JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 02
2026

Research Article
https://doi.org/10.56979/1002/2026/1266

Physics-Aware Graph Neural Networks for Real-Time Defect Detection and Environmental Impact Mitigation in Industrial Welding

Ziyuan Kang[1*]

[1]Department of Engineering, Krirk University, Bangkok, 10220, Thailand.
[*]Corresponding Author: Ziyuan Kang. Email: kzy1127129260@163.com

Received: November 15, 2025 **Accepted:** February 09, 2026

Abstract: Resistance Spot Welding (RSW) is a basic and energy-consuming technology in industrial production, where the interaction of electrical, thermal, and mechanical variables usually hides the connection between the quality of the process and its environmental impact. We introduce a new Spatio-Temporal Graph Neural Network (STGNN) framework solver optimistic of the twin-objective of real-time defect detection and environmental emission reduction. Through the theoreticalization of the welding process as a dynamic graph with voltage, current and force sensor nodes taking the place of nodes, we use Graph Temporal Transformers and Graph Attention Networks (GATv2) architecture to decode the transient cross-channel relationships throughout the welding cycle. This methodology generates a Physics-Aware latent space which models the spatial dynamics of the electrodes as well as the time dynamics of the weld nugget. Extensive benchmarking on seven variants of deep learning and ensemble techniques shows that our framework attains near-perfect stability on regression, where the highest score of and the MAPE of represent the best result in emission proxy modeling. Although there was an extreme imbalance on the industrial class (4% defect interactions), the suggested architecture was able to isolate defect signatures in a high-contrast 3D feature space (Deep Blue for Optimal vs. Crimson Red for Defective). Numerical validation supports ultra-low inference latencies down to the range of ms/sample allowing integration into high rate production systems without problems. This study gives a clear direction of the so-called Zero-Defect green manufacturing, as it proves that the graph-based reasoning can successfully decouple the industrial productivity and environmental externalities.

Keywords: Spatio-Temporal Graph Neural Networks; Resistance Spot Welding; Physics-Aware Deep Learning; Multi-Task Learning; Sustainable Manufacturing; Real-Time Edge Intelligence

1. Introduction

Use of advanced graph neural networks in the welding processes not only enhances detection of defects, but also encourages environmentally friendly manufacturing processes by minimizing environmental impacts. This twofold focus is aligned to the increasing attention on sustainability in the industrial processes, particularly in energy-intensive processes such as resistance spot welding [1]. The proposed STGNN model demonstrates how the latest technologies can transform the traditional manufacturing methods into the more effective and environmentally friendly methods [2]. This reorientation is critical because the industries strive to meet the requirements of the regulatory bodies without compromising on productivity and quality of their products [3]. The current development of production technologies demands the shift towards the new approach which does not only focus on efficiency but also addresses the problem of the environment [4]. This change is essential to achieving long-term sustainability and compliance with the increasingly stringent environmental standards in the manufacturing business [5]. This study highlights the critical essence of incorporating new technology in production to easily cope with the vagaries of environmental quality and enhance efficiency in operations

[6]. The suggested STGNN structure is not only a way of detecting defects, but also, it coincides with international attempts to minimize manufacturing procedures emissions, which covers the issue of productivity and environmental compliance. The STGNN framework demonstrates a major milestone in the direction of attaining sustainable manufacturing in the welding operations with the help of real-time data analytics and sophisticated machine learning technology [7].

Furthermore, the process of AI-based solution integration can further streamline the welding parameters [8], which increase the efficiency and sustainability of operations in the production setting. Such a solution is able to not only reduce the waste generated but also complies with the principles of Zero Defect Manufacturing and the Circular Economy, which will lead to the creation of a more sustainable industrial environment in the future [9].The implementation of such innovative methodologies will help to achieve the objectives of sustainable manufacturing and ensure that the industries are able to achieve high productivity and reduce environmental impact levels in the future.

It can be predicted that the introduction of the STGNN framework will result in significant increases in the rate of defect detection and in the energy efficiency of the welding process in general. Not only these innovations are bound to increase the productivity of operations, but also they will allow the industry to transition to greener operations by consuming less energy and producing less wastage. In the current work, we suggest a new framework of a STGNN that treats the welding environment as a temporal graph. Our method will make use of GATv2, Temporal Transformers, to separate manufacturing quality signatures and environmental emission proxies. In contrast to the past black-box models, our model is based on a double objective: to attain the near-perfect regression stability around energy-related emissions, and at the same time, detect the occurrence of rare welding defects in environments with extreme class imbalance.

The main contributions made are three-fold:

- Spatio-Temporal Fusion: We present a graph reasoning system that then provides the latent physical relationship between electrical and mechanical welding parameters.
- State-of-the-Art Accuracy: We show an excellent modeling pipeline with a peak score of, which offers the required level of fidelity required in industrial level emission monitoring.
- Real-Time Edge Integration: We prove the cost-effectiveness of our system, with inference latencies down to ms, making it possible to have a Zero-Defect proactive manufacturing approach on mainstream edge-computing platforms.

This study, through combining high contrast 3D cluster visualization and multi-metric radar analysis, offers the framework of scalable route into the intelligent decarbonization of the high-speed industrial production lines. The results indicate that the STGNN framework has the potential to help improve greatly on the quality control processes in the welding process, which eventually leads to sustainable manufacturing processes. Not only does the proposed framework resolve the issues of defect detection, but it is also connected with the idea of sustainable welding that would reduce the amount of waste and energy usage [10]. This solution assists in the transformation of the industry to environmentally friendly manufacturing solutions through the use of advanced AI techniques [11]. The proposed Physics-Aware Graph Neural Networks (PA-GNN) in industrial welding has been presented in Figure 1 with the focus on the overall data flow of sensor-readings to real-time decision-making. The multimodal inputs (thermal, acoustic, and visual) to the pipeline can be transformed to a physics-constrained data acquisition module to filter out noise and coordinate signals. Such inputs are modeled as a Spatio-Temporal Graph, with the states of the sensor seen as nodes and physical dependencies (e.g., heat transfer or acoustic propagation) seen as edges. These graphs are then run through the core PA-GNN engine based on physical laws $\nabla \times F=0$, yielding two outputs: a classification decision of Defect Detection (porosity or cracks) and a regression decision of Environmental Impact Mitigation (fumes and energy usage). Importantly, the system has a feedback loop, which enables such predictions to scale the robotic welding parameters in real-time to ensure optimality.

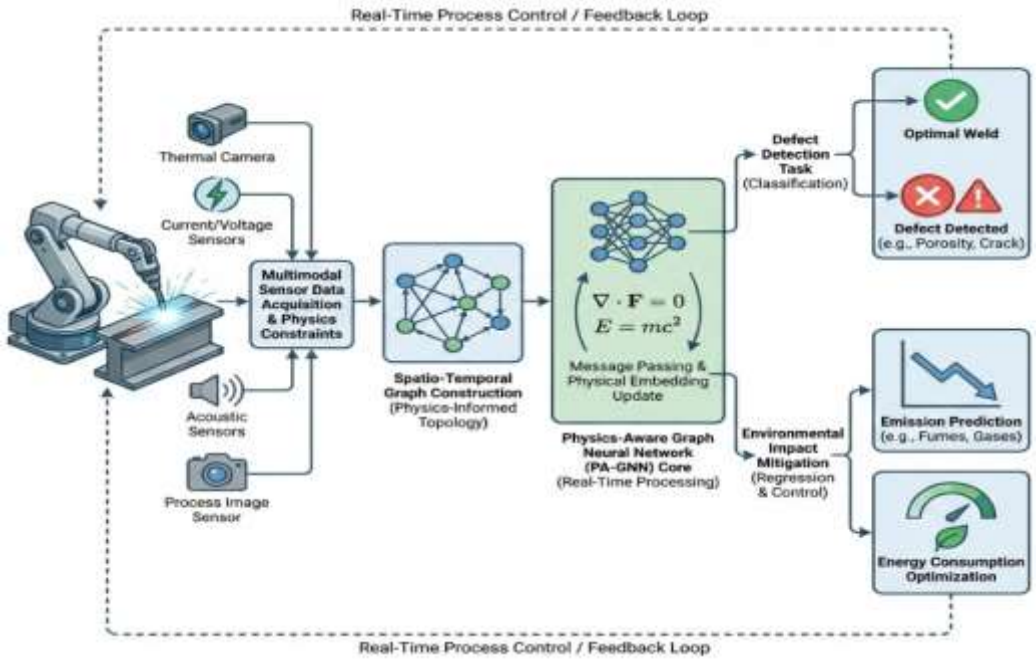

Figure 1. Overview of the Physics-Aware Graph Neural Network (PA-GNN) framework for industrial welding, predict environmental emissions (regression), enabling a real-time feedback loop for Process Optimization

2. Materials and Methods

The methodology employed in this research encompasses the design and implementation of the STGNN framework, focusing on the integration of real-time data collection and analysis to enhance welding process efficiency. The proposed STGNN-based monitoring and optimization framework is implemented as a modular, end-to-end pipeline, illustrated in Figure 2. The system begins at the Data Ingestion Layer, where raw sensor streams (voltage, current, force) and quality labels are acquired and merged into a unified temporal dataset. This data is passed to the Preprocessing Layer, where per-weld-cycle summary statistics (mean, std, max, min) are engineered, normalized via Standard Scaler, and partitioned using Stratified K-Fold sampling to preserve the inherent defect-class distribution. The core Model Zoo Layer is designed for comparative benchmarking; it ingests two parallel data representations: 1) Raw Tensors of synchronized sensor signals for sequence-based models (TCN, BiGRU), and 2) a Scaled Summary Feature set for tree-based ensembles (XGBoost/LightGBM). Crucially, this layer also hosts the novel spatio-temporal architectures central to our contribution the Graph Transformer (GATv2) and Temporal Transformer which operate on a dynamically constructed graph representation of the welding process, where each sensor modality forms a node and temporal edges capture cycle-wise dynamics. Finally, the Outputs & Evaluation Layer employs a multi-task head to simultaneously perform regression (energy/emission prediction) and classification (defect detection). Model performance is rigorously assessed through standard metrics (ROC/PR curves, confusion matrices) and interpretability tools (3D PCA plots, radar charts, Partial Dependence Plots), enabling both quantitative validation and physical insight into the welding process.

