance metrics.

Fault ID	Accuracy (%)	Precision	F1-Score
Fault 1	84	86	84
Fault 2	83	84	83
Fault 3	55	55	55
Fault 4	83	84	83
Fault 5	83	96	89
Fault 6	83	90	87
Fault 7	82	98	89
Fault 8	82	90	86
Fault 9	27	31	29
Fault 10	77	74	76
Fault 11	79	75	77
Fault 12	82	79	80
Fault 13	80	90	85
Fault 14	83	84	84
Fault 15	26	26	26
Fault 16	79	70	74
Fault 17	83	77	79
Fault 18	80	65	71
Fault 19	82	88	85
Fault 20	80	75	77
Average	90	90	90

The model performed particularly well for faults with clear temporal signatures, such as Faults 5, 6, 7, 8, 13, and 19, achieving F1-scores above 85%. Even moderately challenging faults, including Faults 3, 10, 11, and 16–18, showed substantial improvements compared to the baseline, demonstrating enhanced fault discrimination. While a few subtle or overlapping faults (Faults 9 and 15) remained difficult to detect, the overall results indicate that integrating the attention mechanism with GRU effectively strengthens the model's ability to capture critical temporal dependencies and improves robustness in fault detection.

5. Conclusions and future work

This study introduced an Attention-GRU approach for fault detection and diagnosis in the Tennessee Eastman process. By combining a gated recurrent unit network with a self-attention mechanism, the method effectively captured temporal dependencies and adaptively highlighted fault-relevant timesteps, enhancing both interpretability and detection performance. Experimental results indicated that the proposed approach achieved average accuracy, precision, and F1-score of 90%, surpassing the baseline GRU, which recorded approximately 86%, 85%, and 85%, respectively, particularly for faults exhibiting complex or overlapping temporal patterns. While some subtle faults remain challenging, the results highlight the potential of attention-based approaches in enhancing fault detection for complex chemical processes. Future work may explore multi-sensor data fusion and uncertainty quantification to further improve robustness and reliability in industrial process monitoring.

Acknowledgments: The authors gratefully acknowledge Universiti Sains Malaysia (USM) for supporting this work through the Graduate Research Assistantship (GRA-Assist) Scholarship, and the Ministry of Higher Education Malaysia (MOHE) for funding this research under the Fundamental Research Grant Scheme (R503-KR-FRG001-0006071538-K134).

References

1. Bi X, Qin R, Wu D, et al (2022) One step forward for smart chemical process fault detection and diagnosis. Comput Chem Eng 164:107884. https://doi.org/10.1016/j.compchemeng.2022.107884
2. Chen C, Reniers G (2020) Chemical industry in China: The current status, safety problems, and pathways for future sustainable development. Saf Sci 128:104741. https://doi.org/10.1016/j.ssci.2020.104741
3. Ewuzie RN, Gunasekaran S, Ahmad Z, Md Nor N (2025) Design and implementation of deep learning-based framework for multi-class fault diagnosis in complex chemical process systems. Eng Appl Artif Intell 162:112630. https://doi.org/10.1016/j.engappai.2025.112630
4. Hu W, Wang Y, Li Y, et al (2024) A multi-feature-based fault diagnosis method based on the weighted timeliness broad learning system. Process Saf Environ Prot 183:231–243. https://doi.org/10.1016/j.psep.2023.12.071
5. Neupane D, Bouadjenek MR, Dazeley R, Aryal S (2024) Data-driven Machinery Fault Detection: A Comprehensive Review. Neurocomputing 627:129588. https://doi.org/10.1016/j.neucom.2025.129588
6. Amin MT, Khan F, Ahmed S, Imtiaz S (2021) A data-driven Bayesian network learning method for process fault diagnosis. Process Saf Environ Prot 150:110–122. https://doi.org/10.1016/j.psep.2021.04.004