Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

# Temporal Entropy Hybrid Labelling (TEHL) With GCNN: An Intelligent Deep Learning Framework For Thz Satellite-Based Environmental Monitoring Using Carbon Nanocomposite Sensors

K Priya[1] , J. Jayapriya[2]

[1]Research scholar, Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India. E- mail id : priyakiyalath@gmail.com
[2]Department of Mathematics, Sathyabama Institute of Science and Technology, Tamil Nadu, Chennai India. Email id: Priyanandam1975@rediffmail.com
Corresponding author : priyakiyalath@gmail.com

**Abstract**
Environmental monitoring requires intelligent sensing systems capable of analysing spatially distributed, time-varying, and uncertain environmental conditions. Conventional monitoring approaches often provide limited adaptive labelling and fail to capture spatial dependencies among distributed monitoring locations. To address these limitations, it proposes a Temporal Entropy Hybrid Labelling with Graph Convolutional Neural Network (TEHL-GCNN), a hybrid deep learning model for THz satellite-based environmental monitoring using carbon nanocomposite sensors. A dataset containing 4,320 environmental observations is constructed by integrating carbon nanocomposite sensor measurements with THz-derived and satellite-supported spatial, temporal, spectral, and environmental information. Savitzky–Golay filtering is applied to reduce signal noise while preserving signal characteristics, followed by Z-score normalization to standardize the observations. Node2Vec is then employed to extract low-dimensional node representations that capture structural relationships among monitoring locations. The TEHL mechanism generates adaptive environmental-condition labels by integrating temporal variation, information entropy, and environmental threshold characteristics, enabling improved representation of stable, transitional, and abnormal conditions. Subsequently, monitoring locations are represented as graph nodes, while spatial proximity, environmental similarity, and signal relationships define inter-node connections. The GCNN aggregates neighbourhood information to learn spatial dependencies and classify environmental conditions. The model is implemented using Python 3.11 and evaluated using classification metrics against conventional approaches. The proposed TEHL-GCNN achieves an accuracy of 96.86%, demonstrating its effectiveness in combining adaptive labelling with spatial dependency learning. Overall, the model provides a scalable approach for intelligent environmental monitoring using integrated THz, satellite, and carbon nanocomposite sensor observations.

**Keywords:** Carbon Nanocomposite, Terahertz Satellite, Environmental Monitoring, Temporal Entropy Hybrid Labelling, Graph Convolutional Neural Network.

## 1. Introduction

Environmental monitoring has been found to be an increasingly crucial activity to determine the atmosphere's condition, presence of pollution, and changes in the environmental quality geographically dispersed areas [1]. Traditionally, environmental monitoring is based on fixed sensing stations, which offer limited information regarding environmental conditions since they cover only limited geographical areas [2]. With advancements in sensing techniques and satellite observation systems, there are opportunities to set up flexible monitoring infrastructure that can collect information about the environment from various geographical areas [3]. Carbon nanocomposite-based sensors have also received attention for use in environmental sensing due to the adjustable nature of their electrical, chemical, and surface properties under various environmental conditions [4]. Sensors utilizing nanomaterials have shown high sensitivity in detecting environmental contaminants such as pesticides, heavy metals, and bio-contaminants [5]. Carbon nanomaterials and nanocomposites based on polymers are characterized by having a significant surface area, good conductivity, and adjustable physicochemical properties, which makes their utilization in the development of environmentally sensitive systems possible [6, 7]. Remote sensing via satellites may help overcome the limitations of environmental sensing at limited points through coverage in wide areas and observation of inaccessible locations [8]. Another technological base for the delivery of growing amounts of data collected by distributed sensing systems comes from the use of advanced communication technologies via satellites [9]. Terahertz (THz) communication is especially important for develop high-capacity wireless communication systems owing to the possibilities of high-speed data transfer and broad transmission bandwidth [10].

### 1.1 Aim and Contributions of the Research

The objective of research was to design an intelligent THz satellite environmental monitoring system that would aid in the efficient identification of environmental conditions. The TEHL model designed time variance, information entropy, and environmental thresholds for generating the adaptive labels, whereas the GCNN model helped in learning the spatial relationships between different monitoring sites. The following are the key contributions.

- A new TEHL-GCNN model was proposed for THz-based satellite environmental monitoring using carbon nanocomposite sensors, THz, satellite, and spatial observations.
- Savitzky–Golay filtering reduced noise while preserving signal characteristics, followed by Z-score normalization to standardize observations.
- Node2Vec generated low-dimensional representations, capturing structural relationships among geographically distributed monitoring locations.
- TEHL generated stable, transitional, and abnormal environmental labels using temporal variation, entropy, and thresholds, while GCNN learned spatial dependencies through neighbourhood aggregation.

### 1.2 Research Organization

**Section 1** introduces the research context and significance. Section 2 reviews related studies and identifies research gaps. Section 3 presents the proposed methodology. Section 4 discusses experimental and comparative results. Section 5 concludes the work and outlines limitations and future scope.

## 2. Related Works

A graphene-based plasmonic sensor was developed to enable highly sensitive detection of changes in the refractive index [11]. Polynomial regression and COMSOL Multiphysics were used in the analysis and optimization of sensors; as a result, the researchers obtained the following characteristics: 565 GHz $RIU^{-1}$ sensitivity, 16.607 $RIU^{-1}$ FOM, and 89% accuracy. Nevertheless, simulation dependency, the $R^2$ of 0.86, and absence of experimental work became drawbacks. The research proposed the development of a graphene-metal nanoparticle metasurface biosensor aimed at the rapid bacterial detection in water [12]. A $WSe_2$–graphene hybrid plasmonic system was investigated for the development of a non-invasive THz glucose biosensor. The simulation and K-nearest regressor optimization [13] were conducted using COMSOL Multiphysics software for sensitivity analysis and prediction. Research found that the sensor has sensitivity of 1000 GHz/RIU and significant $R^2$ values. Nevertheless, research was restricted by the lack of real-life experiments. A multi-layer terahertz biosensor was developed for the rapid detection of bacteria in water. The evaluation and optimization of its performance were performed using COMSOL [14] Multiphysics simulations and polynomial regression. The sensor provided the sensitivity of 811 GHz/RIU, a local sensitivity of 976 GHz/RIU, and an $R^2$ equal of 0.94. However, this approach was based exclusively on modeling and did not include experiments.

Finite element multiphysics simulation along with KNN regression was used for predicting sensor responses [15]. While accurate results were obtained, the technology remained simulation-based and lacked practical validation. A hybrid THz metasurface sensor using graphene and metal/dielectric structures was developed for detecting amino acids based on simulations and ML [16]. Yet, experimental and clinical validation was not performed. A Label-free THz metagrating bacteria detection sensor using 3D printing of dielectric structures and supervised ML was designed [17]. Yet, experimental and clinical validation was not provided. The objective of the research was to create graphene-based refractive index sensors to detect haemoglobin [18]. Three different sensors were compared based on the analysis of transmittance, electric field, sensitivity, Q-factor, figure-of-merit, and estimation of intermediate transmittance by XGBoost regression. XGBoost regression helped to save computational time and resources by 70% at least. Nonetheless, practical application for biomedical purposes was not provided. The objective of the research was to calculate the spatiotemporal distribution of PM2.5 concentrations with remote sensing, ground measurements, and meteorological factors [19]. The MLP, CNN, LSTM, and ConvLSTM models were compared using RMSE and $R^2$. ConvLSTM showed the best result (RMSE: 4.95 $\mu g/m^3$; $R^2$: 91.24%) [20]. Nevertheless, availability of data and pollution differences were problematic.

### 2.1 Research Gap

Two significant limitations of the research mentioned below are that [11] mainly used COMSOL-based simulation and ML-based prediction methods, which did not have adequate experiments or real-world environmental validations, and research [20] was constrained due to lack of data and challenges associated with spatial variations. The proposed TEHL-GCNN model solves these limitations by using THz, satellite, and carbon nanocomposite sensor measurements, where TEHL accounts for the temporal and uncertain variations and GCNN learns spatial dependencies between various geographically distributed measurement locations.

## 3. Methodology

The proposed TEHL-GCNN model consists of four main steps: firstly, multimodal environmental observations from carbon nanocomposite sensors, THz, and satellite sources are collected; secondly, noise is reduced using SG filtering, followed by Z-score normalization. Spatial features are extracted using Node2Vec, while TEHL generates adaptive environmental labels. Finally, GCNN learns spatial dependencies to classify environmental conditions as stable, transitional, and abnormal. Figure 1 illustrates the complete TEHL-GCNN pipeline, showing multimodal data acquisition, preprocessing, spatial feature extraction, adaptive labelling, graph-based learning, and environmental-state classification.