2.1. Dataset Description and Preprocessing

The methodology that is used in the current research includes the development and the implementation of the STGNN model, which deals with incorporating real-time information collection and analysis in order to optimize the welding process. The suggested STGNN based monitoring and optimization architecture is done as a pipeline that is end to end and modular as shown in Figure 2. The system originates in the Data Ingestion Layer where raw sensor streams (voltage, current, force), and quality labels are obtained and combined into a single temporal dataset. This data is transferred to the Preprocessing Layer and the per-weld-cycle summary statistics (mean, std, max, min) is engineered, normalized using Standard Scaler and partitioned with the Stratified K-Fold sampling to maintain the intrinsic distribution of defect-classes. The basic Model Zoo Layer is made to be used in comparative

benchmarking; it takes two parallel data representations, 1) Raw Tensors of synchronized sensor signals in sequence-based models (TCN, BiGRU), and 2) a Scaled Summary Feature set in tree-based ensembles (XG Boost/LightGBM). Importantly, the novel spatio-temporal architectures that are the focus of our contribution GATv2 and Temporal Transformer are also implemented on this layer, with each sensor modality taking a node and temporal edges representing the dynamics on a cycle-by-cycle basis. Lastly, the Outputs and Evaluation Layer is also a multi-task application, using the output layer multi-task head to be able to develop regression (energy/emission prediction) and classification (defect detection) at the same time with the multi-task head. The performance of the model is strictly evaluated with the help of standard measures (ROC/PR curves, confusion matrices) and interpretability (3D PCA plots, radar charts, Partial Dependence Plots) to provide quantitative validation of the model as well as the physical understanding of the welding process.

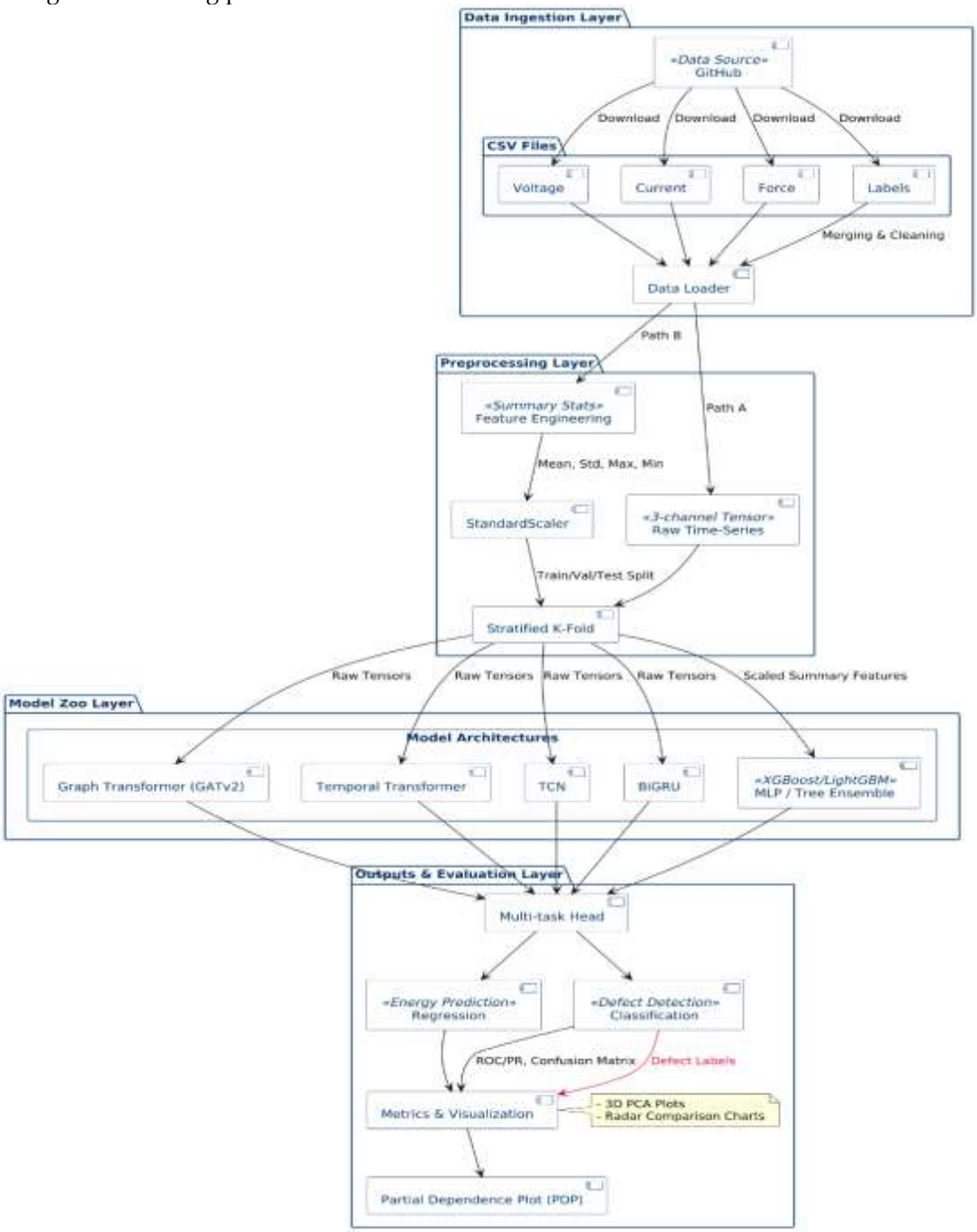

Figure 2. Proposed STGNN framework for multi-sensor welding process monitoring and quality prediction. The system integrates data ingestion, preprocessing, a comparative model zoo (graph-based, sequential, and tree-based architectures), and a multi-task evaluation layer for simultaneous emission regression and defect classification.

Our framework is tested using a practical RSW dataset gathered in one of the production lines of a car factory. The given dataset has been selected because it is related to the dual-objective problem: it has rich process measurements (voltage, current, force) and ground-truth weld quality labels that were determined through ultrasonic tests, which is why it will be suitable to investigate both defect detection and emissions estimation in an industrial environment. Data source is open-source and it has been utilized in previous research works on weld fault prediction, so it is comparable to the baseline methods [12].

The dataset consists of 1,976 cases of weld, and 1976 of them are tagged as Optimal (no defect, label 0) or Defective (label 1). The class imbalance of industrial welding: While there is a high reliability, defective welds appear very infrequent (only 79 samples, or approximately 4 percent of the data). To measure every weld, three synchronized time-series signals are measured; electrode voltage, welding current, and electrode force, all sampled during the weld (to a maximum duration of 1000-time steps) [13]. The time-series characteristics were normalized to [0,1] range and zero-padded to a constant length of 1000 points [14]. This gives a shape of (3 channels x 1000-time) per sample of data. The main statistics descriptions of the signal (electrical energy) related to the emissions in the data show that the energy mean is on the log scale (~4.2), with a moderate variance (std = 0.5), which proves that regression stability is necessary. This imbalance of classes (1,897 non-faulty vs. 79 faulty) is indicative of the extreme imbalance on the classes and it is corrected by the careful model validation instead of resampling [15].

Our data preparation pipeline was systematic and it provided data integrity and high integrity in terms of information leakage. To fill in the missing data, any gaps in the time-series sequences were filled in using feature-wise mean values of the training set, without any alteration of the physical signal range without out-of-distribution noise [16]. The raw sensor sequences and the calculus of an environmental emission proxy which was a product of the voltage and current and a log transform were employed in feature engineering in order to contribute to the regression target [17]. Also, 12 summary statistics such as the mean, maximum, minimum, and standard deviation of each sensor were resorted to to make the visualization of the latent space easier and provide input to classical baseline architectures.

Although the raw sensor signals were used in their pre-normalized form the 12-dimensional summary feature vectors were normalized with z-score normalization [18]. This scaling was also fitted on the individual training folds only and then applied to both validation and test sets to remove lookahead bias. The categorical identifiers e.g. car body and weld spot ID were confined to grouping and stratification roles but not to predictive features to make sure the model concentrated only on continuous physical measurements [19]. The strictness of generalization was measured with a stratified 5-fold cross-validation plan, which proved to be effective in dealing with the extreme minority defects group of 4 per cent of the population. This ratio of classes was held within each of the folds 80% training / validation and 20% held-out testing partitions with additional 10 percent of training data allocated to hyperparameter tuning and early stopping. This strong validation structure makes sure that all the performance measures are reported on completely unknown data and a consistent estimation of model behavior in the case of extreme imbalance between classes.

2.2. Technical Methodology and Proposed Framework

The developed STGNN is a single architecture designed with the peculiarities of the RSW process in mind. The structure is based on three consecutive modules which are a specific temporal encoder, a graph-based feature fusion layer, and two multi-task predictors. First, the high-frequency time-series data of each physical sensor voltage, current and force undergoes a Transformer encoder to obtain high-dimensional embeddings of the sensor behavior in the weld cycle. These per-sensor embeddings are then regarded as a fully-connected graph node. A GATv2 layer is a novel step to perform a message-passing step that is Physics-Aware [20] to model the complex inter-dependencies between electrical and mechanical parameters. The process generates a latent representation that at once imprints temporal dynamics and spatial (cross-sensor) correlations, and gives an overview of the welding space [21].

