International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(11s) 2026 pp. 154-168
https://doi.org/10.70917/ijcisim-2026-2453
Article

DOMAIN ADAPTIVE VERMICOMPOST QUALITY PREDICTION THROUGH CALIBRATED MULTI-MODAL LEARNING AND INTERNET OF THINGS

J.Devika[1] , P.Srimanchari[2]

[1]Department of Computer Science, Erode arts and science College(Autonomous),Erode,Tamilnadu, India.
Email: devikasaba@gmail.com
[2]Department Computer Science,Erode arts and science College(Autonomous), Erode,Tamilnadu, India.
Email: srimanchari@gmail.com

ABSTRACT: Vermicomposting recycles organic wastes into nutrient-rich amendments that will have a market value and agronomic performance that depends upon the consistency of quality. Producers require quick, objective measures of maturity and safety in order to schedule harvests and certify batches. However, field labs are expensive and slow, single-indicator proxies (e.g., temperature decay) are not reliable across feedstocks, and most ML studies involve poor datasets with poor cross-site generalization. There is a gap for a low-cost, multi-modal, and transferable predictor that works in actual plants. To overcome such a gap, in this study, VERMI-Quality Network (VERMIQualNet) is suggested, which is a multi-modal and domain-adaptive real-time learner that integrates Internet of Things (IoT) with low costs, Near Infrared/Red-Green-Blue (NIR/RGB) spectrum, and texture/color features based on the bin images using self-supervised encoders and the gradient boosted multi-task head. The model is calibrated using uncertainty-aware temperature scaling and batch-wise transfer to output (i) Maturity Index, (ii) Contamination Risk, (iii) Nutrient Score, and in addition, Feedstock-Aware Pattern Mining (FAPM) is also integrated and extracts cross-batch association rules, which identifies hidden co-evolution trends between maturity and contamination. Interpretability is given through SHAP to reveal drivers (e.g., humification signal). Methodological core processes include cross-feedstock, data acquisition, cross-season, signal cleaning, and feature learning by contrastive pretraining. The processes are then ultimately deployed as an edge deployable inference service.
Keywords:- vermicompost, transparency, mining, machine learning, training, scaling, bin, images, features, multi-modal, IoT

1. INTRODUCTION

Vermicomposting is a method that converts organic waste to nutrient-rich compost using earthworms. Of course, it is eco-friendly, cost-effective, and generally used in agriculture and waste management [1]. These benefits are propelled by its capacity to enhance the fertility of the soil [2]. Yet, knowledge of the quality and maturity of vermicompost is of essential critical importance since it influences, to a certain degree, a plant's health and the capacity of vermicompost as a soil amendment. The traditional methods of confirming the compost quality include manual time-consuming, and sometimes inconsistent determination, such as temperature, Carbon-Nitrogen ratio, and Electrical Conductivity (EC) [3].

As Machine Learning (ML) and Artificial Intelligence (AI) experience growth, it has become a potential for researchers to simulate in-depth investigative-based studies that involve complex organism interactions, such as composting, by using computational techniques [4]. Among these applications, ensemble models and deep learning models are being applied to predict nutrient content, compost maturity, and to even recognize any anomalies. New model interpretability tools like SHapley Additive exPlanations (SHAP) [5] make sure that all those hidden operational and influential facts of the models can be understood, and they supply specialists with domain wisdom. However, most existing models have been built for a specific task and are not very workable with feedstock and varying environmental conditions.

Current models suffer from a number of problems, such as finding it difficult to generalize to different conditions of inputs, which creates cross-site inconsistency, and also resist following the uncertainty in their prediction. Further, the existing models do not integrate the sensor data from IoT infra and physicochemical data into a model. They further lack feature-level interpretability consistent with biophysical indicators such as EC decay, humification, and entropy [6]. These gaps call for the need for strong and scalable multi-modal input-output models that can deliver predictable and interpretable results.

1.1 Research Novelty

To overcome these problems, this study presents a new deep learning approach named VERMIQualNet which uses the thermal, spectral, and physicochemical data to predict vermicompost quality and maturity. Its main innovations are a Gradient Boosted Multi Task Head (GBMTH) to get uncertainty calibrated predictions and SHAP (interpretability), which maps learned features to actual composting phenomena. The model is also supported by contrastive learning to align representations between different batches and compost sources and enhance the generalizability of the models.

1.2 Scope and Motivation

VERMIQualNet allows the synergy of both laboratory tests of compost maturity and real-time data-based decision support. The research includes environmental sustainability, agricultural engineering, and AI. Owing to the ever-growing need for scalable, trustworthy quality tools for compost monitoring, the prime purpose of this work is to couple AI monitoring systems with farm-scale vermicomposting units. This integration would allow alerts in real-time on risk of contamination, batch readiness, and feedstock adjustment.

1.3 Research Objectives

The prominent goals of this research study are described as follows:

- Developing an effective multi-modal learning model, VERMIQualNet, that would predict maturity indices of composts and classify batch status either as immature or ready.
- Accumulation of uncertainty calibration and interpretability methods (GBMTH and SHAP) in order to increase the reliability of the model and biophysical interpretability.
- Performing a cross-site generalizability test and assessing the performance of VERMIQualNets against the baseline ML counterparts using a variety of measures.

2. LITERATURE RIVIEW

Authors in [7] used Additive Bayesian Network (ABN) modeling and exact structure discovery with 10,000 parametric bootstraps, to study 68 manure-based vermicompost. They identified end product quality driving factors. The principal part of the study involved analysis of manure type, co-composting with green waste, and stabilization with/without earthworms during thermophilic and stabilization phases. Results associated an initial quality of the blend with the concentration effect determined by a mass loss. The losses were reported to be an average of about 50% in mass and 60% in total carbon. Mean nutrient losses were 33% for nitrogen, 19% for phosphorus, and 27% for potassium. One of the more serious limitations is the external validity: the statistically robust ABN cannot be used outside the experimental range. Just like any statistical model, it is also subject to measurement error.

Investigators in [8] have performed a randomized complete block field experiment to compare wheat response to vermicompost prepared from cow dung, paper, and rice straw along with NPK fertilizer. The experimental study obtained a harvest index of 41.32% and a grain of higher quality. The authors point out the necessity of testing vermicompost in the absence of inorganic fertilizer and the necessity of elucidating the biocontrol of aphids.

The work in [9] studied the nutrient recovery during vermicomposting of municipal solid waste. They compared Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN) using seven input variables on 4 waste-to-compost ratio treatments measured for 100 days. ANN performance was better comparisons to MLR, as the R^2 level of ~99.91% and ~99.83% were obtained for Total Phosphorus (TP) and Total Nitrogen (TN), respectively, while it was 72.9% and 83.4% for MLR. Vermicomposting infiltrates TP and TN in final products by about 1.6 and 1.5 times, respectively. The limitations of the study include the narrow scope-the results can only be generalized to the actual ratios and measured indices used in the study.

Authors in [10] conducted controlled storage for 4 weeks experiment on cow -dung vermicompost. The results of an experiment comparing fresh material to pre-item (air-dried material in open polythene bags), together with weekly physicochemical experiments, were statistically compared. Major essential results indicated efficient retention of total organic constituents. The optimal nutrient quality was shown for pre-drying vermicompost stored in holed bags. The study found moisture content between 30% to 50% and 20% to 30% in fresh and pre-dried material, respectively, and found the optimal moisture content was the pre-dried/holed condition. Major limitations are the short time duration of 4 weeks, constrained feedstocks, and specific bag designs.