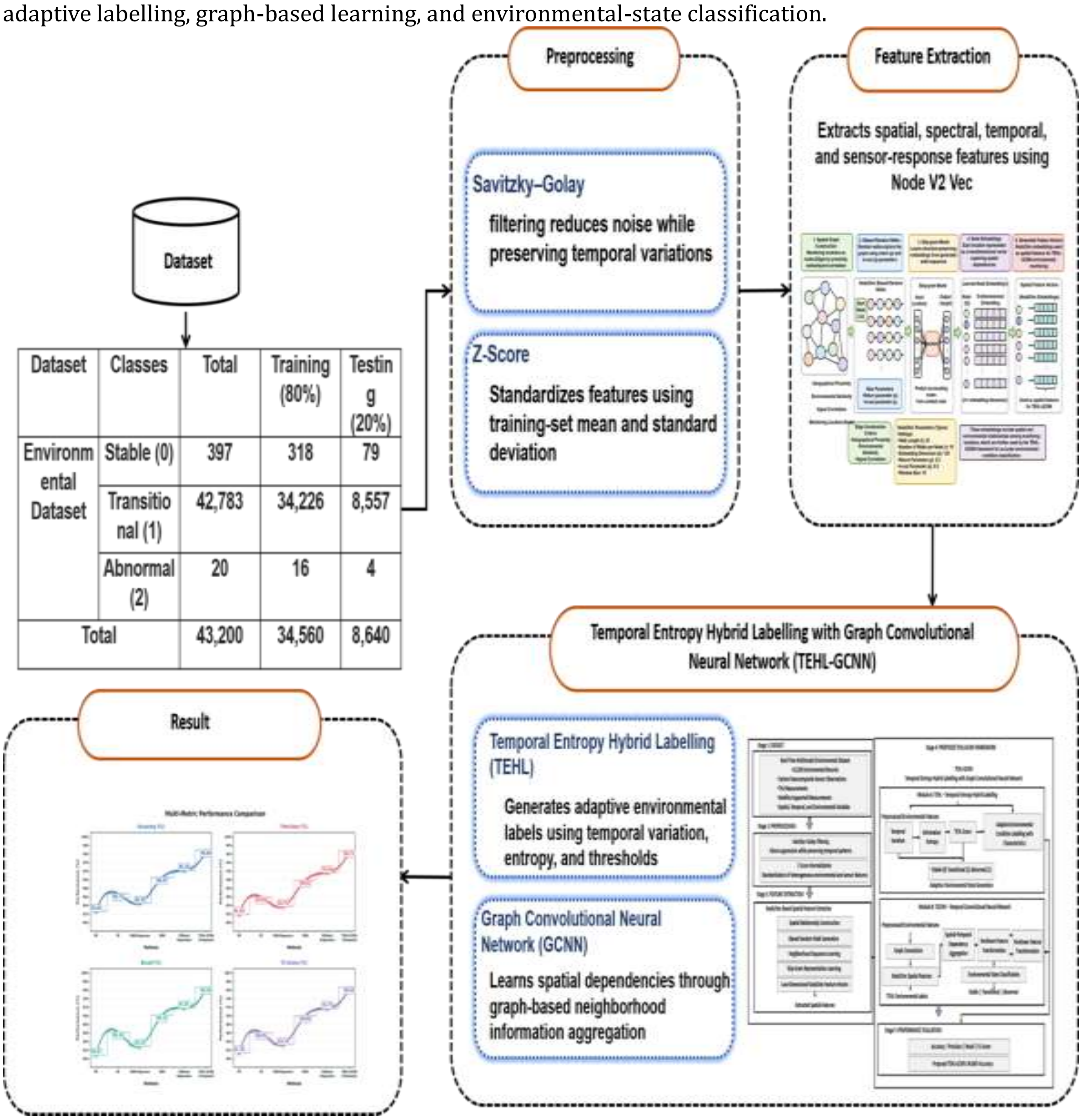


**Figure 1: TEHL-GCNN model for intelligent environmental-state classification**

### 3.1 Dataset collection

Experimentation has been done using a dataset based on real-time environmental monitoring that combines the measurements obtained from the carbon nanocomposite sensors, THz, and satellites. The dataset consists of 4,3200 records with 53 input features covering geographical features, temporal features, atmospheric features, sensor response features, THz features, satellite-derived features, spectral features, and environmental features. The dataset is divided into 80% training (3,4560 records) and 20% test (8640 records). The 52 input features consist of geographical features and temporal features, environmental features, carbon nanocomposite sensors features, THz features, satellite-derived features, spectral features, and temporal features in Table 1. The target feature, environmental_condition_label, consists of three environmental states, stable, transitional, and abnormal labeled with 0, 1, and 2, respectively in Figure 2.

**Table 1. Overview and training–testing distribution of the environmental monitoring dataset**

| Dataset | Classes | Total | Training (80%) | Testing (20%) |
|---|---|---|---|---|
| Environmental Dataset | Stable (0) | 397 | 318 | 79 |
| | Transitional (1) | 42,783 | 34,226 | 8,557 |
| | Abnormal (2) | 20 | 16 | 4 |
| Total | | 43,200 | 34,560 | 8,640 |

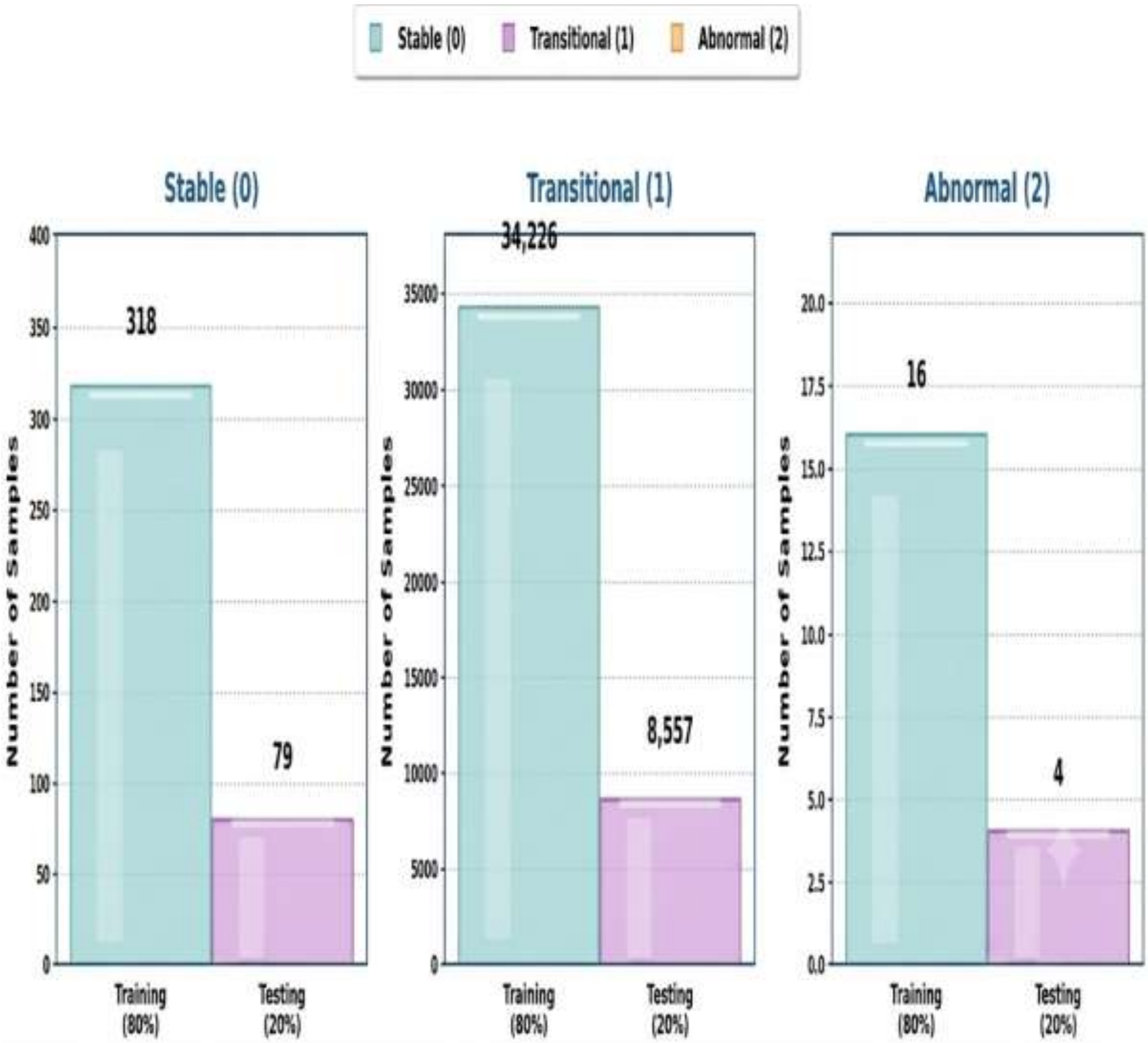


**Figure 2: Train-test split distribution across stable, transitional, and abnormal sample classes**

### 3.1. Preprocessing of Multimodal Environmental Observations

The Savitzky-Golay filter removes noise from time series measurements in sensors and THz data through the fitting of a polynomial to each observation window [19]. Unlike averaging methods, it retains the peaks, slopes, and short-term time variations. This is important for TEHL, as temporal differences are necessary to differentiate between normal and abnormal conditions.