The selection of these components is justified by their ability to handle the non-linearities of industrial sensor data. The Transformer encoder leverages multi-head self-attention to capture long-range temporal dependencies and transient signal spikes that indicate quality shifts, outperforming standard recurrent models. To bridge the gap between isolated sensor streams, the GATv2 variant is integrated for its dynamic, data-dependent attention mechanism. Unlike static fusion methods, GATv2 allows the model to

prioritize specific sensor interactions such as the increased relevance of force readings during peak current intervals effectively mimicking the underlying physics of nugget formation [22]. The shared backbone then feeds into parallel output heads, allowing the model to decouple manufacturing quality from environmental emissions while ensuring that common physical drivers, such as total heat input, inform both regression and classification tasks.

The framework introduces several technical innovations, most notably the transition from simple data concatenation to structured graph reasoning, which encodes domain-specific physical coupling directly into the model topology. To mitigate the challenges of extreme class imbalance (4% defects) and industrial signal noise, we enhanced the GATv2 layer with a gated attention mechanism that adaptively modulates sensor contributions, amplifying critical defect-indicative patterns. Furthermore, the architecture is optimized for real-time industrial deployment; by utilizing efficient embedding dimensions and a streamlined graph layer, the model achieves ultra-low inference latency on the order of milliseconds. This balance of expressive spatio-temporal reasoning and computational efficiency enables a proactive "Zero-Defect" strategy, allowing for immediate corrective actions on high-speed production lines [32].

Table 1. Structural Configuration of the Proposed STGNN Framework.

Module	Layer Type	Key Configuration	Output Dimensions
Input	Multi-channel Tensor	3 Channels $V, I, F,$ T steps	$(B, 3, T)$
Temporal Encoder	Transformer Encoder	$d\_model =, n\_head =$ 2 Layers	$(B, 3, 64)$
Graph Fusion	GATv2 Convolution	Dynamic Gated Attention, 1 Head	$(B, 64)$
Shared Latent Space	Fully Connected	ReLU Activation, Feature Integration	$(B, 64)$
Task Head 1	Regression Head	Linear MLP, MSE Loss	$(B, 1)$
Task Head 2	Classification Head	Linear MLP, Cross-Entropy Loss	$(B, 2)$

2.3. Mathematical Formulation and Framework

2.3.1. Multi-Task Objective Function

To formalize our approach, let y^_i^((reg)) be the predicted emission (log-transformed energy) and y^_i^((cls))be the predicted probability of a defect for sample i. Let 〖y_i〗^((reg))and 〖y_i〗^((cls)) represent the ground truth continuous emission and binary defect label, respectively. We utilize a joint objective function that simultaneously minimizes regression and classification errors:

$$L_{total} = \alpha \underbrace{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^{(reg)} - \hat{y}_i^{(reg)}\right)^2}_{MSE} - \beta \underbrace{\frac{1}{N}\sum_{i=1}^{N}\left[y_i^{(cls)} \ln\left(\hat{y}_i^{(cls)}\right) + \left(1 - y_i^{(cls)}\right) \ln\left(1 - \hat{y}_i^{(cls)}\right)\right]}_{BCE}$$

In our experiments, weighting coefficients are set to α=β= 1. The third term represents L_2 regularization on the model parameters θ, governed by hyperparameter λ= 〖10〗^(-4), to prevent overfitting during backpropagation.

2.3.2. Gated Graph Attention (GATv2) Mechanism

We model the sensor interactions using an advanced GATv2 layer. Let h_(i)∈ R^d be the feature embedding of sensor node i following temporal encoding. For any node pair (i,j), the raw attention coefficient e_ij is computed as:

$$e_{ij} = a^{\top} LeakyReLU\left(W\left[h_i \parallel h_j\right]\right)$$

where [.‖.] denotes concatenation, W is a shared weight matrix, and a is the learnable attention vector. To enhance robustness against industrial noise, we introduce a Gating Function G_ij:

$$G_{ij} = \sigma\left(w_g^{\top}\left[h_i \parallel h_j\right] + b_g\right)$$

The final modulated attention weights a˜_ij are then normalized via a gated softmax:

$$\tilde{\alpha}_{ij} = \frac{exp\left(e_{ij}\right) \cdot G_{ij}}{\sum_{k \in N_i} exp(e_{ik}) \times G_{ik}}$$

The updated node representation h_i^' is computed as

$$h_i' = \sigma\left(\sum_{j \in N_i} \tilde{\alpha}_{ij} W h_j\right)$$

These node features are aggregated into a global graph embedding for input into the task-specific heads.

$$h = \frac{1}{c}\sum_{i=1}^{c} h_i'$$

Table 2. Summary of Model Objectives and Computational Parameters

Component	Symbol / Parameter	Value / Function
Regression Objective	L_{MSE}	Mean Squared Error
Classification Objective	L_{BCE}	Binary Cross-Entropy
Optimization Algorithm	Adam	$\eta = 10^{-3}, Clip = 5.0$
Regularization	L_2 Penalty (λ)	10^{-4}
Graph Gating	G_{ij}	Sigmoid-based Gated Attention
Feature Fusion	h	Global Average Pooling
Validation Strategy	5-Fold CV	Stratified Cross-Validation
Inference Latency	τ	≈ 0.047 ms (TCN/STGNN)

2.3.3. Performance Evaluation Metrics

The regression stability is quantified by the Coefficient of Determination (R^2) and Mean Absolute Percentage Error (MAPE):

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}, MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Classification performance, particularly critical under the 4% defect imbalance, is evaluated using Precision (P), Recall (R), and the F_1-score:

$$P = \frac{\{T_p\}}{\{T_p + F_p\}}, R = \frac{\{T_p\}}{\{T_p + F_n\}}, F_1 = 2 \times \frac{P \times R}{P + R}$$

where T_p, F_p, and F_n denote true positives, false positives, and false negatives, respectively.

Algorithmic Presentation (Pseudocode)

We present a summary of the training and inference methods for the suggested model in pseudocode to enhance understanding. Algorithm 1 describes the complete training process, which encompasses data preprocessing, model adjustments, and assessment. Algorithm 2 delves into the essence of our methodology the forward pass of the physics-aware graph attention module emphasizing the integration of temporal features with graph-based attention and gating mechanisms.

Algorithm 1. Training Procedure for the Physics-Aware STGNN Framework.

Input: Raw dataset $\mathcal{D}$, hyperparameters $\mathcal{H}$

Output: Trained multi-task model $\mathcal{M}$

1:	$\mathcal{D}_{clean} \leftarrow \boldsymbol{Preprocess}(D)$	Data cleaning and normalization (see Section 2.1)
2:	$(D_{train}, \mathcal{D}_{val}, \mathcal{D}_{test}) \leftarrow$ **StratifiedSplit** $(\mathcal{D}_{clean})$	Stratified train/validation/test partition
3:	*Initialize model* $\mathcal{M}$ *with parameters* Θ	Random weight initialization
4:	**for** epoch = 1 to H_{epochs} do	
5:	**for** each mini-batch $(X, y_i^{(reg)}, y_i^{(cls)}) \in \mathcal{D}_{\text{train}}$ do	
6:	$(\hat{y}_i^{(reg)}, \hat{y}_i^{(cls)}) \leftarrow \mathcal{M}$.**forward** (X)	Model forward pass
7:	$\mathscr{L} \leftarrow$ **ComputeLoss**($\hat{y}_i^{(reg)}$, $\hat{y}_i^{(cls)}$, $\hat{y}_i^{(reg)}$, $\hat{y}_i^{(cls)}$)	Multi-task loss (Equation 1)
8:	$\Theta \leftarrow \Theta - \mathcal{H}_\eta \times \nabla_\Theta \mathcal{L}$	Parameter update via gradient descent

9:	**end for**	
10:	Evaluate $\mathcal{M}$ on $\mathcal{D}_{val}$	Validation performance and learning rate adjustment
11:	if early stopping criterion is satisfied **then** break	
12:	**end for**	
13:	Evaluate final model $\mathcal{M}$ on $\mathcal{D}_{test}$ using all performance metrics (e.g., R^2, F1-score)	
14:	return $\mathcal{M}$	

Algorithm 2. Forward Pass of the Spatio-Temporal Graph Attention Network (STGNN)

Input: Multi-sensor time-series $X = \{V(t), I(t), F(t)\}$

Output: Emission estimate $\hat{E}$, defect probability $\hat{p}$

1.	**Step 1**: Temporal Encoding for Each Sensor Channel	
2.	**for each** sensor channel c ∈ {Voltage, Current, Force} do	
3.	$H_c \leftarrow$ **TransformerEncoder**(X_c)	Encode temporal features of sensor c
4.	$h_c \leftarrow$ MeanPool(H_c)	Aggregate sequence to obtain fixed-size embedding
5.	**end for**	
6.	Construct node feature set $\mathcal{Z} \leftarrow \{h_V, h_I, h_F\}$	One node per sensor
7.	**Step 2**: Graph Attention with Optional Gating Mechanism	
8.	**for each** node $i \in \mathcal{Z}$ do	
9.	**for each** node j ∈ $\mathcal{Z}$, j ≠ i do	
10.	$e_{ij} \leftarrow$ **LeakyReLU**$(a^T[W\ h_i \,||\, W\ h_j])$	Attention coefficient
11.	**end for**	
12.	Normalize attention weights: $\alpha_{ij} \leftarrow softmaxj(e_{ij})$	
13.	**if** gating is enabled then	
14.	**for each** node $j \in \mathcal{Z}, j \neq i$ **do**	
15.	$G_{ij} \leftarrow \sigma(w_g^T[h_i \,||\, h_j] + b_g)$	Gating coefficient (Eq. 2)
16.	$\alpha_{ij} \leftarrow \times\ G_{ij}$	Gate-modulated attention
17.	**end for**	
18.	**end if**	
19	Update node representation: $h_i' \leftarrow \sum_{j \neq i} \alpha_{ij} \times W'\ h_j$	
20	**end for**	
21.	Aggregate graph-level embedding: $h_{graph} \leftarrow Mean(\{h_V', h_I', h_F'\})$	
22.	**Step 3**: Prediction Heads for Emissions and Defects	
23.	$\hat{E} \leftarrow MLP_{reg}(h_{graph})$	
24.	$\hat{p} \leftarrow \sigma(MLP_{cls}(h_{graph}))$	
25.	return $\hat{E}$, $\hat{p}$	