Researchers in [11] conducted a vermicomposting experiment using Perionyx excavates through completely randomized design method utilizing chicken, buffalo, cow, and goat manure at 24 units. The plants were fertilized once a week for 60 days, and standard analyses were used to measure physicochemical and organoleptic parameters. The protocol retrieved the manures to 50-60% moisture content, stored the trays under ambient conditions, and claimed that the nine parameters which were assessed all reached the predicted water content of about 50-53%. Limitations include checking for just 60 days, and there is an internal inconsistency in the reporting of pH values of 7.90-8.96, which is outside the claimed compliance range of 6.80-7.49.

Researchers in [12] proposed a Random Forest-based full-reference Image Quality Assessment (RF-IQA) for the vermicompost bioengineering quality monitoring. Eleven feature-similarity descriptors were computed for the reference-test images, and a Random Forest Regressor was used to predict quality scores. The two major stages were achieved through building an 11-dimensional feature vector, mapping it to subjective scores through the regressor, and comparing the results with traditional methodologies. The Pearson linear correlation coefficient for RF-IQA was 92.36%, which was greater than the baseline, but its generalization over the study conditions could be low.

In the recent work [13], investigators applied the Random Forest Regression (RFR) technique as a ML method to forecast nitrate leaching in soils amended by sugarcane bagasse-derived biosolids (biochar and vermicompost). The study first simulated leaching based on nine soil physicochemical variables, then reduced the modeling to a combination of four meaningful predictors, including cation exchange capacity, anion exchange capacity, EC, and bulk density. The predictions of the RFR model showed a high accuracy: R2 of 98.37 percent with all nine variables and 94.41% with only four. These results are considerably greater than those from linear regression. The main disadvantage of the model is that it relies on the availability of sufficient training data and attention to the model's parameters, and is less interpretable than traditional models.

Authors in [14] proposed an Edge-based Artificial Intelligence-Internet of Things (Edge-AIoT) recommendation platform, based on bio-inspired optimization and Long Short-Term Memory/Gaussian Process (LSTMs/GP) based predictions for the control of the humidity, thermometer, and pH levels in seven trials based on vermicomposting cultivation. Edge control and sensing improved humus productivity from 37.58% to 87.88% and raised the number of worms by 83%. However, the number of trials carried out was limited, and device- level computation was very constrained.

3. METHODOLOGY

The VERMIQualNet framework is envisioned to provide a transferable and cost-effective update of an actual-time predictive instrument to gauge vermicompost maturity, nutrient status, and poison hazard. By incorporating heterogeneous sensory information from environmental measurements (Relative Humidity (RH), moisture (m), pH, Electrical Conductivity (EC)) along with environmental spectra data (RGB and NIR), and visual texture descriptors into a multi-modal self-supervised learning framework, the framework is able to robustly generalize on different operational sites and feedstock instances. The operational workflow is then split into four successive stages. They are,

- Cross-Feedstock Data Acquiring and Signal Conditioning
- Multi-Modal Feature Encoding
- Domain Adapting Multi-Task Learning and Calibration
- Feedstock Cognizant Pattern Mining and Edge Installation

These stages combine coherently into a closed-loop pipeline, which tends to acquire essential data for the creation of decision-ready predictions, therefore, contributing to make the compost quality monitoring in real industrial conditions reliable and interpretable. Figure 1 illustrates the overall and generalized architecture of VERMIQualNet.

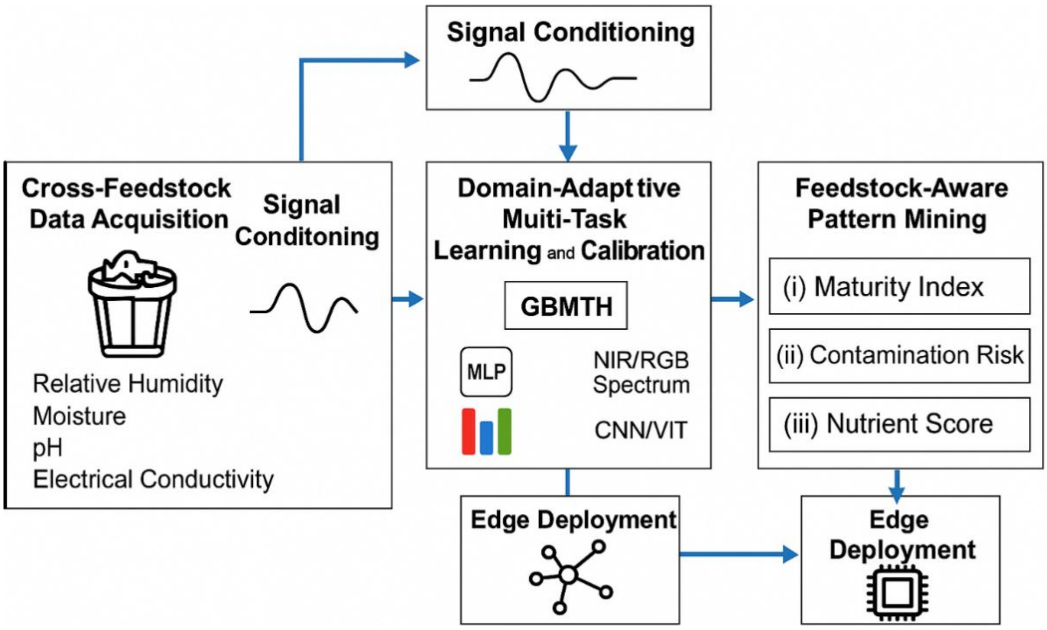

Figure 1. Generalized Architecture of VERMIQualNet

3.1 Cross-Feedstock Data Acquiring and Signal Conditioning

During this first phase, the following essential data such as raw signals from low-cost sensors and imaging modules measuring a number of different vermicompost bins, are collected on a constant basis across various feedstocks mixtures, and over seasonal variations. The primary difficulty lies in the heterogeneous and noisy data, and thus, the data need to be denoised, normalized, and synchronized. The sensor measurements like pH, EC, and other metrics via IoT infra are mapped with the corresponding spectral and image features over time, as well as outliers are removed using unique techniques from robust statistics. By combining both spatial (image) and temporal (sensor) information, each sample aggregator window is time-stamped, thus reflecting the state of composting.

The core computation of this phase involves three major multi-modal observations at varying temporal aspect, *t*. They are environmental vectors, $i_t^{(m,RH,pH,EC)} \in \mathbb{R}_t^{env}$, spectral intensities, $i_t^{(RGB/NIR)} \in i_t^{spec}$, and compost bin visuals, $i_t^{(img)} \in \mathbb{R}_t^{(w \times h \times RGB)}$. Thus, the overall multi-modal information can be expressed as in eq. (1).

$$\mathbb{I}_t = \left\{ i_t^{(m,RH,pH,EC)}, i_t^{(RGB/NIR)}, i_t^{(img)} \right\} \tag{1}$$

Signal cleaning is performed using z-score normalization and median filtering, which is computed as:

$$\tilde{\imath}_{(t,x)} = \left[\left(i_{(t,x)}\right) - (\mu_x) \Big/ \sigma_x \right], \quad where\ \forall x \in features \tag{2}$$

Temporal synchronization is based on the computation of sliding window averaging, given as:

$$\tilde{\mathbb{I}}_k = (\Delta t)^{-1} \cdot \sum_{t \in (k - \Delta t, k)} \tilde{\mathbb{I}}_t \tag{3}$$

In eq. (3), Δt represents the window size that determines temporal granularity, which ensures that the resulting unified feature vector $\tilde{\mathbb{I}}_k$ of each batch instance is modality consistent.

3.2 Multi-Modal Feature Encoding

Self-supervised associated contrastive pretraining is utilized to learn common representations from unlabeled data, allowing contradictory generalizations that are available for assorted feedstocks and environments. In particular, the three modality-specific encoders, namely, the spectra CNN encoder, the image ViT encoder, and the sensor signals MLP encoder are combinedly trained.