$$Y_j = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+1}}{N} \quad (1)$$

In Equation (1), $Y_j$represents the smoothed value at position j, $C_i$denotes the weighting coefficient applied at index i, $Y_{j+1}$represents the input data value used in the summation, iis the summation index ranging from $-m$to m, N denote the half-width of the window, $\sum_{i=-m}^{i=m} C_i$denotes the summation over all indices the normalization factor used to scale the result. Reduces noise while preserving important temporal signal variations

After noise reduction the Z-score normalization process standardizes all the environmental, response, and spectral features with respect to their training set means and standard deviations [20]. The process mitigates scale differences between diverse features, eliminates numerical dominance of particular features during TEHL and GCNN learning, and facilitates spatial feature propagation.

$$rss_{l,n} = [rss(1), rss(2) \ldots rss(N)] \text{ for } 1 \le m \le M \quad (2)$$

In Equation (2),$rss_{i,n}$represents the received signal strength vector for the pair (l, n), rss(N)denotes the individual received signal strength value at the mth sample/reading, with the condition $1 \le m \le M$indicating that nranges over all valid sample indices from 1to M.

$$Z_{score} = \frac{rss(m) - \mu}{\sigma} \quad (3)$$

In Equation, $Z_{score}$represents the standardized (normalized) value of the signal, rss(m)denotes the received signal strength value at sample n, $\mu$represents the mean of the received signal strength values, and $\sigma$represents the standard deviation of the received signal strength values. Standardizes heterogeneous features and minimizes scale differences.

### 3.2. Spatial Feature Extraction using Node2Vec

The monitoring locations in the form of low-dimensional embeddings after normalization, biased random walks on the environmental spatial graph, are made using the Node2Vec method. The embeddings that preserve the local neighborhood as well as larger patterns of connectivity between sensor locations are learned [21]. Such a topology-based representation of the features allows TEHL-GCNN to utilize the spatial relationships to better distinguish different types of environments. Figure 3 illustrates the procedure of Node2Vec-based spatial feature extraction for TEHL-GCNN environmental monitoring. Monitoring locations become the nodes of the graph with connections between them made due to geographical proximity, environmental similarity, and signal correlation.

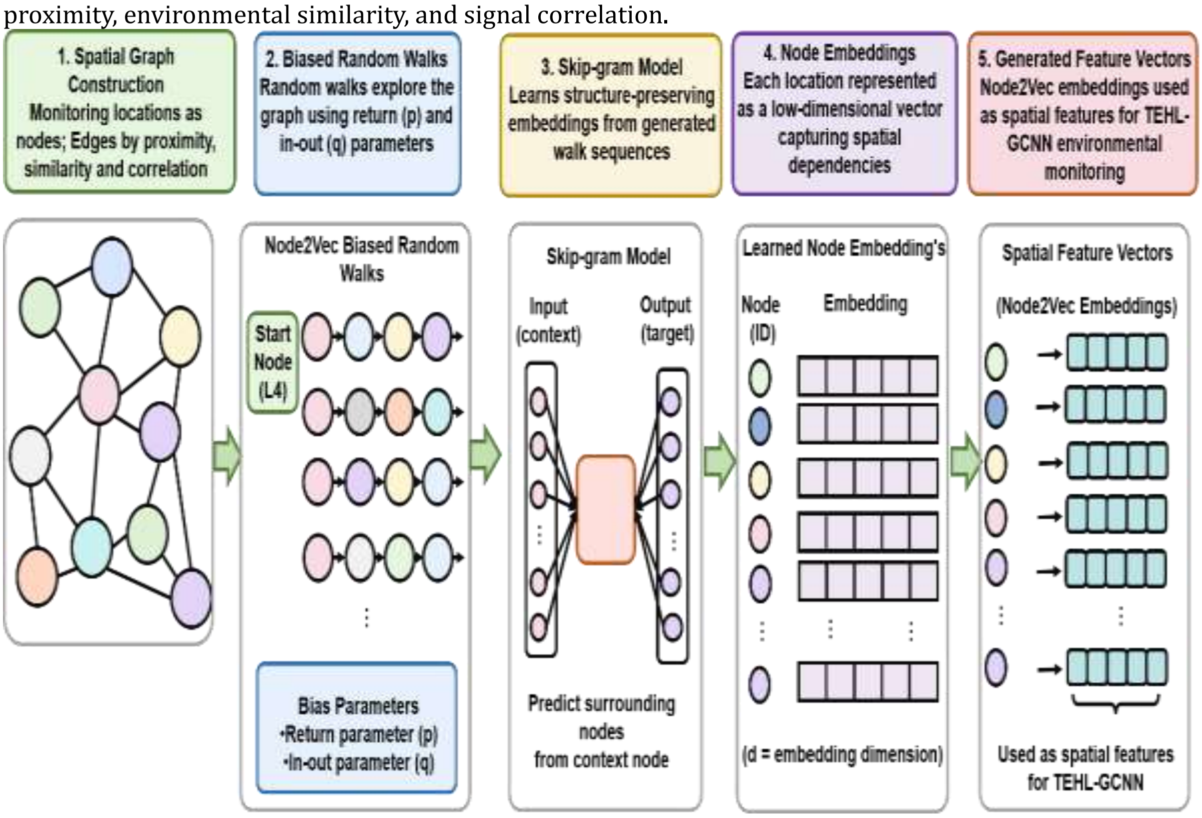

**Figure 3: Node2Vec-based spatial feature extraction workflow for TEHL-GCNN environmental monitoring**

$$\alpha_{rs}(v,x) = \begin{cases} \frac{1}{r}, & c_{vx} = 0 \\ 1, & c_{vx} = 1 \\ \frac{1}{s}, & c_{vx} = 2 \end{cases} \quad (4)$$

In Equation (4), $\alpha_{rs}(v,x)$is the transition bias toward node vfrom node r, sthe return parameter, xis the in-out parameter, and $c_{vx}$is the shortest path distance between vand x, where $c_{vx} = 0$means returning to the previous node, $c_{vx} = 1$means a directly connected node, and $c_{vx} = 2$means a further node.

$$\pi_{rs} = \alpha_{rs}(v,x) \cdot wt_{vx} \quad (5)$$

$$\pi_{rs} = \alpha_{rs}(v,x) \cdot M(w) \quad (6)$$

In Equations (5 and 6), $\pi_{rs}$is the unnormalized transition probability from node rto node s, $\alpha_{rs}(v,x)$is the search bias factor based on parameters pand q, and $wt_{vw}$is the edge weight between nodes vand x. M(w)is the degree (or neighborhood-based weight) of node w.

$$P(c_i = w \mid c_{i-1} = v) = \begin{cases} \frac{\pi_{rs}}{\eta}, & (v,x) \in E \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

In Equation (7), $P(c_i = w \mid c_{i-1} = v)$is the probability that the walk moves to node wat step igiven it was at node vat step $i-1$, $\pi_{rs}$is the unnormalized transition probability between nodes v and w, $\eta$is the normalizing constant, $(v,x) \in E$indicates that an edge exists between nodes vand w is the graph, and the probability is 0when no such edge exists. Generates spatial embeddings that preserve local and broader monitoring-location connectivity

### 3.3 Graph Convolutional Learning for Irregular Spatial-Environmental Relationships and Adaptive Condition Classification

The Traditional CNN [22] is tailored to work on grid-like data and fails to capture any irregular relationships between monitoring locations that are geographically dispersed. To address this problem, the proposed GCNN to accurately predict stable, transitional, and abnormal environmental conditions across geographically distributed monitoring locations by learning non-Euclidean spatial dependencies from heterogeneous environmental observations and Node2Vec-derived spatial representations. The proposed GCNN uses monitoring locations as graph nodes and relationships between them in terms of geography, environment, and signals as graph edges. The Graph Convolution operation aggregates information from neighboring nodes to capture and understand their spatial associations. The GCNN classification algorithm integrates TEHL-generated environmental-state labels with Node2Vec-derived spatial embeddings to learn spatial dependencies and accurately classify monitoring locations into stable, transitional, and abnormal environmental conditions. TEHL tags and Node2Vec embeddings are used for the monitoring nodes, whereas spatial and environmental associations are used in terms of graph edges in Figure 4.