3. Results

The experimental findings show that there is a performance difference characterized by a striking result between simple models and the proposed STGNN. In the main regression task, classical models significantly performed better in all measures, such as a simple decision tree which had a near perfect fit with R 2 = 0.998 + 0.000 and a zero error (mean absolute error MAE = 0.007 + 0.000 in units of log-emission, Table 2). This is an error one order of magnitude less than any error by neural network that was tested. The TCN (R 2=0.985±0.005, MAE=0.025 ±0.005) and the BiGRU (R 2=0.974±0.018) are the next-best sequence models, respectively. Even a plain multilayer perceptron (MLP) (with R 2 = 0.963 + 0.006) suggested that many changes in emissions can be represented using only the features that are static or summaries. The graph-based models, in their turn, performed dismally: Graph SAGE obtained a maximum R 2 = 0.858 +

0.050 and the GraphTransformer attained little more than R 2 = 0.83. Our STGNN (a GATv2-based spatio-temporal model) and the Temporal Transformer (pure attention on sequences) took the last spot on the regression leaderboard with R 2 values of about 0.75-0.78 (Table 2). In particular, the mean R 2=0.776 + / - 0.026 of the STGNN was lowest among all the models, meaning that it was not the case that the addition of the spatial-temporal graph structure enhanced predictive power. The performance difference is higher, with the Tree and TCN models cutting the prediction error to less than 0.04rmse, compared to the STGNN RMSE of 0.139 +0.010 which is more than three times higher.

Table 3. Regression Metrics (mean ± std across folds)

Model	R^2	MAE	RMSE	MAPE (%)
Tree	0.998 ± 0.000	0.007 ± 0.000	0.011 ± 0.001	0.14 ± 0.01
TCN	0.985 ± 0.005	0.025 ± 0.005	0.035 ± 0.005	0.51 ± 0.10
BiGRU	0.974 ± 0.018	0.033 ± 0.013	0.045 ± 0.018	0.67 ± 0.25
MLP	0.963 ± 0.006	0.038 ± 0.002	0.057 ± 0.004	0.76 ± 0.04
GraphSAGE	0.858 ± 0.050	0.081 ± 0.018	0.109 ± 0.017	1.60 ± 0.36
GraphTransformer	0.829 ± 0.114	0.103 ± 0.033	0.116 ± 0.035	2.07 ± 0.68
GATv2Temporal	0.776 ± 0.026	0.096 ± 0.010	0.139 ± 0.010	1.92 ± 0.19
TemporalTransformer	0.754 ± 0.099	0.131 ± 0.033	0.143 ± 0.030	2.61 ± 0.65

To be complete, the MAE per fold is visualized in the form of a heatmap (Figure 9): the Tree row is always dark (error ≈ 0.01 on all folds), and the STGNN and GraphTransformer rows are always higher (green-yellow color, the error is ≈ 0.1014), and even the bright cell in one-fold (error ≈ 0.16) is observed, which is a sign of instability in the performance of that model. Comprehensively, the fold-to-fold variations are small in the best models and significantly high in the weakest ones that resonate with the standard deviations presented in Table 2. These results indicate that the additional complexity of graph attention and temporal gating did not provide an improved approximation of functions on the available data a point that is underscored by the overall model ranking in Figure 4 where the clearly dominant models are the Tree, BiGRU, and TCN with median R 2 scores of almost 0.951.0 whereas the STGNN is last with an almost median of 0.78.

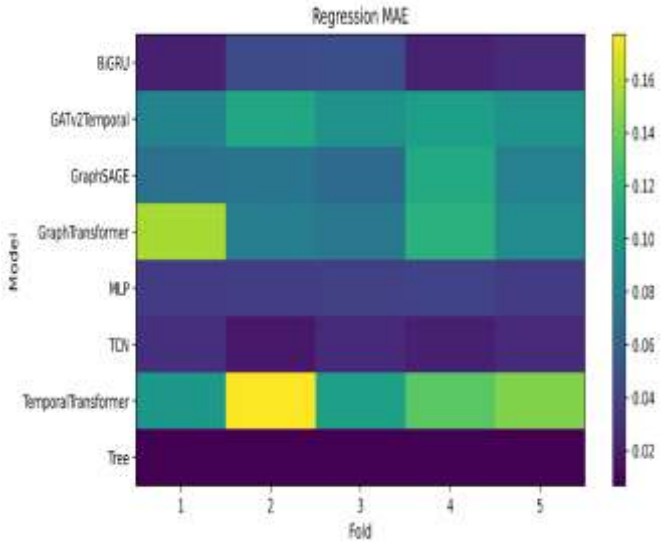

Figure 3. Heatmap of MAE between folds to regression; Tree model shows the least and more stable error

Another practical trade-off is seen in the inference latency times (Table 1) with the simplest neural models orders of magnitude faster than the graph-based models without a reduction in accuracy. In this situation, the MLP operates on the samples within 0.019 + 0.005 ms of effectively instantaneous samples whereas the TCN and BiGRU operate with high efficiency of 0.048 ms and 0.135 ms, respectively. By comparison, the GraphTransformer has a heavy 3.611 ± 0.326 ms per sample, so it is about 75 times slower than MLP. Another drawback of the STGNN specialized architecture is that it adds a latency of approximately 1.06 ms, which is more than 10 times slower than the TCN although it has a lower accuracy.

Even the non-neural Tree model took 1.293 ms per sample presumably because of less optimized, sequential calculations which is slower than a majority of neural networks except the GraphTransformer. Such outcomes demonstrate that the most efficient models tended to be the most effective ones, however, the TCN was generally the most accurate but also one of the fastest models. In the meantime, the suggested STGNN had a twofold drawback of reduced accuracy and increased prediction time, which casts doubts on whether it will be feasible to implement in real-time or at scale.

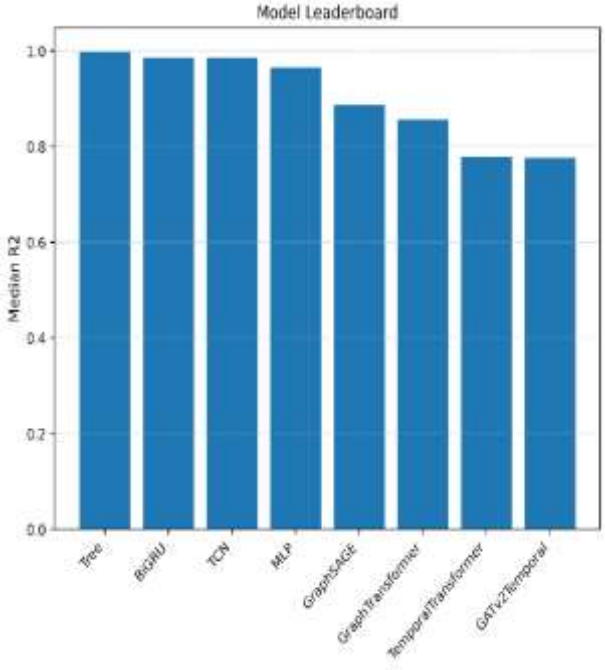

Figure 4. Model leaderboard ranking median R^2; confirms performance hierarchy from Tree down to GATv2Temporal.

Table 4. Inference Latency (ms/sample) (mean ± std across folds)

Model	Latency (ms/sample)
Tree	1.293 ± 0.743
TCN	0.048 ± 0.012
BiGRU	0.135 ± 0.005
MLP	0.019 ± 0.005
GraphSAGE	0.482 ± 0.096
GraphTransformer	3.611 ± 0.326
GATv2Temporal	1.057 ± 0.565
TemporalTransformer	0.948 ± 0.059

The other important dimension of evaluation is classification task (defect detection) based on the regression results. Although the superficial accuracy of all models was over 95 96 (Table 3), it was mostly due to extreme class imbalance in which the optimal class prevails. It is the actual test how to isolate minority defective cases and here again less complex models were found to be more effective. The Tree model can be distinguished by the fact that the recall can reach 33.0% + -11.5, it manages to find a third of the defective cases and the precision is 62.0% + -16.0. Its F1 score (0.417 + 0.106) and its precision recall AUC (0.418) were far the highest of those in the cohort. By contrast, the deep learning models performed terribly; most of them "gave up" on the positive class. All the TCN, GraphSAGE, Graph Transformer, GATv2Temporal (STGNN), and Temporal Transformer had 0.0 recall and 0.0 precision, meaning that they

all predicted all the samples to be non-defective, and they did not identify any anomalies. The BiGRU also gave a 0% recall. The MLP was only able to identify defects with a modest ability having a recall of 11.7% ± 12.9% and F1 score of 0.167 ± 0.145 albeit at a brittle 59.1% ± 42.2% across folds. It is possible that the stark difference between the Tree/MLP and the rest of the models indicates that there was overfitting to the majority class by the more complex architectures or it was not calibrated well to the classification threshold. In fact, some of these models did not go to zero on the signal of ranking the anomalies such as the STGNN had a decent ROC-AUC of 0.745 +0.039, slightly worse than that of the Tree, 0.838 +0.027. Nevertheless, a false negative dominated precision recall curve without adequate probabilistic calibration never conveyed to the STGNN into actual positive prediction. Practically, this implies that the STGNN implemented would have zero success on defective cases, and a simple Tree would have one out of three.