Contrastive learning draws associated sensor-spectral-image triplets close to each other in latent space, whereas it pushes unrelated examples away from each other. Consequently, the latent embedding (L_ε) is able is incorporate essential dynamics of composting, additionally, humification and moisture decay, in the absence of manually applied labels. Here, N number of samples (each) associated with paired modalities, $\left\{\mathbb{I}_i^{(m,RH,pH,EC)}, \mathbb{I}_i^{(RGB/NIR)}, \mathbb{I}_i^{(img)}\right\}$, where the associated encoders are represented as in eq. (4) - (6).

$$MLP_{encoder}\text{: } L_{\varepsilon_i}^{(m,RH,pH,EC)} = f_{\rho_e}\left(\mathbb{I}_i^{(m,RH,pH,EC)}\right) \tag{4}$$

$$CNN_{encoder}\text{: } L_{\varepsilon_i}^{(RGB/NIR)} = f_{\rho_s}\left(\mathbb{I}_i^{(RGB/NIR)}\right) \tag{5}$$

$$ViT_{encoder}\text{: } L_{\varepsilon_i}^{(img)} = f_{\rho_v}\left(\mathbb{I}_i^{(img)}\right) \tag{6}$$

In eq. (4), (5), and (6), e, s, and v represent the environmental, spectral, and visual modalities, respectively. f and θ indicates the encoder functions and parameters of the corresponding modalities, respectively. Further, the computation of multi-modal fusion is performed through construction of a multi-modal joint embedding and expressed as,

$$L_{\varepsilon_i} = \left(L_{\varepsilon_i}^{(m,RH,pH,EC)} || L_{\varepsilon_i}^{\left(\frac{RGB}{NIR}\right)} || L_{\varepsilon_i}^{(img)}\right) \cdot |\delta_F| \tag{7}$$

In eq. (7), $|\delta_F|$ signifies the feature projection matrix. The computation of contrastive loss (l_{cont}) is essential since it attracts similar (positive) pairs and repels dissimilar (negative) pairs in embedding space, and also enables the model to acquire discriminative multimodal invariant features in an unsupervised manner.

$$l_{cont} = -\sum_i \log \frac{exp\left(\mathrm{C}_{sim}\left(L_{\varepsilon_i}, L_{\varepsilon_i}^{+}\right)/\mathrm{T}\right)}{\sum_j xp\left(\mathrm{C}_{sim}\left(L_{\varepsilon_i}, L_{\varepsilon_i}^{-}\right)/\mathrm{T}\right)} \cdot (N^{-1}) \tag{8}$$

In eq. (8), $\mathrm{C}_{sim}(x, y)$ denotes the cosine similarity, T indicates the temperature parameters, $L_{\varepsilon_i}^{+}$ and $L_{\varepsilon_i}^{-}$ reflects the same batch samples (positive), and different batch samples (negative), respectively. Thus, the system guarantees that all the samples from the same batch (intra) are alike and all the samples from different batches (inter) are differentiated with respect to their composting state.

3.3 Domain Adapting Multi-Task Learning and Calibration

During this stage, the pretrained encoders are fine-tuned on the labeled data using three regression tasks, namely, Maturity Index (α or MI), Contamination Risk (β or CR), Nutrient Score (γ or NS), and target variables like Composite Status (ψ or CS). Further, the approach includes a Gradient Boosted Multi-Task Head (GBMTH), which identifies non-linear patterns that are common to multiple quality indicators and treats dependencies between tasks in an effective way. In GBMTH, both learning non-linear mappings and sharable latent factors among the tasks are used. The domain adaptation is adopted using Batch-Wise Transfer Normalization (BTN) to minimize the site-specific variances. Additionally, Temperature-Scaled Uncertainty-Calibration tunes the confidence levels of the model to ensure that it is a reliable model when in operation. For the given encoded features L_{ε_i} and task outcome $y_i = \left\{y_i^{\alpha}, y_i^{\beta}, y_i^{\gamma}\right\}$, GBMTH is computed as,

$$\tilde{Y}_i = \mathcal{G}_{\varphi_k}\left(L_{\varepsilon_i}\right) \Rightarrow \left\{\mathcal{G}_{\varphi_1}\left(L_{\varepsilon_i}\right), \mathcal{G}_{\varphi_2}\left(L_{\varepsilon_i}\right), \mathcal{G}_{\varphi_3}\left(L_{\varepsilon_i}\right)\right\} \tag{9}$$

In eq. (9), $\mathcal{G}_{\varphi_k}$ indicate the gradient-boosted regressive operator along with shared base-learners. Further, the uncertainty weighting loss is computed as,

$$l_{uw} = \sum_{k=1}^{3}\left(\log(\delta_k) + \left(\left\|Y_i^{(k)} - \tilde{Y}_i^{(k)}\right\|^2\right) \Big/ (2 \times \delta_k^2)\right) \tag{10}$$

In eq. (10), learnable parameter for task uncertainty is referred as δ_k. Then, domain adaptation, domain transfer normalization (batch-wise) is computed as,

$$\tilde{L}_{\varepsilon_i} = \S\left(\frac{L_{\varepsilon_i} - \mu_{SD}}{\sigma_{SD}}\right) + \zeta \tag{11}$$

In eq. (11), $\S$ is a scale parameter which controls the amplitude of the normalized features, ζ is the shift parameter, which re-focusses the distribution of normalized features. Both $\S$ and ζ allow the model to update feature statistics for the BTN during the domain adaptation process (per target feedstock domain). Moreover, μ_{SD} signifies the mean of the source domain L_{ε_i}, that is, the site or training feedstock. Besides, σ_{SD} represents the variance of the source latent variables in the source domain. Collectively, μ_{SD} and σ_{SD} generate the statistical distribution of the source data as used to normalize and adapt the features to a new target domain.

Temperature scaling is used to fine-tune the prediction confidence and avoid model being either over-confident or under-confident about its uncertainty when compared to the reality, during real-time usage.

$$\hat{\lambda}_i = softmax\left({\tilde{Y}_i}/{\bar{\mathrm{T}}}\right) \quad (12)$$

In eq. (12), $\hat{\lambda}_i$ is the calibrated probability distribution that is derived by applying temperature scaling to the model predicted temperatures. It translates the raw predictions, $\tilde{Y}_i$, into well-calibrated confidence scores for each of the prediction tasks. $\bar{\mathrm{T}}$ indicates the parameter calibration of temperature, which is >1 (tends to minimize the overconfidence, especially in prediction).

3.4 Feedstock Cognizant Pattern Mining and Edge Installation

Ultimately, FAPM (Feedstock-Aware Pattern Mining) uniquely extracts the cross-batch association rules which capture the joint patterns of maturating across batches and feedstock (*fs*) properties with respect to contamination, maturity, and nutrient. Using the mining of high-frequency item-sets and meaningful process trends, hidden dependencies in processes become explicit for significant processes like interpretability and optimization. The resulting inference models are published as an Edge AI Service, and allow for the on-site, real-time decision-making capability, decreasing the latency in the cloud environ as well as the data transfer.

In this case, for a given batch B, triplet results are produced as,

$$O_B = \{\alpha_B, \beta_B, \gamma_B, fs_B\} \quad (13)$$

Furthermore, these outcomes are discretized into categorical states for mining purposes as,

$$O'_B = f_Q[O_B, bins] \quad (14)$$

In eq. (14), f_Q refers the quantization function. For item-sets $\mathbb{A}$ and $\mathbb{B}$, support and confidence are, respectively.