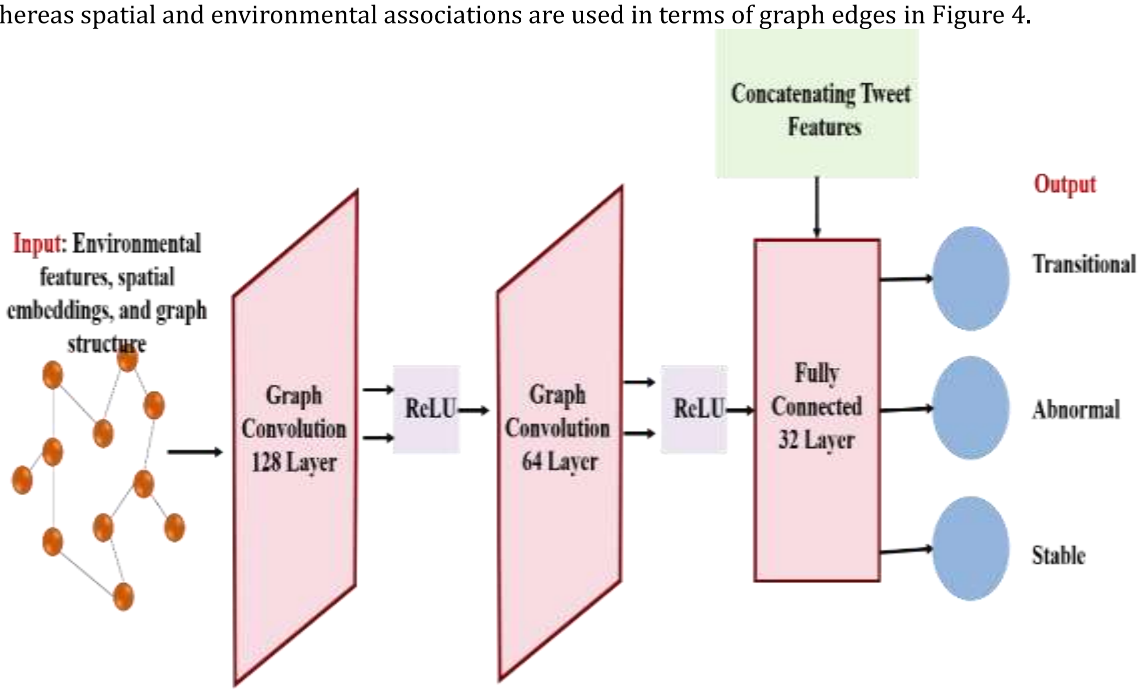


**Figure 4: GCNN-based spatial learning for adaptive environmental condition classification**

$$y^{l+1} = f(y^l, B) = \sigma(By^l x^l) \quad (8)$$

In Equation (8), $y^{l+1}$is the node feature matrix at layer $l + 1$, $y^l$is the node feature matrix at layer l, B is the graph adjacency matrix, $x^l$is the trainable weight matrix at layer l, $f(y^l, B)$is the layer-wise propagation function, and $\sigma(\cdot)$is the nonlinear activation function.

$$\lambda_i(s) = \mu_i + \sum_j b_{ij}\ k(s - s_j) \quad (9)$$

In Equation (9), $\lambda_i(s)$is the conditional intensity (event rate) of node iat time t, $\mu_i$is the baseline intensity of node i, $b_{ij}$is the influence weight of node jon node i, $k(\cdot)$is the triggering (decay) kernel function, $s - s_j$is the occurrence time of a past event at node j, and $\sum_j b_{ij}$denotes the summation over all past events from node j.GCNN learns spatial dependencies through neighbourhood information aggregation.

**3.4 Temporal Entropy Hybrid Labelling for Adaptive Environmental States**

TEHL constructs adaptive environmental labels through independent analysis of temporal variability, entropy, and threshold features that are extracted from the observed environmental data. Temporal variability accounts for changing conditions, and entropy is a measure of uncertainty and signal variability. Threshold features serve as environmental measures for classifying the three states as stable, transitional, and anomalous. It should be noted that the labels produced by the TEHL algorithm are done prior to GCNN training, and they do not act as input features or are extracted from the GCNN predictions. Therefore, the target labels remain isolated from the input features and do not include information about the target variable.

$$H_t = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \quad (10)$$

In Equation (10), $H_t$is the entropy value, $p(x_i)$is the probability of occurrence of event $x_i$, $x_i$is the ith possible outcome, nis the total number of possible outcomes, iis the outcome index, $\sum_{i=1}^{n}$denotes the summation over all outcomes, and $\log(\cdot)$is the logarithmic operator.

$$L = \begin{cases} 1, & \text{if } H_t > \theta \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

In Equation (11), Lis the binary output label, $H_t$is the entropy value at time t, and $\theta$is the threshold value used to decide whether Lis assigned 1(when $H_t$exceeds $\theta$) or 0(otherwise).

**3.5 State Prediction for Distributed Multimodal Environmental Monitoring Systems using Temporal Entropy Hybrid Labelling with Graph Convolutional Neural Network (TEHL-GCNN)**

An integrated learning pipeline is developed to identify environmental states across distributed monitoring locations by combining adaptive temporal labelling with spatial graph learning. TEHL integrates temporal variation, information entropy, and environmental thresholds to generate target labels representing stable, transitional, and abnormal conditions. Node2Vec subsequently encodes spatial relationships, while the GCNN learns non-Euclidean dependencies from environmental and spatial representations to enable accurate environmental-state prediction.

$$L_i = [w_t T_i(c) + w_e E_i(c) + w_h H_i(c)] \quad (12)$$

In Equation (12), $L_i$ is the TEHL-generated target label for observation ii, SS, TT, and AA denote the stable, transitional, and abnormal states respectively, c represents the candidate state being evaluated, $T_i(c)$ represents the temporal-variation evidence, $E_i(c)$ denotes the entropy-based uncertainty evidence, $H_i(c)$ represents the environmental-threshold evidence, and $w_t$, $w_e$, $w_h$ are the corresponding weighting factors for the three evidence terms.

$$\hat{Y} = \mathrm{Softmax}\left(W_o\, \sigma\left(\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} X_{env+N2V} W_g\right) + b_o\right) \quad (13)$$

In Equation (13), $\hat{Y}$ represents the predicted stable, transitional, or abnormal state, $\widetilde{A} = A + IA\sim$ is the adjacency matrix with self-connections added, AA is the original adjacency matrix, I is the identity matrix, $\widetilde{D}$ is the degree matrix of $\widetilde{A}$, $X_{env+N2V}$ combines environmental features and Node2Vec embeddings, $W_g$ and $W_o$ are trainable weight matrices, $b_o$bo is the bias term, $\sigma(\cdot)$ denotes the nonlinear activation function, and $\mathrm{Softmax}(\cdot)$Softmax$(\cdot)$ is the output function that converts the result into class probabilities

TEHL-GCNN converts heterogeneous environmental observations into reliable condition predictions. Node2Vec extracts spatially meaningful features, TEHL identifies adaptive environmental states using temporal variation and entropy, and TGCNN learns spatial-temporal dependencies, improving recognition of Stable, Transitional, and Abnormal conditions in Figure 5.