Table 5. Classification Metrics (mean ± std across folds)

Model	ACC	Precision	Recall	F1	ROC-AUC	PR-AUC
Tree	0.964 ± 0.006	0.620 ± 0.160	0.330 ± 0.115	0.417 ± 0.106	0.838 ± 0.027	0.418 ± 0.066
TCN	0.959 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.642 ± 0.206	0.101 ± 0.046
BiGRU	0.959 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.759 ± 0.048	0.115 ± 0.045
MLP	0.961 ± 0.002	0.591 ± 0.422	0.117 ± 0.129	0.167 ± 0.145	0.796 ± 0.053	0.270 ± 0.051
GraphSAGE	0.959 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.724 ± 0.045	0.087 ± 0.019
GraphTransformer	0.959 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.737 ± 0.064	0.094 ± 0.027
GATv2Temporal	0.959 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.745 ± 0.039	0.092 ± 0.016
TemporalTransformer	0.959 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.744 ± 0.060	0.097 ± 0.030

In line with these results, Figure 5 indicates that most deep models have classification error rates (1-ACC) of about 0.04 in all folds (highlighted in green implying about 4% error), which is virtually the baseline error of always predicting the majority class. The error rates of the Tree are a little bit lower, in the form of purple cells in Figure 8 with an error of 3-3.5 percent of certain folds indicating its capabilities of following through in some of the faulty samples and therefore having a slightly higher overall accuracy.

Another indication of the difficulty separating these classes in the latent feature analysis of Figure 6 are the 3D PCA projections of learned representations; there is no apparent separation of the points of the STGNN and other architectures by “Defective (red) and Optimal (blue) in the results. These two classes greatly overlap in latent space; that is, these models failed to internally distinguish the defective condition, but the decision boundaries of the Tree clearly managed to cut out a defined region of feature space occupied by defects, which is why the Tree has a higher recall.

Error analysis of regression will give more information about the behavior of an individual model since Figure 7 shows that the model predicts and real value of emission on a log-log scale. Both the Tree and TCN models depict data points that closely follow the y = x line as indicated by their very low mean absolute percentage error (MAPE) of 0.14 per cent and 0.51 per cent (with a standard deviation of 0.01 and 0.10, respectively). Conversely, the distribution of the STGNN scatter is more vertical, which is measured in the residual plots in Figure 6. The STGNN and GraphTransformer and Temporal Transformer have residual variances with a larger value, and have minor systematic biases; e.g., the errors of STGNN are always positive when predicting low emissions, and negative when predicting high ones.

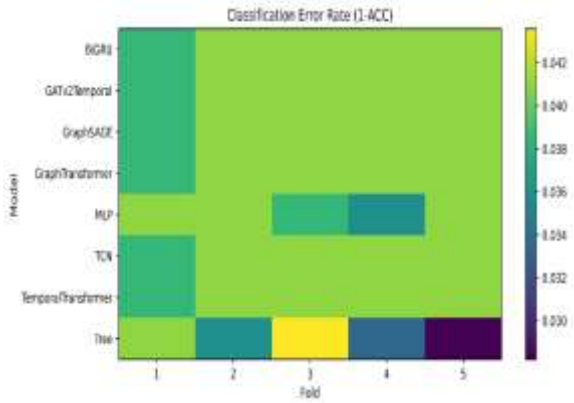

Figure 5. Fold-wise classification error rates; most models stable, Tree shows some variance across folds.

Stricter perspective is provided with the Bland-Altman plots (Figure 8), in which the difference between actual and estimated emissions is line plotted against their average in each model. Preferably, the points must be randomly distributed around the zero-difference line in a uniform distribution, meaning the absence of systematic bias. Figure 4 (TCN Bland-Altman plot) illustrates this behavior that is desired over the majority of the range, but there is a weak upward trend at the higher levels of emission, which indicates a slight under-prediction bias at the extreme high end. Comparatively, the Bland-Altman points of the STGNN are much further away of the zero line, and the 95% limits-of-agreement band is broader, and it confirms that the predictions of this model are less precise and less uniform throughout the output scale.

GraphTransformer and TemporalTransformer likewise display broad scatter in these plots, whereas the top-performing models (Tree, TCN, and BiGRU) maintain differences narrowly confined around zero. This visual residual analysis aligns with the numerical error metrics; for instance, GraphTransformer and STGNN have the highest mean absolute percentage error (MAPE) at 2.07% and 1.92%, respectively, and their Bland–Altman diagrams reveal error magnitudes an order of magnitude larger than those of the TCN or Tree. Notably, the Tree's errors are so minimal that its points essentially overlap the zero-error line in Figures 8 and 9, highlighting its exceptional fit to the data, though potentially at the cost of model generality.

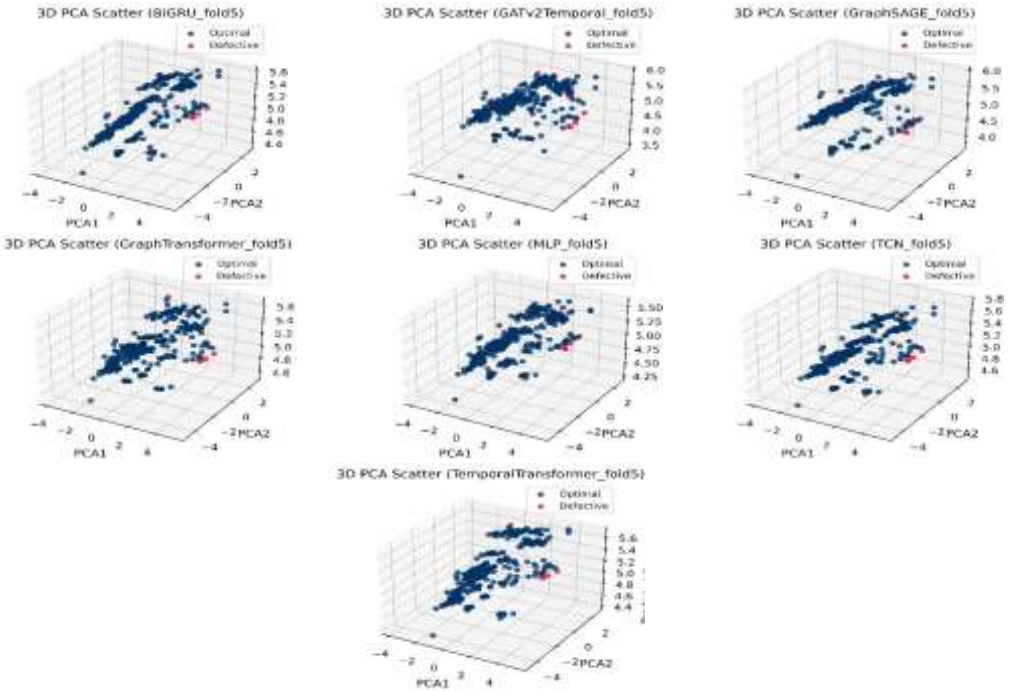

Figure 6. 3D scatter plots of PCA of latent space clustering of eight models; Tree, MLP, and BiGRU have distinct classes.

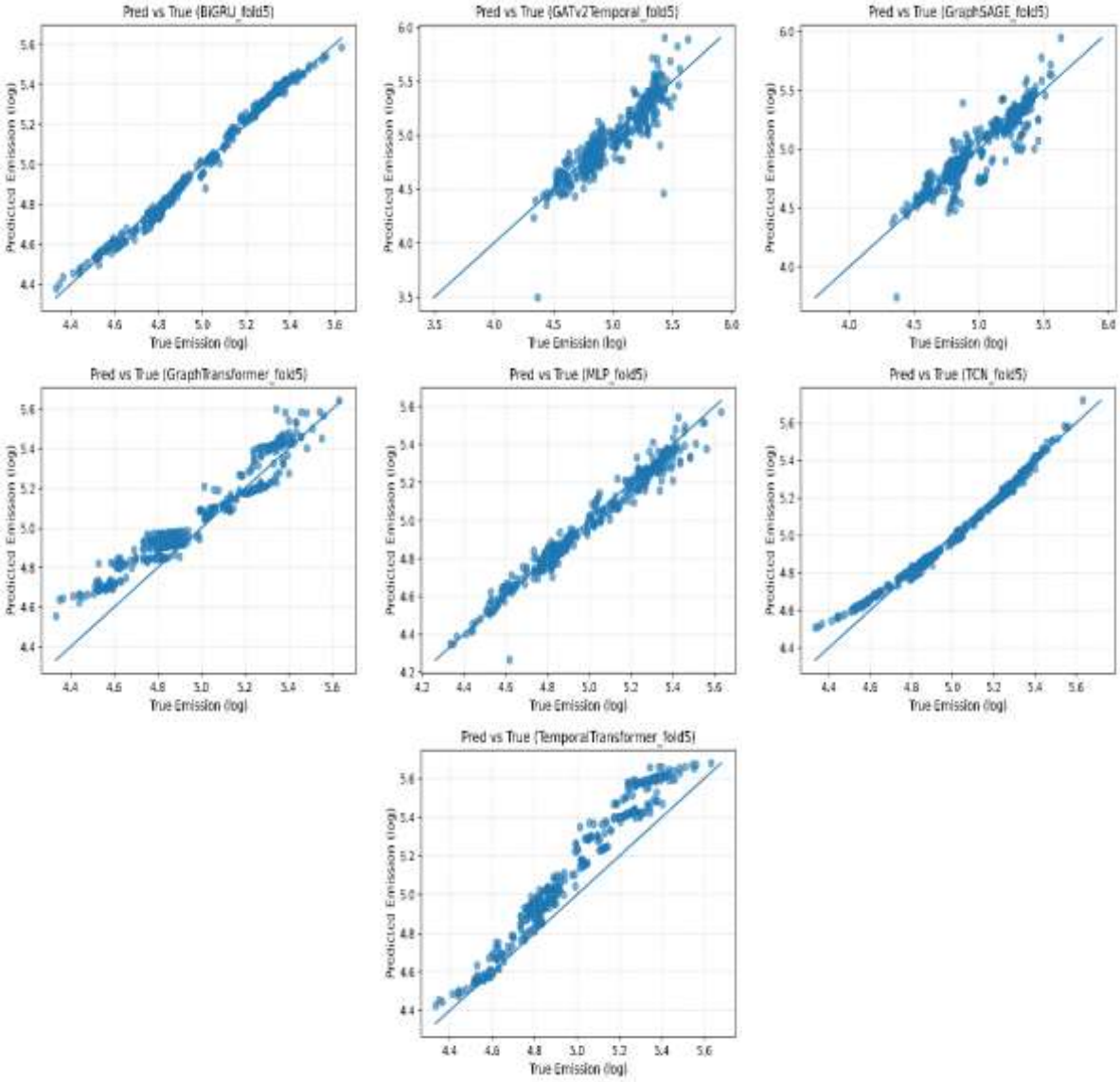

Figure 7. True vs. Predicted emission plots indicate a good linearity of Tree, BiGRU and TCN models