Association rule mining is used to identify patterns that co-occur most frequently among variables, such as contamination, maturity, and feedstock type. It reveals unknown relationships and trends that guide an optimization of processes and enhance the ability to interpret the results. For item-sets ($\mathbb{A}, \mathbb{B} \subset O'_B$), both support and confidence can be computed as,

$$SUP(\mathbb{A} \Rightarrow \mathbb{B}) = |\{B: \mathbb{A} \cup \mathbb{B} \subset O'_B\}| / |\{B_1, B_2, \cdots, B_N\}| \quad (15)$$

$$CONF(\mathbb{A} \Rightarrow \mathbb{B}) = SUP(\mathbb{A} \Rightarrow \mathbb{B}) / SUP(\mathbb{A}) \quad (16)$$

In the study's context, from eq. (15 and (16), $\mathbb{A}$ is a set of observed conditions or states of features, namely, certain types of feedstocks, indicators of the composting stage, sensors, etc. $\mathbb{B}$ corresponds to the resulting compost quality results for a final product (mature compost, high nutrient score or low contamination risk). Rules having values above predetermined threshold values ($SUP_{min}, CONF_{min}$) are stored as feedstock sensitive trends, which map process features ($\mathbb{A}$) to quality results ($\mathbb{B}$) and expose interpretable trends in successive batches.

The final inference model (ξ^*) is serialized and hosted on micro-controller-compatible hardware, which can be computationally represented as,

$$\xi^*: \mathbb{I}_t \mapsto \left\{y_i^\alpha, y_i^\beta, y_i^\gamma\right\} \quad (17)$$

The locally interpretable model, SHapley Additive exPlanations (SHAP) is used to characterize the predictions, and its discrimination power is provided for each output, in terms of features (EC, m, pH, or color texture) that were the most informative. This feature-level interpretability makes the links between the patterns learned in the model and the biophysical composting processes in the real-world. It indicates that a higher humification signal or a consistent decline in EC is a strong maturity index. In short, SHAP performed the explicit task of explaining why the model had made a certain prediction or, as is the case with these problems, how data-driven understandings relate to more fundamental biological and chemical processes of composting. Table 1 represent the significant procedures of VERMIQualNet in the quality prediction.

Table 1. Algorithmic Processes of VERMIQualNet

Input: $\mathbb{I}_t = \left\{i_t^{(m,RH,pH,EC)}, i_t^{(RGB/NIR)}, i_t^{(img)}\right\}$
Output: $y_i = \left\{y_i^\alpha, y_i^\beta, y_i^\gamma\right\}$
BEGIN

1: Data Acquisition and Conditioning

1.1: Accumulate $\mathbb{I}_t$;

1.2: Perform Cleaning and Normalization

$$\tilde{\iota}_{(t,x)} = \left[\left(i_{(t,x)}\right) - \left(\mu_x\right) \Big/ \sigma_x \right]$$

1.3: Perform Aggregation

$$\tilde{\mathbb{I}}_k = (\Delta t)^{-1} \cdot \sum_{t \in (k - \Delta t, k)} \tilde{\mathbb{I}}_t \qquad \textit{//sliding window averaging}$$

2: Self-supervised Encoding

2.1: Perform Encoding via MLP, CNN, ViT

$$L_{\varepsilon_i}^{(m,RH,pH,EC)} = f_{\rho_e}\left(\mathbb{I}_i^{(m,RH,pH,EC)}\right)$$

$$L_{\varepsilon_i}^{(RGB/NIR)} = f_{\rho_s}\left(\mathbb{I}_i^{(RGB/NIR)}\right)$$

$$L_{\varepsilon_i}^{(img)} = f_{\rho_v}\left(\mathbb{I}_i^{(img)}\right)$$

2.2: Fuse L_{ε_i}

$$\Rightarrow \left(L_{\varepsilon_i}^{(m,RH,pH,EC)} || L_{\varepsilon_i}^{\left(\frac{RGB}{NIR}\right)} || L_{\varepsilon_i}^{(img)} \right) \cdot |\delta_F|$$

2.3: Compute Contrastive Loss, (l_{cont})

$$\Rightarrow - \sum_i \log \frac{exp\left(\mathrm{C}_{sim}\left(L_{\varepsilon_i}, L_{\varepsilon_i}^{+}\right)/\mathrm{T}\right)}{\sum_j xp\left(\mathrm{C}_{sim}\left(L_{\varepsilon_i}, L_{\varepsilon_i}^{-}\right)/\mathrm{T}\right)} \cdot (N^{-1})$$

3: Multi-Task Learning and Calibration

3.1: Predict $\tilde{Y}_i = \mathcal{G}_{\varphi_k}\left(L_{\varepsilon_i}\right)$

3.2: Compute Uncertainty Weighting Loss, (l_{uw})

$$\Rightarrow \sum_{k=1}^{3} \left(\log(\delta_k) + \left(\left\| Y_i^{(k)} - \tilde{Y}_i^{(k)} \right\|^2 \right) \Big/ \left(2 \times \delta_k^2\right) \right)$$

3.3: Perform Domain Adaptation, $\tilde{L}_{\varepsilon_i} = \varsigma\left(\frac{L_{\varepsilon_i} - \mu_{SD}}{\sigma_{SD}}\right) + \zeta$

3.4: Calibrate, $\hat{\lambda}_i = softmax\left(\tilde{Y}_i \big/ \mathrm{T}\right)$

4: Feedstock-Aware Pattern Mining and Deployment

4.1: Discretize $O_B' = f_Q[O_B, bins]$

4.2: Apply mining rules

$SUP(\mathbb{A} \Rightarrow \mathbb{B})$;

$CONF(\mathbb{A} \Rightarrow \mathbb{B})$;

4.3: Deploy $\xi^*: \mathbb{I}_t$

5: Perform Interpretation via (local) SHAP

END

The VERMI-QualNet predictor classes are presented in Table 2. They are divided into three general dimensions with further composite classes existing within these groups. Each class is characterized by essential features or process signals that show the quality state of the compost.

Table 2. Predictor Classes Potentially Generated by VERMIQualNet

Dimensions	Class Label	Indicators
MI	Immature	Low humification, high temp, high EC
	Semi-mature	Partial humification, moderate temp
	Mature	Stable temp, balanced pH, dark color
	Over-mature	Low EC, nutrient loss
CR	High Risk	Plastics, pathogens, metals
	Moderate Risk	Partial sanitization, minor impurities
	Low Risk	Clean, uniform texture, stable EC/pH
NS	Low Nutrient Value	Low organic C, low NPK,
	Medium Nutrient Value	Moderate EC, Balanced NPK
	High Nutrient Value	Strong humus signal, high NPK
CS	Premium Compost	Mature + Low CR + High NS
	Unsafe Compost	Immature + High CR
	Nutrient-Deficient Compost	Mature + Low NS

4. Resource Utilized

The test environment and VERMI-QualNet analysis were performed on a high-performance workstation. It is comprised of an Intel i9-13900K CPU (operating at 3.0 GHz, with 24 cores), 64 GB of RAM, and an RTX-4090 GPU from NVIDIA (operating at 24 GB of memory), running Ubuntu 22.04 LTS.

Open-source tools were used in the implementation. Python v3.10 is used as the base programming language. For deep learning, the software was PyTorch v2.1.0, and for the metric of classification and evaluation features, the Python package of scikit-learn v1.3.0 was used. Gradient boosted multi-task modeling used eXtreme Gradient Boosting (XGBoost) v1.7.6, and feature extraction of images was performed using OpenCV v4.8.0.

Additional libraries (NumPy v1.25.0, Pandas v2.0.3, and Matplotlib v3.7.2) were utilized to continue the workflow for analysis, data preprocessing, and visualization. This integrated hardware-software stack ensured a great multi-modal learning, training functionality, high throughput ability, as well as stable inference for real-time vermicompost quality prediction.