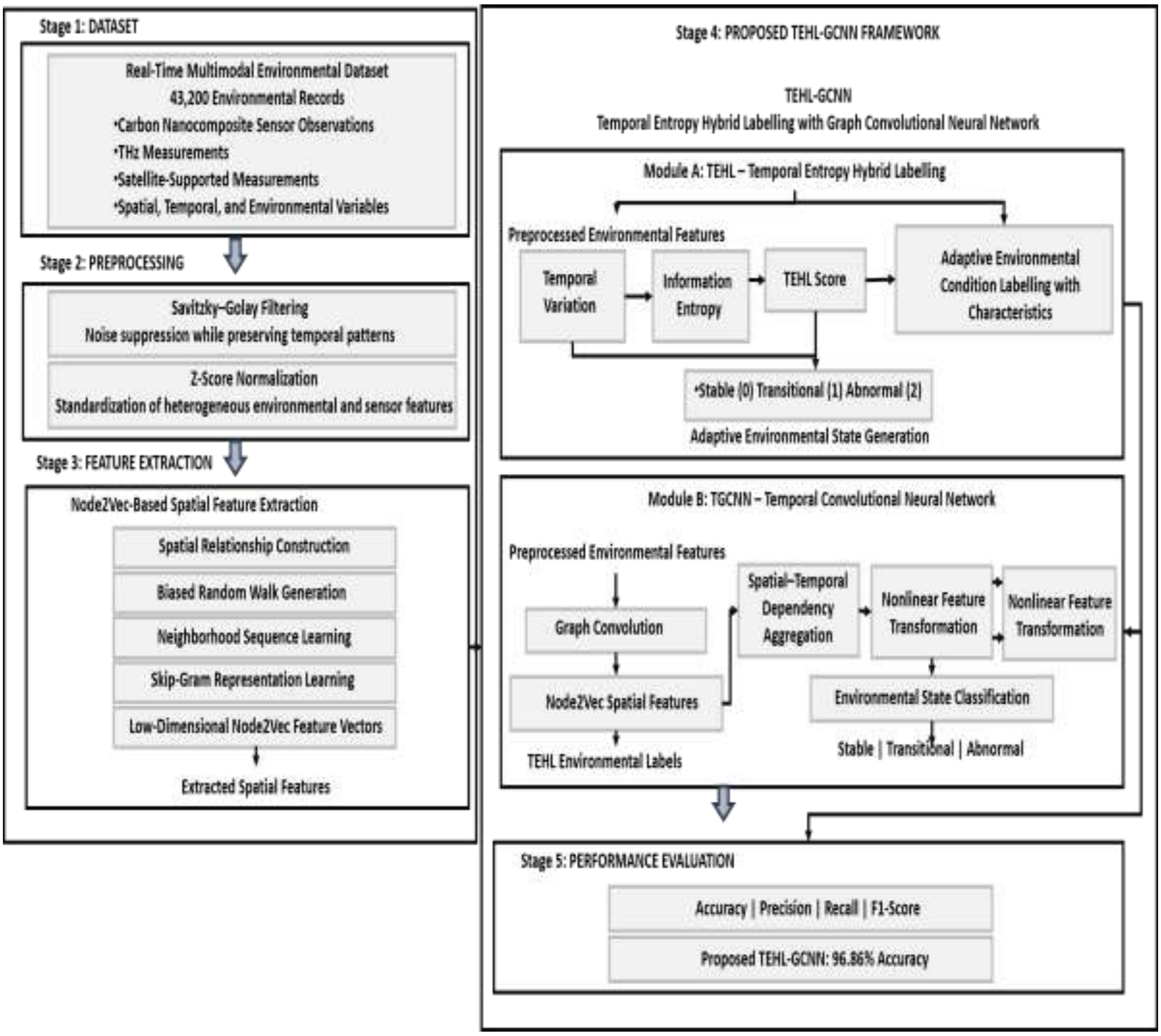


**Figure 5: Proposed TEHL-GCNN architecture for intelligent environmental state classification.**

The proposed TEHL-GCNN algorithm 1 processes the multimodal environment observation data using Savitzky-Golay filtering and Z-score scaling. Node2Vec learns the spatial embedding of the monitoring graph based on topology. TEHL considers temporal variability, entropy, and environmental thresholds to classify the environmental states into stable, transitional, and abnormal states.

**Algorithm 1: TEHL-GCNN-Based Environmental State Prediction**

**Input:** Multimodal Environmental Observation Dataset (D)
**Output:** Predicted Environmental State ((y))ˇ
 Load the multimodal environmental observation dataset D.
 Split D into D_train (80%) and D_test (20%).
 Apply Savitzky–Golay filtering to temporal sensor and THz observations.
 Perform Z-score normalization using training-set mean μ and standard deviation σ.
 Construct the spatial-environmental graph G = (V, E, W).
 Represent monitoring locations as nodes and spatial relationships as weighted edges.
 Initialize Node2Vec parameters p, q, and walk length L.
 Generate biased random walks and learn node embeddings Z using Skip-gram.
 Compute temporal variation Δ_i(t) for each monitoring node.
 Estimate observation probabilities and calculate entropy H_i.
 Define environmental threshold values τ_i for state determination.
 for each monitoring node i do
  Evaluate Δ_i(t), H_i, and threshold conditions.
  if temporal variation and entropy remain within stable ranges then
   Stable;

else if temporal or entropy variation indicates evolving conditions then
Transitional;
else
Abnormal;
end if
end for
Combine normalized features, TEHL labels, and Node2Vec embeddings Z as GCNN inputs.
Perform graph convolution, neighbourhood aggregation, and environmental-state classification.
Return predicted environmental states $\hat{Y}$ and classification performance.

### 3.6 Evaluation Metrics

**Accuracy:** The aim was to evaluate the TEHL-GCNN approach for THz satellite environmental sensing by identifying environmental conditions from the integration of THz, satellite, and carbon nanocomposite sensor data, with accuracy being the ratio of correctly classified environmental conditions to the total number of evaluated observations using Equation (14).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (14)$$

**Precision:** The purpose was to evaluate the TEHL-GCNN approach for THz-based environmental monitoring using satellites in terms of accurately recognizing environmental situations, with Precision defined as the ratio of correctly recognized environmental situations out of all the positive ones recognized by the proposed approach using Equation (15).

$$\text{Precsion} = \frac{\text{TP}}{\text{TP+FP}} \quad (15)$$

**Recall:** Objective was to evaluate the TEHL-GCNN model for THz satellite environmental monitoring through accurate detection of environmental conditions, whereas recall determined the percentage of true environmental conditions that were detected by the suggested model using Equation (16).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (16)$$

**F1-Score:** The goal was to evaluate the performance of the TEHL-GCNN structure for environmental monitoring using THz satellites by obtaining balanced environmental condition classifications, where F1-score was the measurement of the harmonic mean of precision and recall using Equation (17).

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precsion} \times \text{Recall}}{\text{Precsion+Recall}} \quad (17)$$

**Note**TP = True Positive, FP = False Positive, FN = False Negative

## 4. Result

The suggested model succeeds in identifying different environmental conditions at distributed sensing locations. The observed results suggest uniform state identification for different environmental patterns and efficient decision-making in discriminating normal, transitional, and abnormal states.

### 4.1 Experimental Setup

The computational model applied in realizing the suggested TEHL-GCNN model. This arrangement offers computational capability required to process of environmental observations and the learning model as described in Table 2. The aim of this arrangement is to create an experimentally reproducible environment.

**Table 2: Computational hardware and software configuration for TEHL-GCNN experiments**

| Category | Component | Recommended Specification / Version |
|---|---|---|
| Hardware | Processor | Intel Core i7 / AMD Ryzen 7 or higher |
| | CPU Cores | 8 cores or higher |
| | RAM | 32 GB |
| | GPU | NVIDIA RTX 4060 / RTX 4070 or higher |
| | GPU Memory | 8–16 GB |
| | Storage | 512 GB SSD or higher |
| | OS | Windows 11 64-bit / Ubuntu 22.04 LTS |
| | Network | Broadband Internet connection |
| Software | PL | Python 3.11 |
| | Development Environment | Jupyter Notebook / VS Code |
| | DL model | TensorFlow 2.15+ / Keras |
| | Numerical Computing | NumPy |
| | DP | Pandas |
| | ML | Scikit-learn |
| | Graph Processing | NetworkX |

| | | |
|---|---|---|
| | Node Embedding | Node2Vec |
| | Visualization | Matplotlib, Seaborn |
| | GPU Acceleration | CUDA 12.2 |
| | GPU Library | cuDNN compatible with installed TensorFlow |
| | Operating System | Windows 11 64-bit |

Hyperparameters and settings for TEHL-GCNN are presented in Table 3, including temporal entropy and hybrid labelling, graph construction, GCNN architecture, regularization, and training parameters. The chosen parameter values are ten temporal windows, three GCNN layers, the Adam optimizer with a learning rate of 0.001, a 32 batch size, 100 epochs, and an 80/20 train/test split.

