GPR-Former: Context-Aware Moisture Detection

Kevin Lee[1] **and Chen Feng**[1,2]

[1]Building Diagnostic Robotics, Inc., USA
[2]Tandon School of Engineering, New York University, USA

k.lee@nyu.edu, cfeng@nyu.edu

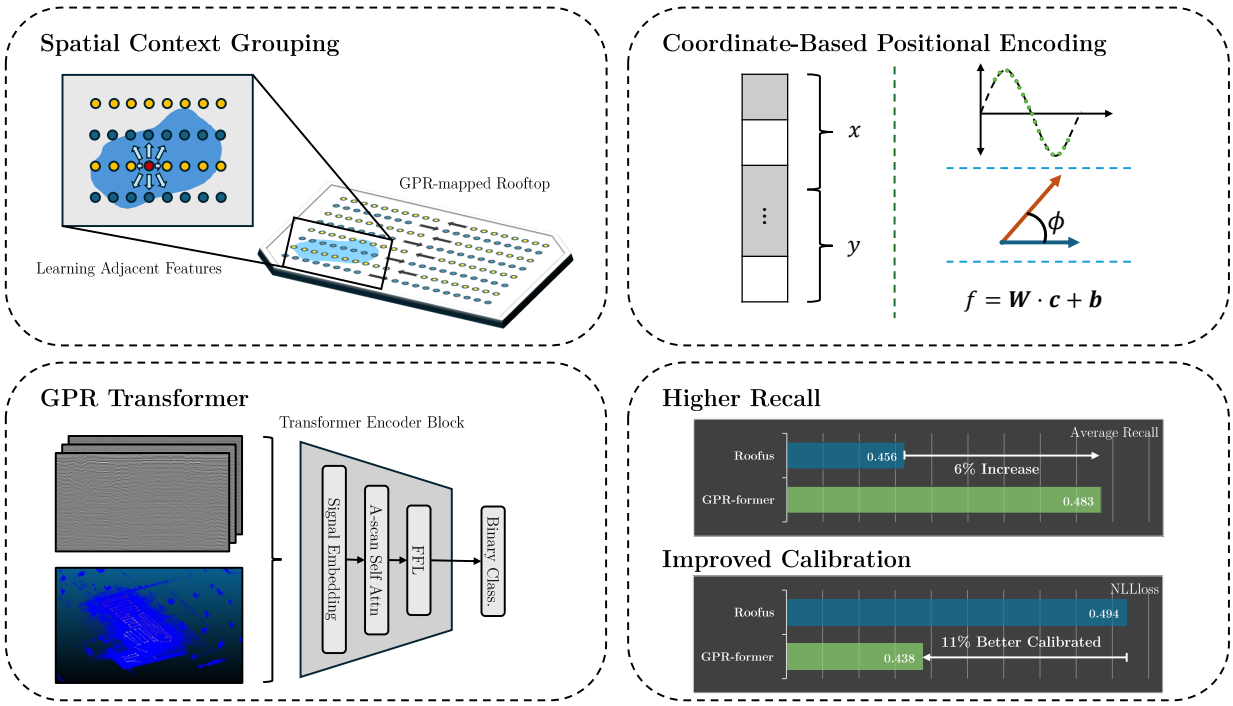


Figure 1. **Overview of GPR-former.** We take inspiration from the complex patterns in which moisture propagates and design a 2D coordinate-based spatial context grouping strategy to more effectively detect the presence of moisture damage on building rooftops. Compared to our previous method, GPR-former demonstrates superior performance in key metrics such as Average Recall as well as having better calibrated outputs.

Abstract -

Moisture damage in roofing systems poses a significant challenge to building operation and maintenance, with far-reaching implications for energy efficiency, structural integrity, health, safety, and sustainability. This study introduces *GPR-former*, a novel transformer-based architecture designed to leverage spatial context in ground-penetrating radar (GPR) data, enhancing the accuracy of moisture detection. Unlike our previous approach that analyzes isolated B-scans, GPR-former incorporates spatial groupings of A-scans with coordinate-based positional encoding, enabling the model to detect complex, nonlinear moisture patterns. Comprehensive experiments across diverse roofing materials demonstrate that GPR-former outperforms our previous method by up to 6% across performance metrics and achieves improved calibrated results. These findings highlight the potential of GPR-former as an improved, transformative tool for sustainable building maintenance, contributing to embodied carbon reduction and extending roof lifecycle management.

Keywords -

Moisture detection, ground-penetrating radar, transformers, spatial context, sustainability

1 Introduction

Moisture damage in roofing systems is a critical issue that impacts energy efficiency, structural safety, and repair costs. It also undermines sustainability efforts by increasing the embodied carbon footprint of buildings through both premature and overdue replacements. Early and ac-

curate detection of moisture damage is essential to mitigate these impacts, support lifecycle management, and enable circular economy initiatives in building renovation and maintenance.