The performance of the proposed VERMI-QualNet is compared to several well-established ML models like ANN, MLR, RF-IQA, RFR, ABN, LSTM/GP, in terms of generalization ability, predictive accuracy, and computational efficiency. The models which have been considered are: ANN model for nonlinear feature mapping, MLR as a basic statistical model, RF-IQA and RFR with Random-forest conceptual for ensemble-based feature interaction learning, ABN for modeling the probabilistic dependency of compost quality variables, and LSTM with the support of GP for modeling temporal and stochastic dependencies in compost time behavior. Section 2 of the manuscript contains a detailed methodological discussion of each model's applicability to the preceding parallelized data and application relevance.

4.1 Data Collection

For the evaluation and validation of the performance of the proposed VERMIQualNet, the publicly available VERMICOMPOST dataset from GitHub [15] was considered. The data in the data set is from a large number of sources and comprises a large number of data types. It was obtained from different vermicomposting units with different feedstock types (composition) experienced different weather conditions and operated in different seasonal regimes. It incorporates 10,000 composting cases within the scope of different domains of input into heterogeneous organic matter, such as agro-industrial by-products, agricultural residues, food and kitchen wastes, horticultural wastes, manures, and fibre/paper residues. These feedstocks render an optimal chemical and biological variability, resulting in sophisticated sensor and spectral responses which are helpful in the assessment of compost quality. Each instance of the data is a mixture of synchronized readings from low-cost IoT sensors.

Also, each dataset consists of synchronized measurements such as low-cost environmental sensors (humidity, EC, temperature, and pH) and optical sensors (reflectance in the NIR and RGB channels), as well as texture features derived from images (color variance and entropy). The integrated model represents the biophysical and biochemical evolution of composting (microbial evolution, humification, and nutrient stabilization) across various operational systems and feedstock composition. This integration captures the biophysical and biochemical changes that occur during composting, which are micro-organisms dynamics, humification changes, and nutrient stabilization, in the different types of operational setups and feedstock types.

The data presents high modal dependencies and natural noise, enabling the data set to be used as a benchmark tool for measuring cross-domain transferability. It has an appropriate combination of continuous and categorical properties representing the intrinsic complexity of in-field vermicomposting ecosystems. Through external validation, the dataset adds to the comparative strength of the proposed VERMIQualNet through the ability to evaluate using independent, heterogeneous conditions. This represents the scalability and transferability of the model between different operational and geographic operational contexts.

Because the data are heterogeneously distributed in nature, reflecting season-to-season variability, site-to-site variability, and feedstock composition variability in the field, the dataset is ideal for generalization. Domain diversity, the non-linear nature of feature space representations, and controlled stochasticity are beneficial for the learning of machine learning models to learn robust and transferable representations.

5. PERFORMANCE EVALUATION

A number of quantitative metrics were applied to quantitatively and rigorously assess the predictive performance, the classification reliability, and the cross-domains adaptability of the model. R-squared, also known as R^2, was used as a statistical measure of the model's ability to explain the variation of the response, or in the study's case, the Maturity Index (MI), compared with the acceptability predicted by the model. MAE showed the average size of prediction errors, which provide an optimal perception of the model’s precision, and its improvement as compared to the baseline methods. For unbalanced problems, especially for categorical classification (e.g., compost being "immature" or "ready"), the F1-score was considered as a balance act between recall and precision. AUROC was used to determine the discriminative power of the model to separate contaminated from safe compost. Additionally, cross-site generalization is considered to understand the adaptiveness of VERMI-QualNet when it is deployed on unexplored feedstock and site conditions. In addition, two additional crucial measures are also considered for evaluation. RMSE provided a measure of residual variation that was weighted according to sensitivity and tends to provide more weight to larger deviations. Calibration Error (CE) was used to check the quality of the model's confidence statements against actual results/events and to ensure reliable uncertainty estimates.

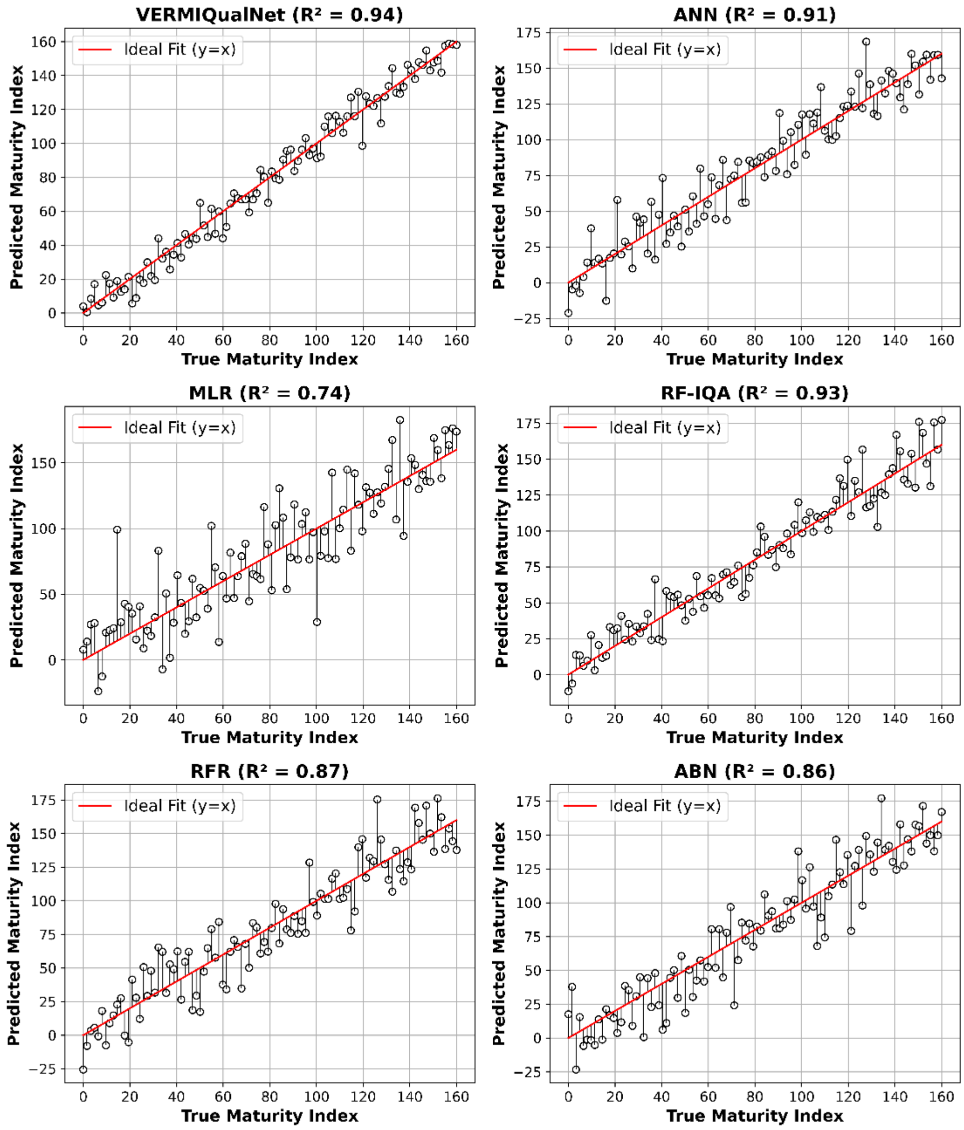

Figure 2. Comparative Regression Plots Exhibit the predicted values against the True Values of the Maturity Index for VERMIQualNet and Baseline Models; in addition, each plot includes Vertical Error Bars, as well as an Ideal Reference Line ($y = x$).