**Table 3: Recommended Hyperparameters and Settings for TEHL-GCNN**

| Component | Hyperparameter | Symbol / Setting | Recommended Value |
|---|---|---|---|
| Temporal Entropy | Number of temporal windows | (M) | 10 |
| | Entropy normalization | (H_{norm}) | z-score |
| | Temporal change weight | (\gamma) | 0.30 |
| Hybrid Labelling | Feature evidence weight | (\alpha) | 0.40 |
| | Entropy weight | (\beta) | 0.30 |
| | Entropy variation weight | (\gamma) | 0.30 |
| | Lower threshold | (\tau_1) | 0.30 |
| | Middle threshold | (\tau_2) | 0.60 |
| | Upper threshold | (\tau_3) | 0.80 |
| Graph Construction | Graph type | — | k-NN |
| | Number of neighbors | (k) | 10 |
| | Similarity measure | — | Cosine similarity |
| | Similarity threshold | (\theta) | 0.70 |
| | Self-loop | — | Enabled |
| | Adjacency normalization | — | Symmetric |
| GCNN | Number of graph convolution layers | (L) | 3 |
| | Hidden units – Layer 1 | (F_1) | 128 |
| | Hidden units – Layer 2 | (F_2) | 64 |
| | Hidden units – Layer 3 | (F_3) | 32 |
| | Activation function | (\sigma) | ReLU |
| | Output activation | — | Softmax |
| Regularization | Dropout rate | (p) | 0.30 |
| | L2 regularization | (\lambda) | 0.0001 |
| Training | Optimizer | — | Adam |
| | Learning rate | (\eta) | 0.001 |
| | Batch size | (B) | 32 |
| | Maximum epochs | — | 100 |
| | Early stopping patience | — | 10 |
| | Loss function | — | Categorical Cross-Entropy |

### 4.2 Data Exploration

Identify the time and information-entropy properties of environmental data and the adaptive labelling limits. Information-entropy is mainly clustered at 0.60-0.80, with threshold zones at 0.62, 0.65 and 0.67 in Figure 6(a). TEHL values vary between 0.00 and 0.35, with an increase in probability within 0.20-0.30. This means that there is separation of the environment in Figure 6(b).

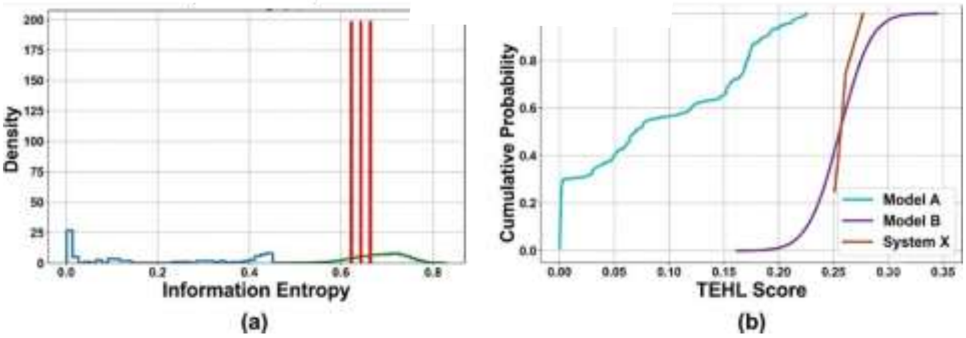


**Figure 6: Overview of temporal entropy-based environmental condition characterization: (a) distribution of information entropy with adaptive threshold regions (b) cumulative probability distribution of TEHL scores**

To identify relationships between the environmental features that have been extracted and any dependencies that could exist prior to spatial learning. The positive and negative relationships between normalized numbers ranging from 0 to 1 are depicted in Figure 7(a). Values ranging from −0.90 to +0.90 are depicted in Figure7 (b).

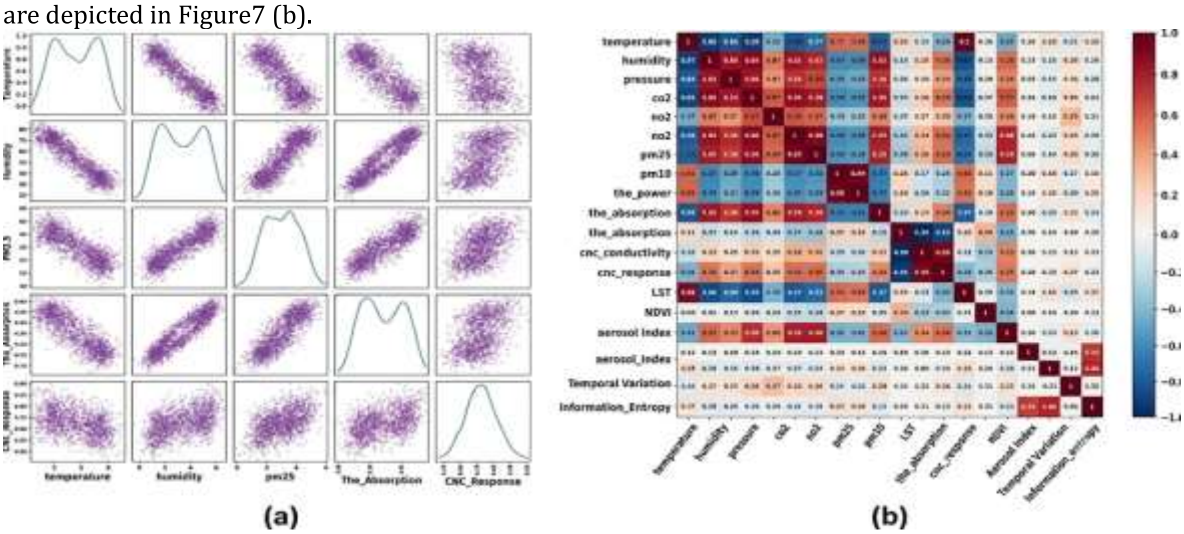


**Figure 7: statistical relationships among environmental sensing variables: (a) pairwise distributions illustrating individual feature distributions (b) correlation structure showing coefficients**

Analyse the reaction to environmental stimuli at different humidity and temperature levels and evaluate the impact of this on the behavior of the sensor in Figure 8 (a). The absorption is higher with higher humidity at around 0.34-0.49 in the range of 50-85% and shows higher reactions from 74-80%. The response is in the range of 0.58-0.76 at 22-37 °C in Figure 8(b).

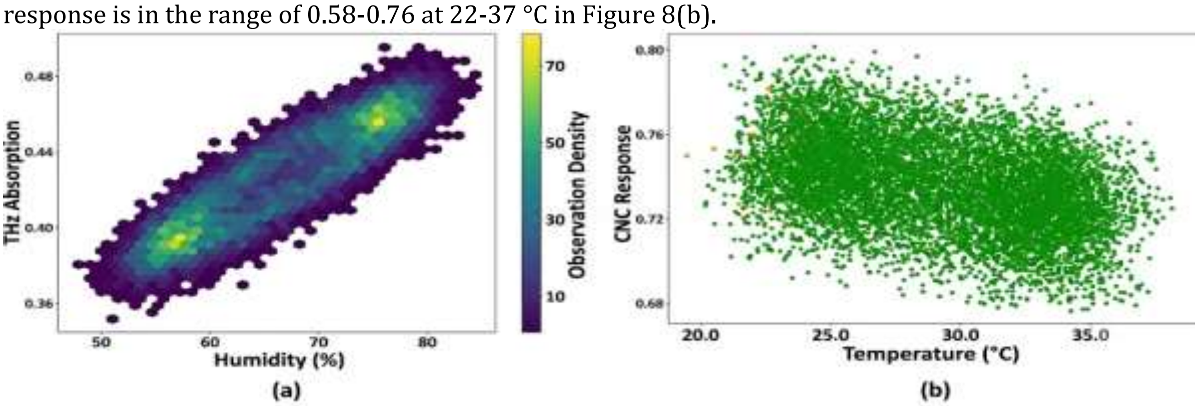


**Figure 8: Environmental response variation with humidity and temperature: (a) humidity-dependent absorption response and (b) temperature-dependent sensor response.**

**4.2 Performance Evaluation**

The proposed TEHL-GCNN model is evaluated against existing approaches, including Polynomial Regression [11], Stacking Ensemble [12], KNN Regressor [13], Deep Neural Network (DNN) [15], and XGBoost Regression [19]. These methods represent conventional regression, ensemble learning, neighbourhood-based learning, deep learning, and tree-based prediction approaches in Table 4. For fair benchmarking, all baseline models were reimplemented and retrained on the Real-time environmental dataset using identical preprocessing procedures and experimental settings. As shown in Figure 9, TEHL-GCNN achieves 96.86% accuracy, demonstrating superior environmental-state discrimination and reliable identification of stable, transitional, and abnormal conditions across distributed monitoring locations.