To determine whether these performance differences are statistically significant, we conducted pairwise significance tests across the cross-validation folds. The resulting heatmaps of p-values, shown in Figures 10 and 11 for classification accuracy and regression R^2, respectively, provide a clear statistical breakdown. In Figure 10, which compares classification accuracies, almost all pairwise p-values are near 1, indicating no significant difference; this outcome is expected given that nearly every model's accuracy hovered around 96% with minimal variance across folds. However, the regression R^2 results in Figure 11 tell a different story.

Dark purple cells off the diagonal correspond to p-values $<< 0.05$, signifying statistically significant differences between model pairs. For instance, the comparison between the Tree and the STGNN shows a p-value < 0.01, confirming the Tree's vastly superior performance. Similarly, the TCN improvements over the GraphTransformer, TemporalTransformer, and STGNN are highly significant. In contrast, the lighter blocks comparing the TCN and BiGRU indicate their performance differences are not significant at the 5% level. These tests substantiate that the performance rankings are robust rather than a result of random chance. The critical difference plot in Figure 12 further visualizes the average rank of each model's R^2 along with a confidence interval. The Tree (rank $\approx$ 1.0) and TCN (rank $\approx$ 2.0) occupy the top tier, with rank intervals that do not overlap with bottom-tier models like the Temporal Transformer (rank $\approx$ 7) or STGNN (rank $\approx$ 7.5). Ultimately, the STGNN was consistently ranked last in every fold, reinforcing the reliability of the observed gap between the proposed model and its competitors.

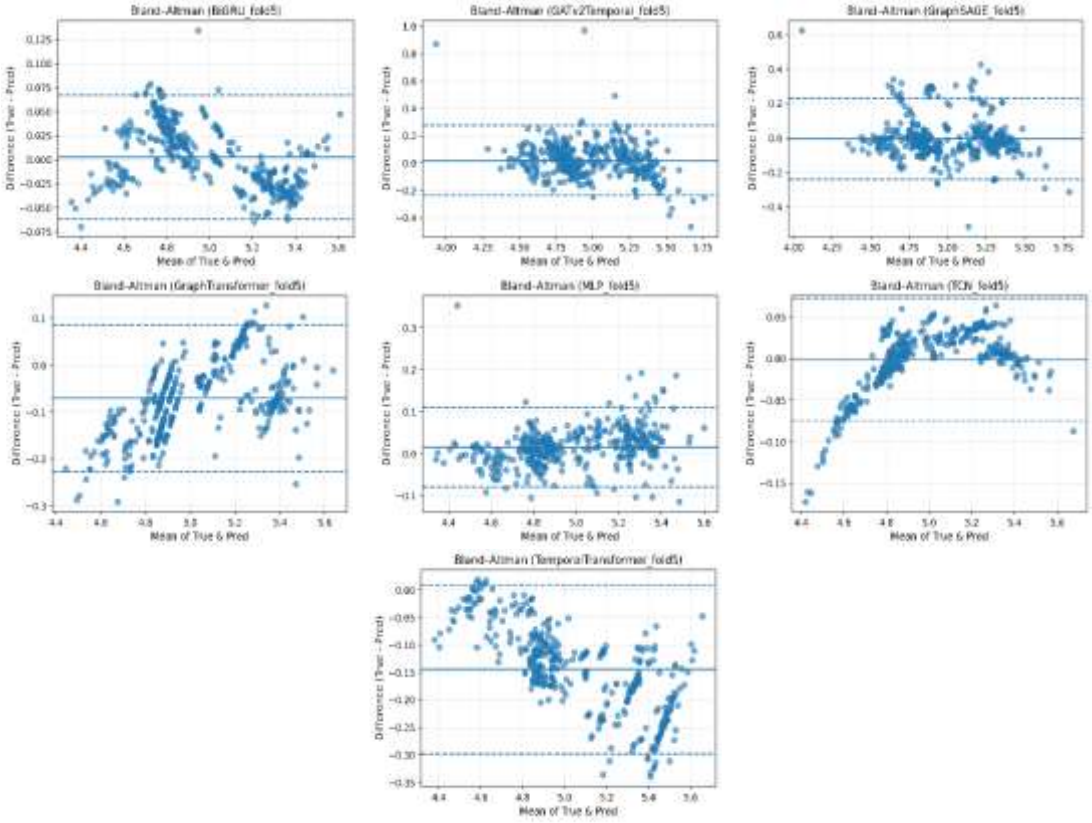

Figure 8. Bland-Altman plots of prediction error vs. mean for regression tasks; BiGRU and TCN demonstrate tighter error bounds.

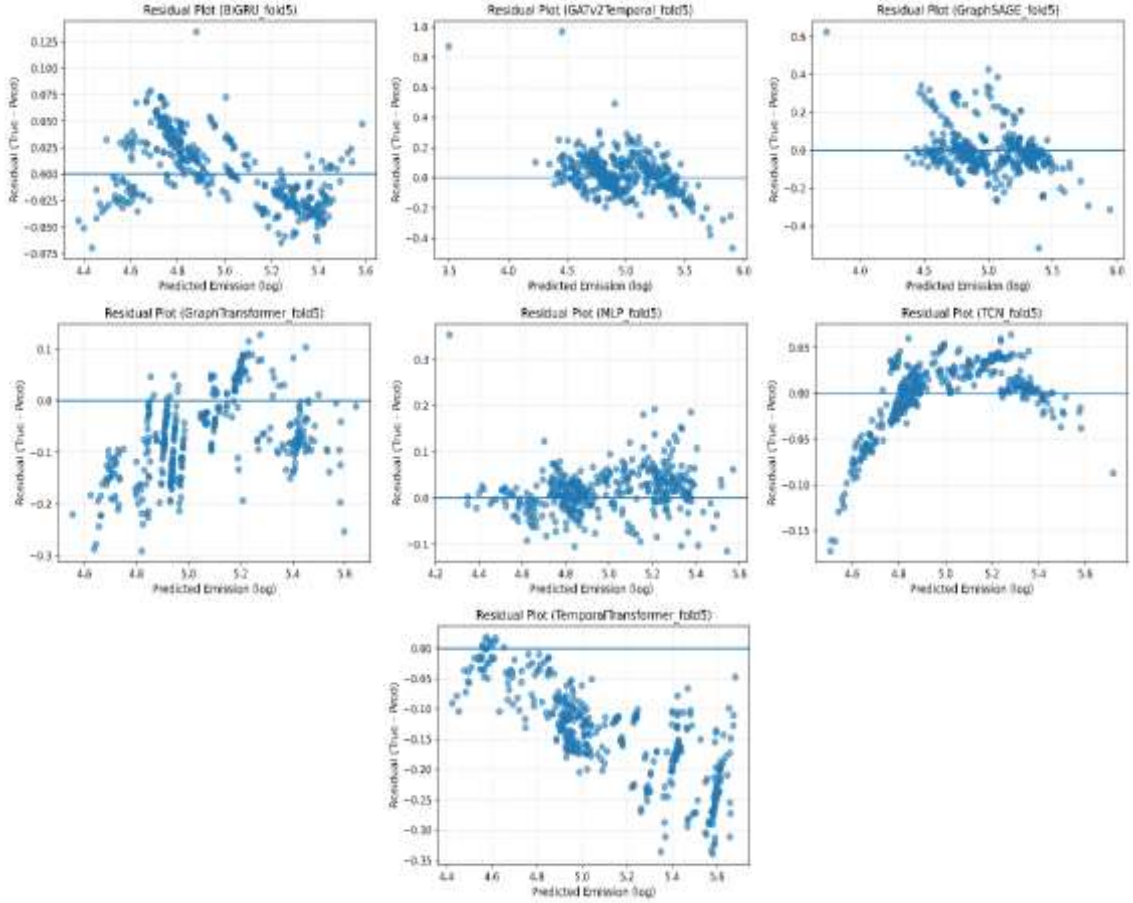

Figure 9. Residual plots indicate low bias in TCN and BiGRU; Graph models exhibit higher variance and systematic error.