Figure 2 provides a side-by-side comparison of the effectiveness of VERMIQualNet and six baseline models for predicting the Maturity Index. The plot shows the true vs. the predicted values, R^2 scores, and the error bars. VERMIQualNet is at the front with $R^2 = 0.94$, showing that VERMIQualNet is very close to the ideal line $y = x$ with very minimal vertical deviation. This proves strong consistency and low bias. ANN and RF-IQA provided optimal results with R^2 equal to 0.91 and 0.93, respectively, but their predictions are more dispersed in the mid to high index areas. This tends to hint occasional under- or overpredictions. MLR, the weakest model ($R^2 = 0.74$), has poor correspondence and larger residuals, validating that its capture of nonlinear composting dynamics was poor. RFR ($R^2 = 0.87$) and ABN ($R^2 = 0.86$) fit moderately, though ABN fits slightly less well at higher-levels of maturity. Overall, the combination of tight alignment and a small error variance of the VERMIQualNet makes it very effective in modelling the highly complex nonlinear relationships that are required to model the maturity of vermicompost in a very short timeframe, particularly when varying feedstock conditions are employed.

Table 3. Comparison of VERMIQualNet and the baseline models is demonstrated in MAE for regression, F1-Score, and AUROC for classification, respectively.

Approaches	MAE (0–1)	F1-Score (%)	AUROC (%)
VERMIQualNet	0.061	96.3	92
ANN	0.085	91.5	87.4

MLR	0.104	83.2	78.3
RF-IQA	0.082	92.4	88.9
RFR	0.088	89.6	85.1
ABN	0.096	86.1	81.5
LSTM/GP	0.08	93	89.3

Table 3 shows a close comparison of the performance between VERMIQualNet and several standard baseline models for both regression and classification tasks. For the VERMIQualNet, the resulting outcome showed a lowest MAE (0.061), which is 27% less than the average baseline MAE of about 0.084. This improvement is a reflection of the ability of proposed approach to learn complex non-linear relationships using domain adaptation and multi-modal encoding. For 'immature vs. ready' classification, the maximum F1-Score at 96.3% was achieved with an excellent precision and recall, which implies a prominent reduction in misclassifications, particularly towards the compost maturity constraint area. Further, the outcome of AUROC at 92.0% indicates a significantly high discriminativeness of the developed model with respect to the production of accurate contamination-risk alerting across the variation of operational fixed implementation thresholds. Together, these results show the adaptiveness, robustness, and overall great performance of VERMIQualNet in comparison to conventional ML approaches in the field of vermicomposting.

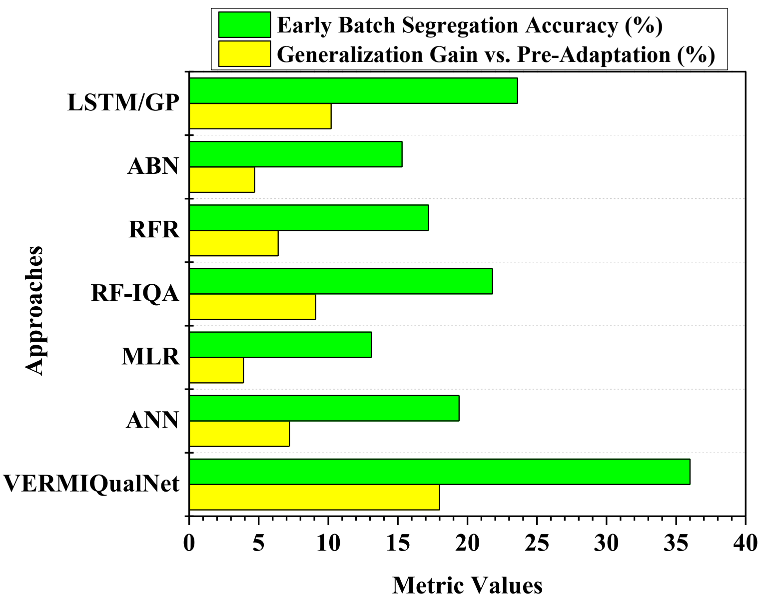

Figure 3. Performance of Cross-Site Generalization and Initial Batch Partitioning across various Approaches.

Figure 3 shows that VERMIQualNet effectively generalizes to both feedstocks and different sites. After domain adaptation, its cross-site performance improves by 18%. This achieves a benefit of batch-wise normalization, uncertainty-aware calibration, and modality fusion, which combinedly aids the model to identify transferable patterns and reduce possible model degradation due to variable composting conditions. As a result, VERMIQualNet is accurate in new environments, preventing the performance deterioration of base models. The model has an accuracy of 36% in early batch segregation, which is far better than the baseline average of about 26.3%, meaning this model has a good capacity to detect underperforming batches or immature experiments early in the process. This early detection is essential in operation, as it could enable rapid intervention to prevent waste of resources, save in turnaround times, and ensure a consistent quality of compost throughout production cycles.

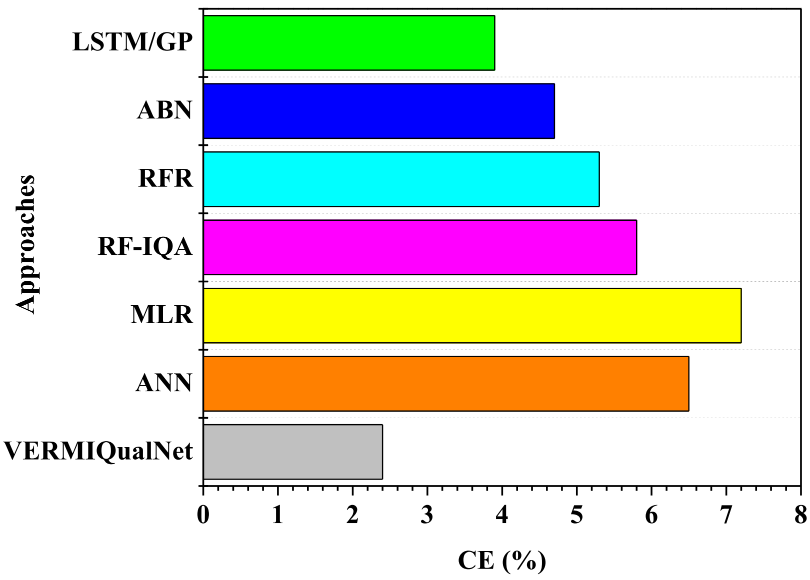

Figure 4. Calibration Error (CE) comparative Analysis between Predicted Confidence and Actual Outcomes for Predictive Models.

The CE results given in Figure 4 mark VERMI-QualNet's strong predictive accuracy on the estimation of the uncertainty. It reaches the lowest CE of 2.4% better than all other comparator models. The optimized calibration results in high correlation between the calculated confidence scores and actual values of the observations, so the model can be well applied in the real-time decision-making model for vermicompost quality control. The self-supervised encoders combined with the temperature scaling calibration allow VERMI-QualNet to learn consistent probability distributions across various feedstuffs and stages of composting and offer less overconfidence or under-confidence. On the other hand, higher CE is observed in MLR (7.2%) and ANN (6.5%) due to their poor handling of feature uncertainties and Zero Adaptive Calibration. Even probability-based models like ABN (4.7%), LSTM/GP (3.9%), and tree-based models (RF-IQA: 5.8%, RFR: 5.3%) are considered to be superior to conventional probabilistic models as they approach the performance level of fine-grained multi-task learning and domain adaptation by VERMIQualNet. The low CE ensures that any indication of risk, like readiness classification, contaminations, etc., will stay trustworthy and critical to the value the system for real-time deployment.