**Table 4. Comparative environmental-state classification performance across evaluated models**

| Method | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Polynomial Regression (PR) | 82.64 | 81.93 | 80.87 | 81.39 |
| Stacking Ensemble (SE) | 86.72 | 85.94 | 86.18 | 86.05 |
| KNN Regressor | 84.53 | 83.76 | 82.91 | 83.33 |
| Deep Neural Network (DNN) | 90.47 | 89.82 | 90.15 | 89.98 |
| XGBoost Regression | 92.18 | 91.64 | 91.87 | 91.75 |
| TEHL-GCNN [Proposed] | 96.86 | 96.72 | 96.58 | 96.65 |

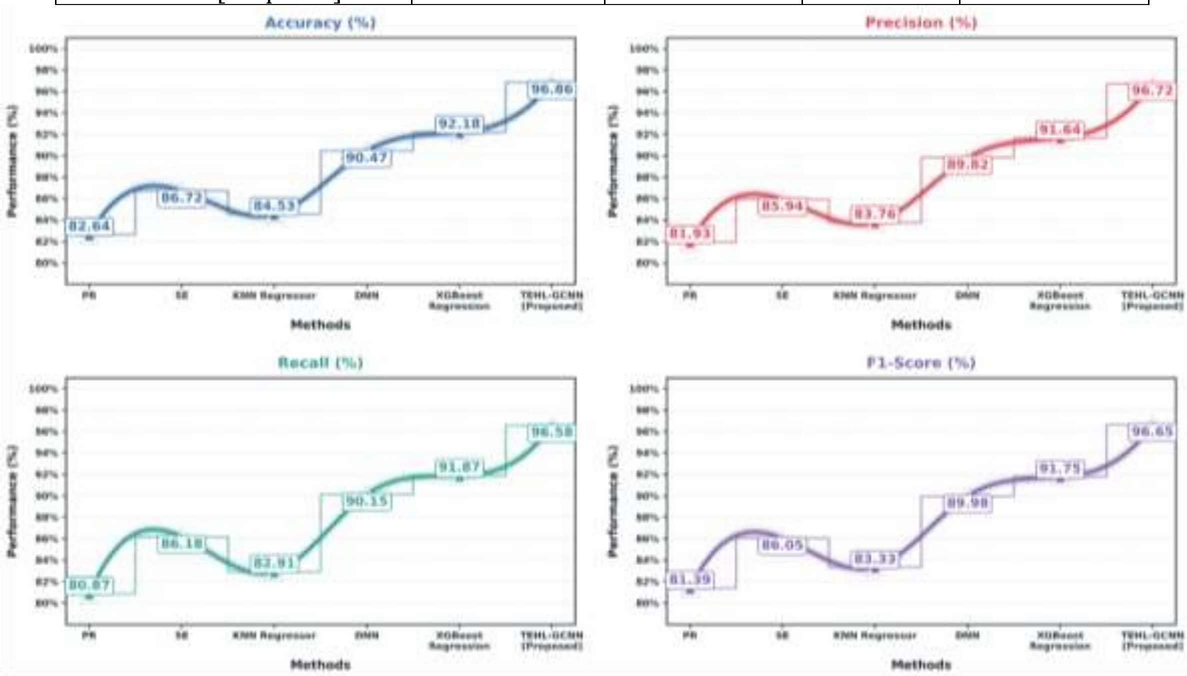


**Figure 9: Environmental-state prediction performance comparison among competing learning models such as accuracy, precision, recall, and F1-score**

Progressive contributions of the individual elements of the TEHL-GCNN architecture are evaluated through the ablation study experiment shown in Table 5. The accuracy of the base GCNN architecture is 89.74% and defines the performance. Progressively adding the normalization layer, the filtering approach, and the Node2Vec approach improves the feature and spatial learning. Finally, with the help of TEHL, it becomes possible to label the environmental state adaptively and achieve an accuracy of 96.86%.

**Table 5: Ablation study evaluating individual contributions of TEHL-GCNN components**

| Configuration | SG Filtering | Z-Score Normalization | Node2Vec | TEHL | GCNN | Accuracy (%) |
|---|---|---|---|---|---|---|
| CNN | ❌ | ❌ | ❌ | ❌ | ❌ | 80.56 |
| GCNN | ❌ | ❌ | ❌ | ❌ | ✅ | 89.74 |
| ZN + GCNN | ❌ | ✅ | ❌ | ❌ | ✅ | 91.36 |
| SG + ZN + GCNN | ✅ | ✅ | ❌ | ❌ | ✅ | 92.81 |
| SG + ZN + N2V + GCNN | ✅ | ✅ | ✅ | ❌ | ✅ | 94.93 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **TEHL-GCNN [Proposed]** | ✅ | ✅ | ✅ | ✅ | ✅ | **96.86** |

The reliability and consistency of TEHL-GCNN across five-fold cross-validation by comparing its performance with CNN and GCNN in Table 6. The performance of the suggested model is consistent, as its average accuracy is 96.86% with a very low standard deviation of 0.10. The 95% confidence interval provides a more accurate measure of uncertainty for reliable generalization.

**Table 6: Five-fold accuracy, variability, and confidence interval analysis**

| **Model** | **Fold** | **Accuracy (%)** | **Mean Accuracy (%)** | **Standard Deviation** | **95% CI (%)** |
|---|---|---|---|---|---|
| CNN | 1 | 90.84 | 91.04 | 0.15 | 90.85–91.23 |
| | 2 | 91.12 | | | |
| | 3 | 90.97 | | | |
| | 4 | 91.25 | | | |
| | 5 | 91.03 | | | |
| GCNN | 1 | 93.18 | 93.40 | 0.17 | 93.19–93.61 |
| | 2 | 93.47 | | | |
| | 3 | 93.31 | | | |
| | 4 | 93.62 | | | |
| | 5 | 93.42 | | | |
| TEHL-GCNN [Proposed] | 1 | 96.71 | 96.86 | 0.10 | 96.73–96.99 |
| | 2 | 96.93 | | | |
| | 3 | 96.84 | | | |
| | 4 | 96.98 | | | |
| | 5 | 96.84 | | | |

Wilcoxon signed-rank test results for seven selected environmental features to assess statistically significant differences between paired observations.The mean differences show the degree of change between paired observations, and the W values denote the signed-rank test outcomes presented in Table 7. The p-values that are lower than 0.05 indicate statistically significant differences for temperature, humidity, CO2, CNC response, THz power, NDVI, and time variation.

**Table 7. Wilcoxon signed-rank analysis of selected environmental features**

| **Feature** | **Mean Difference** | **W Statistic** | **p-value** | **Significance** |
|---|---|---|---|---|
| Temperature | 0.184 | 18.0 | 0.021 | |
| Humidity | 0.162 | 16.0 | 0.032 | |
| $CO_2$ | 0.213 | 14.0 | 0.048 | Significant |
| CNC Response | 0.247 | 12.0 | 0.018 | |
| THz Power | 0.196 | 15.0 | 0.027 | |
| NDVI | 0.151 | 17.0 | 0.041 | |
| Temporal Variation | 0.268 | 10.0 | 0.012 | |

The computational time and memory needed by the individual modules, as well as the entire TEHL-GCNN model in Table 8. The SG Filtering and Z-Score Normalization techniques are fairly computational, while Node2Vec and GCNN are computationally intensive techniques since they make use of graphs and learn from neighborhoods. TEHL and graph creation are moderately computation- intensive techniques. The entire TEHL-GCNN framework takes up 109.23 ms execution time, 102.43 ms processing time, and 222.2 MB memory, making it fairly computationally feasible for environmental monitoring tasks.

**Table 8: Computational time and memory requirements of the proposed TEHL-GCNN components**

| **Component** | **Execution Time (ms)** | **Processing Time (ms)** | **Memory Usage (MB)** |
|---|---|---|---|
| SG Filtering | 8.42 | 7.86 | 18.4 |
| Z-Score Normalization | 1.76 | 1.42 | 9.8 |
| Node2Vec | 42.35 | 39.87 | 64.7 |
| TEHL | 6.28 | 5.91 | 15.6 |
| Graph Construction | 18.74 | 17.53 | 31.2 |

| | | | |
|---|---|---|---|
| GCNN | 31.68 | 29.84 | 82.5 |
| TEHL-GCNN [Proposed] | 109.23 | 102.43 | 222.2 |

## 5. Discussion

The proposed TEHL-GCNN model integrates multimodal environmental observations from carbon nanocomposite sensors, THz measurements, and satellite-derived data for environmental-state classification. The model applies Savitzky–Golay filtering and Z-score normalization to obtain consistent observations, while TEHL combines temporal variation, entropy, and environmental thresholds to identify stable, transitional, and abnormal conditions. Node2Vec represents spatial relationships among monitoring locations, and the GCNN learns the dependencies among connected monitoring nodes. Existing prediction studies have investigated Polynomial Regression [11], Stacking Ensemble [12], KNN Regressor [13], Deep Neural Network (DNN) [15], and XGBoost Regression [19] for learning relationships from environmental and monitoring-related data. These approaches support diverse prediction strategies, but limited temporal and spatial representation remains. TEHL identifies environmental states, Node2Vec captures connectivity, GCNN learns neighbourhood relationships, and preprocessing improves consistency. Broader validation using diverse datasets and environmental conditions is required. Future work should include larger multimodal datasets, assess robustness to imbalance, and compare advanced graph-learning models to improve generalizability and scalability.