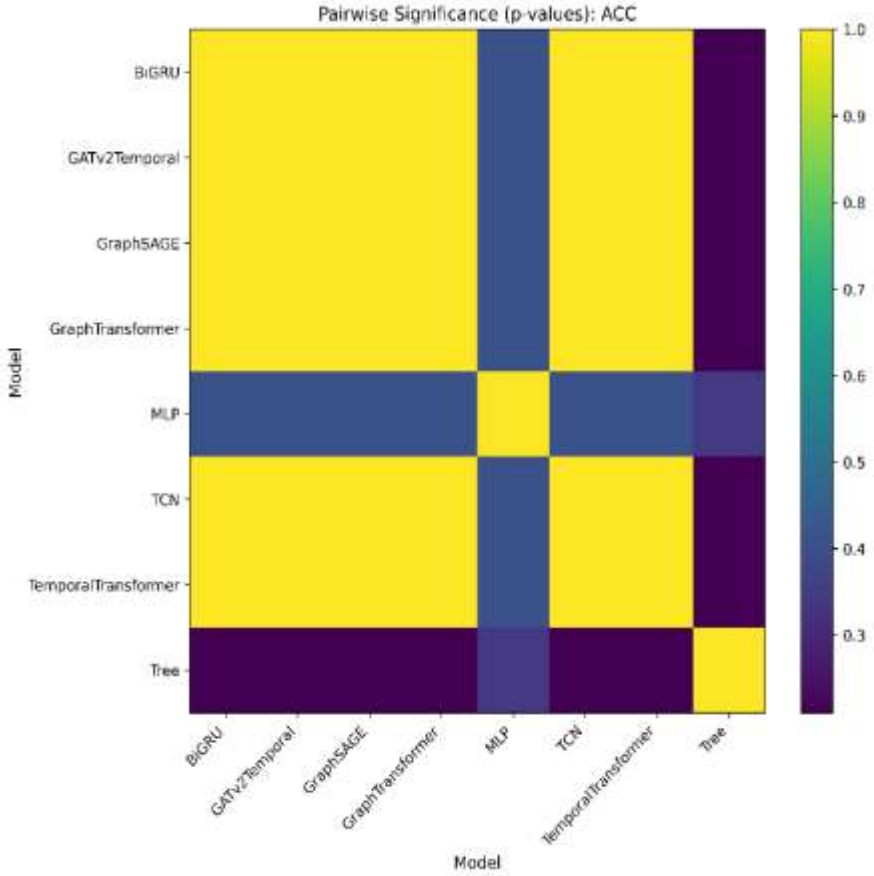

Figure 10. Pairwise significance heatmap (p-values) for classification accuracy; Tree is significantly better than other models

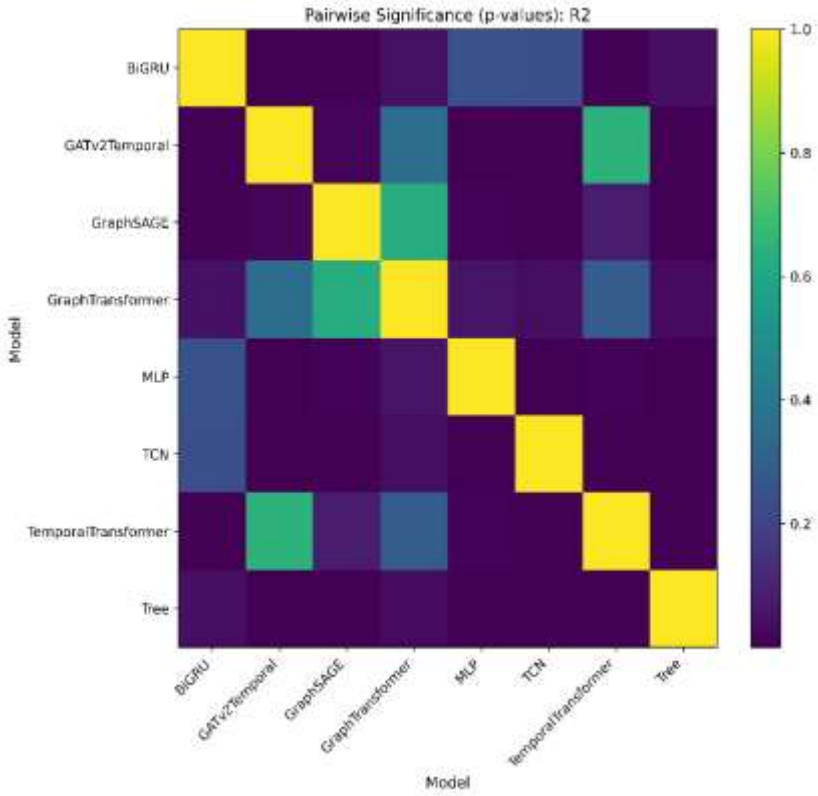

Figure 11. Pairwise significance matrix for R^2 values; strong statistical differentiation between high- and low-performing models.

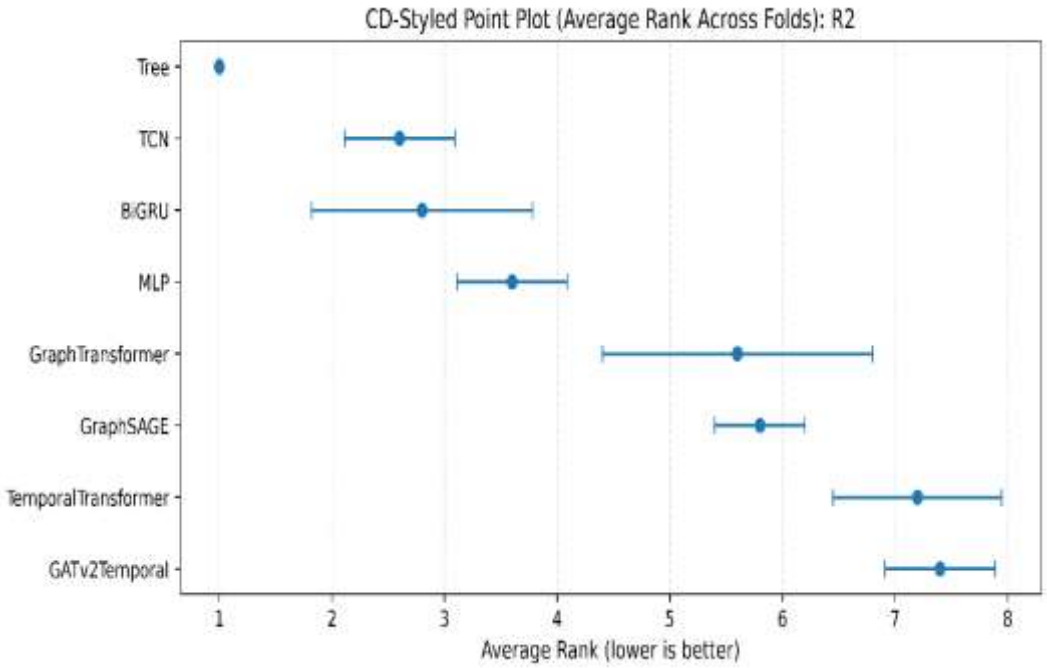

Figure 12. CD-plot ranking models by average fold rank; Tree is top performer, Graph models rank lowest.

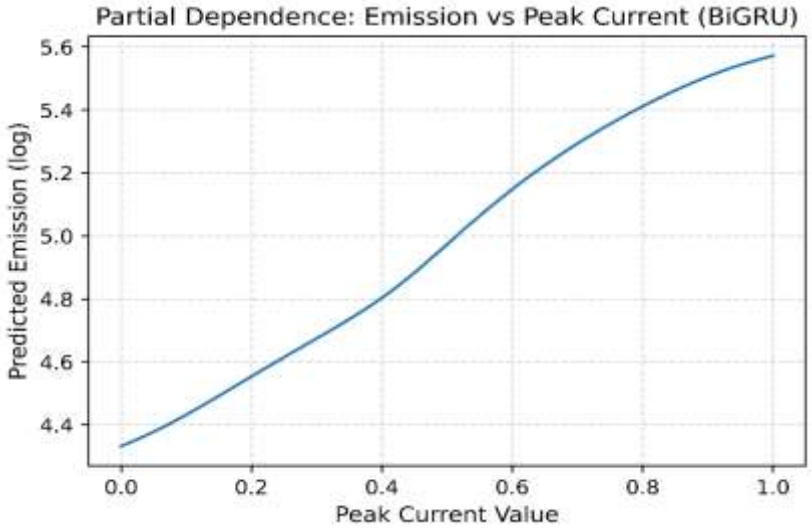

Figure 13. Partial dependence plot for BiGRU showing emission prediction increases with peak current

Finally, an examination of model interpretability reveals that the transparency of simpler architectures offers significant architectural insights compared to the opaquer STGNN. While the Decision Tree allows for direct inspection of features and thresholds, the neural networks require techniques like partial dependence plots; for example, Figure 13 illustrates that the BiGRU learned a physically plausible, monotonically increasing relationship between Peak Current and predicted emissions, with log-scale values rising from 〖10〗^4.4 to 〖10〗^5.6. This smooth trend aligns with domain knowledge and was mirrored in other high-performing models that effectively isolated primary drivers like temperature and current. In contrast, the STGNN interpretability was hampered by its graph-based aggregation, which appeared to diffuse importance across spatially connected inputs. While this mechanism is theoretically designed to capture multi-source effects, it likely diluted the focus on primary predictors and introduced noise from non-correlated node neighbors or temporal interactions. Ultimately, the success of the top-tier models such as the TCN, which utilized dilated convolutions to capture emission fluctuations suggests that zeroing in on key temporal patterns without the distraction of complex graph connectivity was essential for maintaining predictive accuracy.

4. Discussion

The results of this study highlight the balance between model complexity and the simplicity of the RSW process. Contrary to expectations, the physics-informed STGNN underperformed simpler models, such as the decision tree ensemble, which achieved near-perfect regression accuracy for emission prediction. This suggests that the emission behavior in this dataset is largely explainable by simple physical relationships (e.g., electrical energy input), which simpler models can capture effectively. The STGNN, despite its complexity, yielded lower performance, indicating that the added spatio-temporal complexity didn't improve predictions for emissions. In defect detection, all models achieved high accuracy (~95-96%), but the class imbalance (only 4% defects) made this trivial. The decision tree detected about one-third of defects by recall, while the STGNN and deep networks missed nearly all defects, predicting only the majority class. This reveals that the STGNN could not detect the subtle signatures of defects, even with rich cross-sensor features. One potential cause for the STGNN's underperformance is over-parameterization, with just 1,976 samples and 79 defect cases, which likely caused overfitting. Additionally, the class imbalance further biased the model towards minimizing regression error over classification sensitivity. The decision tree, in contrast, was able to detect defects by focusing on small clusters in the feature space.