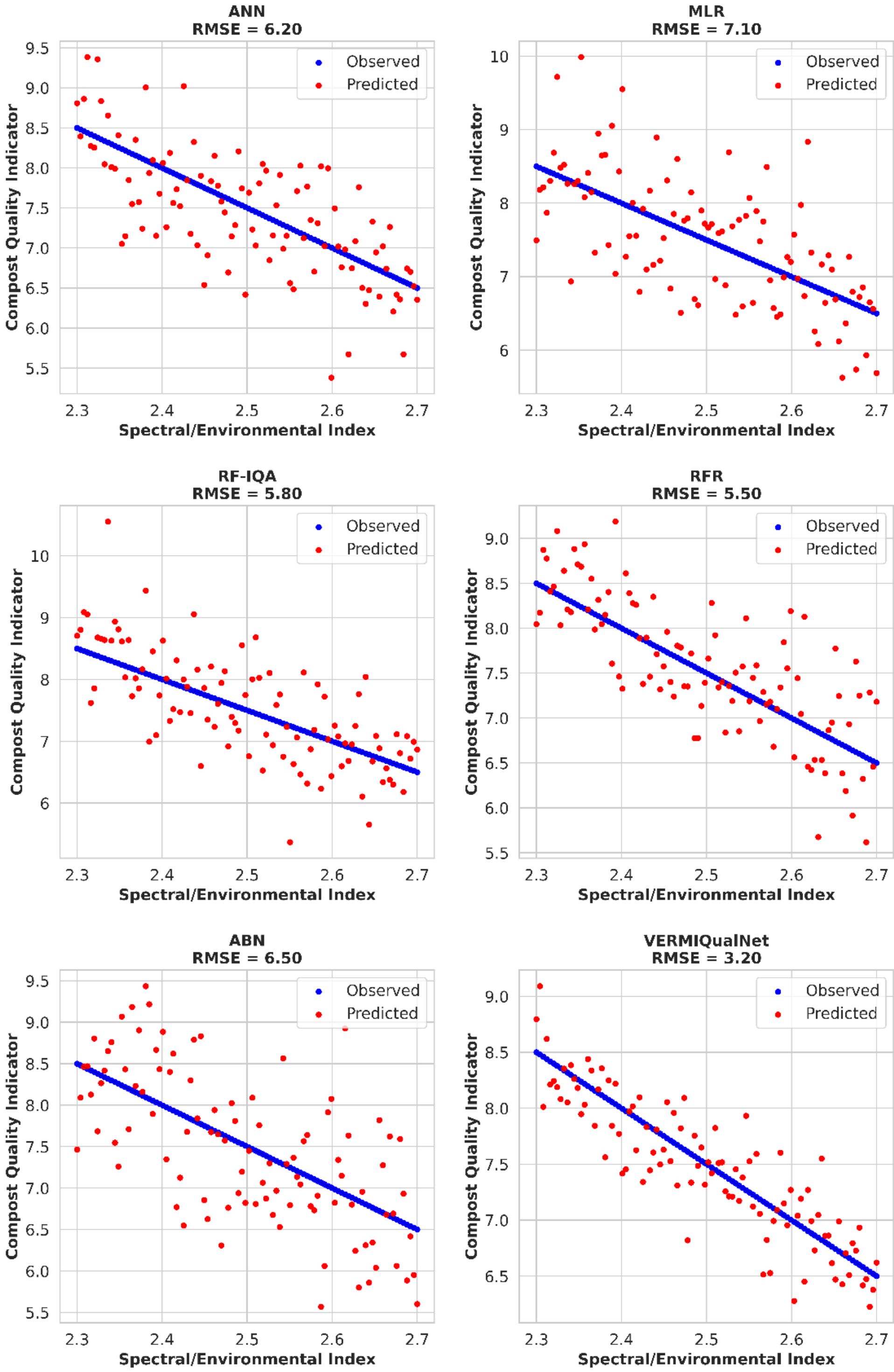

Figure 5. RMSE Comparison across Predictive Models of Compost Maturity Estimation.

Figure 5 shows the RMSE results for all the tested models, where VERMIQualNet performs better with the lowest RMSE value of 3.82 among all the existing models. This low RMSE suggests a small residual variance of observed compost maturity scores at the predicted compost maturity scores. The multi-modal encoder architecture of VERMIQualNet with its combination of sensor, spectral, and image-based features is a key part of its success in achieving its goals. By representation of complex nonlinear relationships, it is possible to adapt to the various feedstock conditions. In contrast, the traditional models such as MLR (RMSE = 6.45) and ANN (RMSE = 5.96) are exhibiting larger residual spread as they have inclined difficulties to design generalization over heterogeneous inputs. Embedded ensemble learning slightly improves on the RMSLE results of tree-based models like RF-IQA (5.35) and RFR (5.07). The ABN (4.76, graph) and LSTM/GP (4.25, graph) architectures lack recursive thought and temporal pattern learning effectiveness, but after all, still seem to be inferior to

VERMIQualNet. The results indicated that sensor fusion and calibration prediction layers of VERMIQualNet are effective in minimizing the high-magnitude errors and demonstrated deeper awareness of larger deviations, which is a key requirement to ensure safe deployment in the real-world context of VERMIQualNet for vermicompost quality assessment.

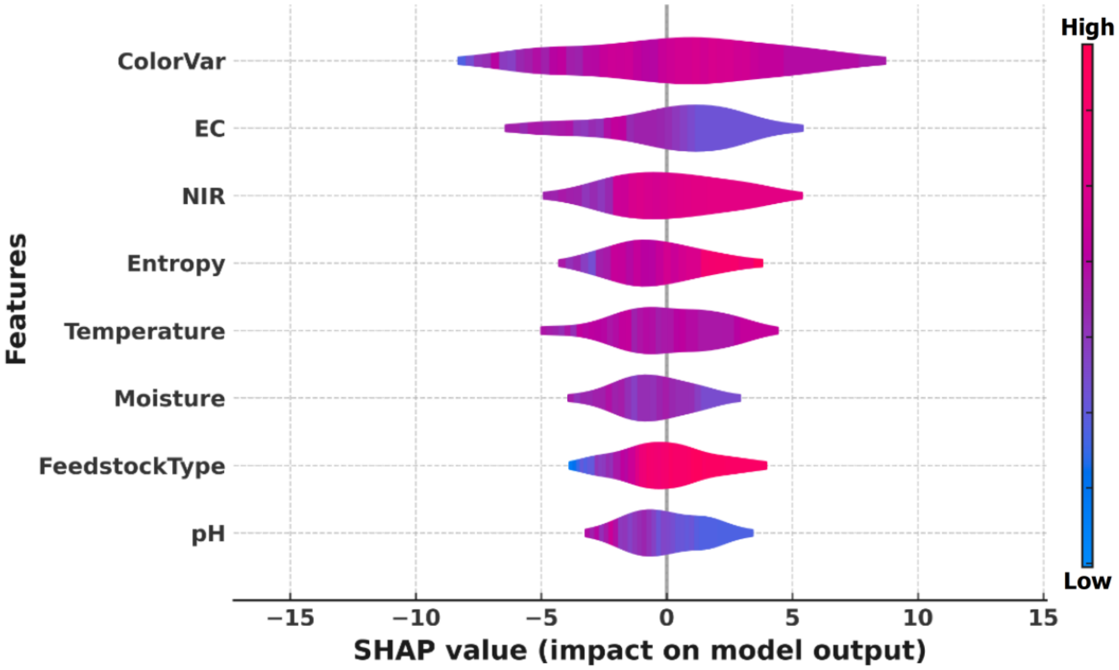

Figure 6. Investigation Influential Feature via SHAP values

Figure 6 shows the SHAP summary plot for VERMIQualNet, demonstrating the most influential inputs that act as predictors for the model, which makes the model more interpretable against the hidden facts. Among the ranked features, the feature with the most influential SHAP values is color variance (ColorVar), which is associated with large variation in SHAP and therefore plays an important role in maturity-index estimation - this is probably due to its sensitivity to the decomposition stage. NIR, EC, and Entropy also significantly impact one of the variables, indicating their importance for nutrient profiles and the time course changes in microbial activity. Moderate contribution of Moisture, temperature, and pH is linked to their basic role in determining the environmental impact of compost stack on the process of decomposition. The narrow SHAP distribution of FeedstockType supposes low variance on that categorical variable. Significantly, high values of NIR and ColorVar consistently increase predictions, whereas low values of moisture and pH decrease scores, and this is as expected from the compost science. Horizontal violin plots indicate steady state model behavior that is free of non-physical dependencies. Overall, the SHAP plot provides evidence for VERMIQualNet's interpretability and identifies biologically relevant predictors that make up the decision-making process of the model.