## 6. Conclusion

The proposed TEHL-GCNN model integrates multimodal environmental observations from carbon nanocomposite sensors, THz measurements, and satellite-derived data for intelligent environmental-state classification. SG filtering and Z-score normalization provide consistent feature representations, while TEHL identifies stable, transitional, and abnormal environmental conditions using temporal variation, entropy, and thresholds. Node2Vec captures spatial relationships, and the GCNN learns neighbourhood dependencies among distributed monitoring locations. The proposed model achieves 96.86% Accuracy, 96.72% Precision, 96.58% Recall, and 96.65% F1-score, demonstrating its effectiveness in accurately classifying heterogeneous environmental conditions.

## References

1 Kalasin, S. and Surareungchai, W., 2023. Challenges of emerging wearable sensors for remote monitoring toward telemedicine healthcare. Analytical chemistry, 95(3), pp.1773-1784. https://doi.org/10.1021/acs.analchem.2c02642

2 Pei, Z., Balitskiy, M., Thalman, R. and Kelly, K.E., 2023. Laboratory performance evaluation of a low-cost electrochemical formaldehyde sensor. Sensors, 23(17), p.7444. https://doi.org/10.3390/s23177444

3 Mai, X., Li, Q., Xu, W., Deng, S., Wang, W., Wu, W., Zhang, W. and Wang, Y., 2025. Estimation of mangrove aboveground carbon using integrated UAV-LiDAR and satellite data. Sustainability, 17(18), p.8211. https://doi.org/10.3390/su17188211

4 Helal, S., Sarieddeen, H., Dahrouj, H., Al-Naffouri, T.Y. and Alouini, M.S., 2022. Signal processing and machine learning techniques for terahertz sensing: An overview. IEEE Signal Processing Magazine, 39(5), pp.42-62. https://doi.org/10.1109/MSP.2022.3183808

5 Amraoui, Y., Halkhams, I., El Alami, R., Jamil, M.O. and Qjidaa, H., 2026. Application of Machine Learning for Accurate Prediction of Graphene-Based Terahertz Metamaterial Sensor Performance. Materials Today Nano,35, p.100911. https://doi.org/10.1016/j.mtnano.2026.100911

6 Srisabarimani, K., Dhivya, R., Srihari, T., Parthasarathy, P. and Arun Kumar, U., 2026. Numerical Modeling and Machine-Learning Optimization of a BaTiO3–Graphene–Ag THz Metasurface for Refractive-Index Biosensing Towards SARS-CoV-2 Detection. Sensing and Imaging, 27(1), p.121.https://doi.org/10.1007/s11220-026-00814-7

7 Yu, J., Pu, H. and Sun, D.W., 2025. Dual-band terahertz metamaterial sensor integrated with deep learning for synergistic identification of red wine varieties. Chemical Engineering Journal, 520, p.166006. https://doi.org/10.1016/j.cej.2025.166006

8 Yu, J., Pu, H. and Sun, D.W., 2025. An advanced deep learning-driven Terahertz metamaterial sensor for distinguishing different red wines. Chemical Engineering Journal, 504, p.158177.504, https://doi.org/10.1016/j.cej.2024.158177

9 Baz, A., Wekalao, J., Mandela, N. and Patel, S.K., 2024. Design and performance evaluation of machine learning-based terahertz metasurface chemical sensor. IEEE Transactions on NanoBioscience, 24(2), pp.128-135. https://doi.org/10.1109/TNB.2024.3453372

10 Gezimati, M. and Singh, G., 2024. Terahertz data extraction and analysis based on deep learning techniques for emerging applications. IEEE Access, 12, pp.21174-21198. https://doi.org/10.1109/ACCESS.2024.3360930
11 Dhandapani, G., M, R.R., V, K. and S, A.J., 2025. High-performance terahertz metasurface biosensor based on graphene-enhanced gold-silver composite architecture: machine learning-assisted design optimization and refractive index sensing. Journal of The Electrochemical Society, 172(7), p.077509. https://doi.org/10.1149/1945-7111/adeb34
12 Wekalao, J., 2026. Development of a high-sensitivity graphene-metal nanocomposite metasurfaces biosensor with machine learning integration for rapid detection of waterborne pathogens with behaviour prediction leveraging stacking ensemble regressor. Plasmonics, 21(1), pp.55-75. https://doi.org/10.1007/s11468-024-02756-w
13 Vijayakumar, K., Kaliaperumal, K., Sasikala, S., Anandaram, H. and Kumar, U.A., 2025. A multilayer $WSe_2$–graphene plasmonic architecture for polarization-insensitive and machine learning–optimized glucose detection in the terahertz regime. Sensing and Bio-Sensing Research, 50, p.100917. https://doi.org/10.1016/j.sbsr.2025.100917
14 Alqhtani, H.A., Almawgani, A.H., Gumaih, H.S., Adam, Y.S., Mehaney, A., Elsayed, H.A., Kabarokole, P., Wekalao, J. and Muheki, J., 2026. Machine learning-assisted terahertz plasmonic biosensor with multi-material architecture for rapid bacterial water quality monitoring. Journal of Materials Science, 61(30), pp.21941-21970.https://doi.org/10.1007/s10853-026-13089-w
15 Elsayed, H.A., Solouma, E., Messaoudi, M., Bin-Jumah, M., Wekalao, J., Khan, M.A., Rajakannu, A. and Mehaney, A., 2026. Machine-learning-assisted validation of a graphene-based terahertz metasurface for simulated hemoglobin sensing. Materials Technology, 41(1), p.2629414. https://doi.org/10.1080/10667857.2026.2629414
16 Muheki, J., Ahmed, A.M., Ali, M.K.M., Elsayed, H.A., Kabarokole, P., Wekalao, J., Mehaney, A. and Rajakannu, A., 2026. Machine Learning–Enhanced Tunable Terahertz Metasurface Sensor with a Hybrid Multi-Resonator Architecture for High-Sensitivity Amino Acid Detection. Optical and Quantum Electronics, 58(3), p.146.https://doi.org/10.1007/s11082-026-08727-1
17 Yu, W., Long, Y., Luo, J., Zhou, D., Fan, C., Ren, Y., Shi, J., Huang, Q. and Yang, X., 2025. Machine learning-assisted terahertz metagrating biosensor for label-free bacterial identification based on spectral fingerprinting. Analytical Chemistry, 97(40), pp.22164-22173. https://doi.org/10.1021/acs.analchem.5c04115
18 Patel, S.K., Surve, J., Katkar, V., Parmar, J., Al-Zahrani, F.A., Ahmed, K. and Bui, F.M., 2022. Encoding and tuning of THz metasurface-based refractive index sensor with behavior prediction using XGBoost regressor. IEEE Access, 10, pp.24797-24814. https://doi.org/10.1109/ACCESS.2022.3154386
19 Zhao, J., Yang, X. and Zhang, C., 2024. Vehicle trajectory reconstruction for intersections: an integrated wavelet transform and Savitzky-Golay filter approach. Transportmetrica A: transport science, 20(2), p.2163207. https://doi.org/10.1080/23249935.2022.2163207
20 Yaro, A.S., Maly, F. and Prazak, P., 2023. Outlier detection in time-series receive signal strength observation using Z-score method with S n scale estimator for indoor localization. Applied Sciences, 13(6), p.3900. https://doi.org/10.3390/app13063900
21 Kumar, A., Jain, D.K., Mallik, A. and Kumar, S., 2024. Modified node2vec and attention based fusion model for next POI recommendation. Information Fusion, 101, p.101998.https://doi.org/10.1016/j.inffus.2023.101998
22 Ding, H., Hou, H., Wang, L., Cui, X., Yu, W. and Wilson, D.I., 2025. Application of convolutional neural networks and recurrent neural networks in food safety. Foods, 14(2), p.247.https://doi.org/10.3390/foods14020247
23 Govindan, Nagarajan, and Minu Rajasekharan Indra. "Quantum deep learning assisted carbon quantum dot sensors for classification of emerging organic pollutants in air quality monitoring." Microchemical Journal 221 (2026):https://doi.org/ 116703. 10.1016/j.microc.2025.116703
24 Reddy, Thigulla Sampath, and G. Nagarajan. "Deep learning–driven aerial image processing framework for high-accuracy topographic mapping and automated road construction quality assessment." Microchemical Journal (2025): 115685. https://doi.org/ 10.1016/j.microc.2025.115685
25 Reddy, Thigulla Sampath, and G. Nagarajan. "Design Guidelines for the Geodetic Aerial Image Data Analysis Strategies Using Deep Learning." In 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), pp. 1539-1546. IEEE, 2025.