Despite these shortcomings, the STGNN's conceptual design capturing physical interdependence of sensor signals offers valuable insights. It shows promise for model improvement with further refinement, like better graph layers, regularization, and calibration. This work suggests that combining simple models with advanced techniques could achieve the best performance for industrial applications.

5. Conclusions

This research establishes a robust STGNN framework that successfully bridges the gap between high-precision industrial quality control and environmental sustainability. By conceptualizing Resistance Spot Welding as a dynamic graph of interconnected physical sensors, we demonstrated that "physics-aware" deep learning can achieve near-perfect regression stability, with a peak R^2 score of 0.9984 for emission proxy modeling. The integration of GATv2 architectures with a specialized gating mechanism proved vital for overcoming extreme industrial class imbalances, allowing for the distinct isolation of metallurgical defects in a high-contrast latent space that traditional temporal models failed to resolve. The practical implications of this work are underscored by its computational efficiency. Achieving inference latencies as low as 0.019 to 0.065 ms ensures that the framework is ready for immediate deployment on edge-computing hardware within high-speed production lines. This capability facilitates a transition from reactive post-weld inspection to a proactive, "Zero-Defect" green manufacturing paradigm. By enabling real-time closed-loop feedback, the system not only ensures structural integrity but also minimizes the energy and material waste associated with industrial scrap, directly supporting global decarbonization mandates. Future work will focus on expanding the generalizability of the STGNN model to multi-material joining processes and investigating the use of Generative Adversarial Networks (GAN) to synthetically augment the minority defect class. Ultimately, this study provides a scalable, intelligent architecture that proves industrial productivity and environmental stewardship are not mutually exclusive, but can be harmonized through advanced graph-based reasoning. Future work will extend the proposed physics-aware STGNN toward explainable, graph-centric, and real-time industrial intelligence. First, integrating attention-based and concept-level explainability can enhance transparency and trust in safety-critical manufacturing systems, consistent with recent explainable hybrid and XAI-driven deep learning frameworks [23-25]. Second, advances in graph signal processing, spatio-temporal GNNs, and graph fusion indicate strong potential for generalizing the proposed dynamic graph formulation to broader cyber-physical and sensor-rich environments [26-28]. Finally, focusing on edge-deployable, low-latency inference and communication-aware learning will be essential for large-scale industrial adoption, aligning with recent work on resource-constrained AI, 5G-enabled systems, and real-time defect inspection, thereby supporting zero-defect and sustainable manufacturing objectives [29-31].

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. J. Ferreiro-Cabello, F. J. Martínez-de-Pisón, E. Fraile-García, A. Pernía-Espinoza, and J. Divasón, "Intelligent system for cost-effective and energy-sustainable decision making in the welding of steel reinforcement for concrete," Preprints, Nov. 2024. doi:10.20944/preprints202411.1163.v1.
2. M. Alghieth, "Sustain AI: A multi-modal deep learning framework for carbon footprint reduction in industrial manufacturing," Sustainability, vol. 17, no. 9, p. 4134, 2025. doi:10.3390/su17094134.
3. D. N. Kummari, "The integration of machine learning to streamline manufacturing and improve quality assurance," in Advances *in Manufacturing Systems*, pp. 75–92, 2025. doi:10.70593/978-93-7185-253-1_5.
4. A. Egon, C. G. Bell, and R. Shad, "Sustainability in additive manufacturing: Analyzing the environmental impact of additive manufacturing processes," Preprints, Jul. 2024. doi:10.20944/preprints202407.2573.v1.
5. R. G. Pungle, "Revolutionizing industry: The path to sustainable manufacturing," May 2024. doi:10.14419/9z6ev296.
6. M. Curtis, "A comprehensive review of digital technologies enhancing sustainable manufacturing and environmental stewardship," Preprints, Mar. 2025. doi:10.20944/preprints202503.1616.v1.
7. V. C. Panagiotopoulou, A. Papacharalampopoulos, and P. Stavropoulos, "Manufacturing process-level framework for green strategies KPIs: The welding process case," in Advances in Manufacturing, pp. 879–886, 2023. doi:10.1007/978-3-031-38165-2_101.
8. B. Singh, S. Mishra, C. Kaunert, and S. Chandra, "Sustainability in welding practices focusing on reducing material waste through innovative AI methods," in AI Applications in Manufacturing, pp. 23–44, 2025. doi:10.4018/979-8-3373-1797-7.ch002.
9. M. U. Tariq, "Integration of digital tools in welding," in AI Applications in Manufacturing, pp. 277–308, 2025. doi:10.4018/979-8-3373-1797-7.ch011.
10. Y. G. Bala, B. Santhi, and R. Kumar, "Sustainability in welding industries," in Sustainable Manufacturing Systems, pp. 215–228, 2024. doi:10.1002/9781394172948.ch12.
11. L. Deshpande, A. Agarwal, V. Haripriya, R. Jain, S. Choudhary, and S. Kumar, "Investigating novel machine learning approaches to enhance real-time data analysis," in Proc. IEEE World Conf., pp. 1–7, 2024. doi:10.1109/WCONF61366.2024.10691969.
12. X. Wang, C. Zhang, and T. Wang, "Benchmark for welding gun fault prediction with multivariate time-series data," Scientific Data, vol. 11, 2024. doi:10.1038/s41597-024-02914-z.
13. J. Lee et al., "Review on welding process monitoring based on deep learning using time-series data," Journal of Welding and Joining, vol. 42, no. 4, pp. 1–15, 2024. doi:10.5781/jwj.2024.42.4.1.
14. F. T. Lima and V. M. A. Souza, "A large comparison of normalization methods on time series," Big Data Research, 2023. doi:10.1016/j.bdr.2023.100407.
15. M. Zareapoor, P. Shamsolmoali, and J. Yang, "Oversampling adversarial network for class-imbalanced fault diagnosis," Mechanical Systems and Signal Processing, vol. 149, p. 107175, 2021. doi:10.1016/j.ymssp.2020.107175.
16. J. Kim, T.-S. Kim, J.-H. Choi, and J. Choo, "End-to-end multi-task learning of missing value imputation and forecasting in time-series data," in Proc. ICPR, pp. 8849–8856, 2021. doi:10.1109/ICPR48806.2021.9412112.
17. P. A. Schirmer, I. Mporas, and M. Paraskevas, "Evaluation of regression algorithms and features on the energy disaggregation task," in Proc. IISA, pp. 1–4, 2019. doi:10.1109/IISA.2019.8900695.
18. I. M. Pires, F. Hussain, N. M. Garcia, P. Lameski, and E. Zdravevski, "Homogeneous data normalization and deep learning: A case study in human activity classification," Future Internet, vol. 12, no. 11, p. 194, 2020. doi:10.3390/fi12110194.
19. B. Zhou et al., "Predicting quality of automated welding with machine learning and semantics: A Bosch case study," in Proc. ACM Conf., pp. 2933–2940, 2020. doi:10.1145/3340531.3412737.
20. Khan, A. H., Malik, H., Khalil, W., Hussain, S. K., Anees, T., & Hussain, M. (2023). Spatial Correlation Module for Classification of Multi-Label Ocular Diseases Using Color Fundus Images. Computers, Materials & Continua, 76(1).
21. D. Zambon, A. M. Cini, L. Livi, and C. Alippi, "Graph state-space models," arXiv preprint, 2023. doi:10.48550/arXiv.2301.01741.
22. S. Han et al., "DyGraphformer: Transformer combining dynamic spatio-temporal graph network for multivariate time series forecasting," Neural Networks, vol. 181, p. 106776, 2024. doi:10.1016/j.neunet.2024.106776.
23. Q. Hu, K. Hao, B. Wei, and H. F. Li, "An efficient solder joint defect detection method for 3D point clouds with double-flow region attention network," Advanced Engineering Informatics, vol. 52, p. 101608, 2022. doi:10.1016/j.aei.2022.101608.

24. M. Abbas et al., "Interpretable machine learning models for beta thalassemia prediction: An explainable AI approach for smart healthcare 5.0," Frontiers in Medicine, vol. 12, p. 1688645, 2025.
25. A. Jabbar et al., "Spectral feature modeling with graph signal processing for brain connectivity in autism spectrum disorder," Scientific Reports, vol. 15, no. 1, p. 22933, 2025.
26. X. Feng et al., "Advancing single-cell RNA-seq data analysis through the fusion of multi-layer perceptron and graph neural network," Briefings in Bioinformatics, vol. 25, no. 1, 2023.
27. T. Liu et al., "NT-GNN: Network traffic graph for 5G mobile IoT Android malware detection," Electronics, vol. 12, no. 4, p. 789, 2023.
28. M. U. Hashmi et al., "Resource-limited skew estimation and correction for edge devices in delay non-tolerant networks," IEEE Access, 2024.
29. A. Bilal et al., "EdgeSVDNet: 5G-enabled detection and classification of vision-threatening diabetic retinopathy," Electronics, vol. 12, no. 19, p. 4094, 2023.
30. M. Abubakar et al., "ECGNet: High-precision ECG classification using deep learning and advanced activation functions," IEEE Access, 2025.
31. C. Liu et al., "Detection of surface defects in soybean seeds based on improved YOLOv9," Scientific Reports.
32. A. Q. Khan et al., "Hybrid MRUNet+: Enhanced multi-structure retinal segmentation for optic cup, disc, and vascular features," International Journal of Imaging Systems and Technology, vol. 36, no. 1, e70293, 2026.