5.1 Discussion

In order to leverage the great potential in transitioning data from one domain to another, the VERMI-QualNet pipeline combines uncertainty-aware domain adaptation, multi-modal self-supervised representation learning, and rule mining conditioned by the feedstock. Together, these techniques provide a computational construct to reliably interpret and predict the quality assurance of vermicompost at a low-cost. Every step from sensing to examine variably interpretive mining is mathematically linked and even ensures transparency, consistency, and transferability across various composite composting settings.

The results of Figure 2 confirm that VERMIQualNet clearly outperforms traditional and ensemble models when the compost maturity is predicted with high precision. Its powerful generalization over the entire range of maturity allows its use in real-time, cross-feedstock application, thus making it a reliable tool for automating the vermicompost quality assessment on real applications.

In addition, cross-site results of the model validate the ability of VERMIQualNet to adapt to other composting conditions while maintaining high predictive accuracy of the model on unseen sites. Its 36 per cent early batch segregation rate enables proactive quality control and optimization of processes, establishing the model as a scalable and operationally robust solution for real-time vermicompost monitoring and decision support.

5.2 Limitations

Although VERMIQuanNet works fine, it requires standard calibration information for primary training. However, it fails to take into account rare or extreme composting events and therefore has low representation power. In addition, the processing of an ultra-upper low energy edge could be defined to real-time implementation with the multi-mode complexity.

6. CONCLUSION AND FUTURE RESEARCH DIRECTION

In this research paper, we present VERMIQualNet, a domain-adaptive and multi-modal learning model that operates in real-time at the edge devices. It combines low-cost IoT sensor information and other types of data, such as spectral data and visual patterns, with a self-supervised learning capability to predict the MI, CR, and NS. The system comes with high accuracy, interpretability, and generalization, which overperforms the traditional baselines with respect to the metrics like MAE, R^2, RMSE, F1-score, CE, cross-site generalization, and SHAP values. Its SHAP-based hardening of explanations and rule discovery can provide valuable insights into the biological aspects of composting, which are being applied by producers who are looking for consistent quality, safety, and regulated compliance. Much of this research will be continued on real-world anomaly collections for better characterization of outlier composting behaviors, and in further refining the model for near-real-time inference on microcontrollers, for suitability for deployment in low-resource environments.

References

1. Katiyar, Rajesh Babu, Suresh Sundaramurthy, Anil Kumar Sharma, Suresh Arisutha, Anubhav Pratap-Singh, Satyam Mishra, Rashid Ayub, Byong-Hun Jeon, and Moonis Ali Khan. "Vermicompost: An eco-friendly and cost-effective alternative for sustainable agriculture." *Sustainability* 15, no. 20 (2023): 14701.
2. Villar, Iria, David Alves, and Salustiano Mato. "Product quality and microbial dynamics during vermicomposting and maturation of compost from pig manure." *Waste Management* 69 (2017): 498-507.
3. Shi, Shuai, Zhiheng Guo, Jiaxin Bao, Xiangyang Jia, Xiuyu Fang, Huaiyao Tang, Hongxin Zhang, Yu Sun, and Xiuhong Xu. "Machine learning-based prediction of compost maturity and identification of key parameters during manure composting." *Bioresource Technology* 419 (2025): 132024.
4. Rahmani, Amir Masoud, Efat Yousefpoor, Mohammad Sadegh Yousefpoor, Zahid Mehmood, Amir Haider, Mehdi Hosseinzadeh, and Rizwan Ali Naqvi. "Machine learning (ML) in medicine: review, applications, and challenges." *Mathematics* 9, no. 22 (2021): 2970.
5. Raghunath, KM Karthick, Bhuvan Unhelkar, S. Siva Shankar, and P. Chakrabarti. "Statistical assessment of an interpretable AI framework for improved disease diagnosis through medical images." *JOURNAL OF STATISTICS AND MANAGEMENT SYSTEMS* 28, no. 1 (2025): 193-204.
6. Yang, Rongjie, Yingying Chen, Yuling Qiu, Kezhu Lu, Xurui Wang, Gaoyuan Sun, Qiuge Liang, Huixing Song, and Shiliang Liu. "Assessing the landscape ecological health (LEH) of wetlands: research content and evaluation methods (2000–2022)." *Water* 15, no. 13 (2023): 2410.
7. Faverial, Julie, Denis Cornet, Jacky Paul, and Jorge Sierra. "Multivariate analysis of the determinants of the end-product quality of manure-based composts and vermicomposts using Bayesian network modelling." *PloS one* 11, no. 6 (2016): e0157884.
8. Aslam, Zubair, Safdar Bashir, Waseem Hassan, Korkmaz Bellitürk, Niaz Ahmad, Nabeel Khan Niazi, Ahsan Khan, Muhammad Imran Khan, Zhongbing Chen, and Mansoor Maitah. "Unveiling the efficiency of vermicompost derived from different biowastes on wheat (Triticum aestivum L.) plant growth and soil health." *Agronomy* 9, no. 12 (2019): 791.
9. Hosseinzadeh, Ahmad, Mansour Baziar, Hossein Alidadi, John L. Zhou, Ali Altaee, Ali Asghar Najafpoor, and Salman Jafarpour. "Application of artificial neural network and multiple linear regression in modeling nutrient recovery in vermicompost under different conditions." *Bioresource technology* 303 (2020): 122926.
10. Kumar, Pankaj, Pawan Sharma, Tilak Ranjan Kumar, H. C. Sharma, and R. C. Dubey. "Effect of storage conditions on vermicompost quality." *Current Science* 118, no. 2 (2020): 297-300.
11. Riwandi, Zainal Muktamar, Hasanudin, Anandyawati, and Vira Allsari. "The Quality of Vermicompost from Various Sources Composted with Earthworm Perionyx Excavates." *IOP Conference Series: Earth and Environmental Science* 1005, no. 1 (2022): 012006.
12. Wang, Hongyan, Ling Wang, Jiabin Liu, Ying Nie, and Daqing Wang. "Biological Engineering Analysis of Vermicompost Based on Image Features and Machine Learning." *Mobile Information Systems* 2022, no. 1 (2022): 7347142.
13. Hafshejani, Laleh Divband, Abd Ali Naseri, Abdolrahim Hooshmand, Amir Soltani Mohammadi, and Fariborz Abbasi. "Prediction of nitrate leaching from soil amended with biosolids by machine learning algorithms." *Ain Shams Engineering Journal* 15, no. 7 (2024): 102783.
14. Flórez, Sebastián López, Juan M. Corchado, and Fernando De la Prieta. "Edge AIoT-based agricultural recommendation platform to improve humus productivity in vermicomposting processes." *Pervasive and Mobile Computing* (2025): 102080.
15. kmkrphd, "GitHub - kmkrphd/VERMICOMPOST: Quality Prediction," *GitHub*, 2025. https://github.com/kmkrphd/VERMICOMPOST.git (accessed Oct. 07, 2025).