Using GPR to detect moisture in roofing materials presents unique challenges:

- **Signal Obfuscation:** GPR signals are inherently affected by environmental noise, material heterogeneity, and structural interference, making it difficult to identify moisture-related patterns with high confidence.
- **Complex Moisture Patterns:** Moisture often propagates nonlinearly, driven by factors like capillary action and material saturation, complicating interpretation when relying on isolated scans.

Recent advancements in deep learning for GPR [1, 2, 3], particularly the use of transformers and attention mechanisms for B-scan analysis [4, 5, 6], have demonstrated the potential of neural networks in GPR data interpretation. Moreover, progress in the development of mobile robotic systems for GPR data acquisition and processing [4, 7] motivates the development and integration of these technologies. However, methods like our previous approach [4] typically operate within localized linear windows, limiting their ability to leverage spatial relationships critical for identifying moisture patterns across adjacent scans. This gap underscores the need for models that integrate spatial context and enable holistic analysis of GPR data.

To address these challenges, we propose **GPR-former**, a transformer-based architecture that introduces the following key contributions:

- **Spatial Context Grouping:** A novel grouping mechanism that aggregates A-scans based on spatial proximity, providing a richer contextual understanding of moisture propagation.
- **Coordinate-Based Positional Encoding:** A mechanism to encode spatial coordinates directly into the model, allowing GPR-former to effectively map relationships between grouped scans.
- **Comprehensive Evaluation:** Extensive experiments across diverse roofing materials validate the efficacy of the proposed approach.

2 Related Work

2.1 GPR-Based Moisture Detection

Ground-penetrating radar has been widely applied in various domains for moisture detection. For example, [8] employs a drone-mounted GPR sensor to map soil moisture across agricultural fields, demonstrating the versatility of GPR in remote sensing applications. Similarly, [9] integrates GPR with thermal imaging to detect moisture in building walls, highlighting the potential of GPR-based moisture detection specifically in the domain of building diagnostics and maintenance.

In the context of roofing systems, our previous work [4, 10] establishes a foundation for GPR-based moisture analysis. [10] explores self-supervised learning to extract semantically meaningful features from GPR A-scans, enabling downstream tasks like moisture detection with minimal reliance on labeled data. Our investigation indicates the possibility of accurately identifying and isolating GPR scans taken over moisture saturated surfaces, but fails to produce a scalable and robust classifier. Building on this, Roofus [4] applies supervised learning as well as the transformer architecture to classify moisture in real-world rooftop environments using a large-scale labeled dataset of GPR B-scans. Although effective, Roofus focuses on localized analysis within sample B-scans, failing to account for broader spatial relationships that may be critical for detecting moisture propagation across B-scans.

2.2 Transformers in Coordinate-Based Applications

Transformers [11] have gained prominence in tasks requiring spatial reasoning, with positional encoding enabling the modeling of spatial relationships in structured data. In medical imaging, for instance, [12] employs coordinate-based embeddings to enhance transformer-based models for spatially aware tasks. Similarly, CoordFormer [13] introduces a Coordinate-Aware Attention module to encode spatial-temporal coordinates, preserving dependencies critical for analyzing video and image datasets. Li et al. [14] investigates the efficacy of explicit positional encoding by concatenating 2D positional coordinates to the end of each token's embedding dimension among other strategies. Despite these advancements, the application of transformers to specifically GPR data remains underexplored, leaving untapped potential for leveraging spatial relationships in GPR-based analysis.

2.3 Limitations of Existing Approaches

Existing GPR-based methods, including Roofus [4], primarily analyze localized B-scans and overlook the spatial propagation of moisture across adjacent scans. This limitation often results in the model struggling when encountering strongly reflective objects like horizontal rebar. Similarly, while transformers with positional encoding have been effectively applied to coordinate-based tasks, their use in GPR data analysis lacks exploration. Addressing these gaps requires a model capable of capturing the unique spatial relationships inherent in GPR datasets to identify complex moisture diffusion patterns in real-world

rooftop environments.

3 Model Overview

3.1 Task Definition

The primary goal is to classify Ground Penetrating Radar (GPR) A-scans into binary categories indicating the presence or absence of moisture. Formally, given a collection of A-scans grouped by spatial proximity, denoted as $\{S_1, S_2, \ldots, S_\omega\}$, where each scan $S_i \in \mathbb{R}^{655}$ represents a feature vector derived from radar signals, the task is to output a binary prediction $P_i \in \{0, 1\}$ for each scan S_i. Here, $P_i = 1$ signifies the presence of moisture, while $P_i = 0$ indicates its absence.

Binary classification is particularly suited for the application domain, where client-facing diagnostic reports prioritize actionable outcomes. While additional insights such as moisture depth or severity could enrich the analysis, this study focuses on binary classification as a practical first step. The detection of moisture ingress typically necessitates remediation regardless of severity, aligning with the goal of actionable and interpretable results.

Unlike previous approaches that process A-scans or B-scans in isolation, GPR-former leverages spatial context by analyzing grouped A-scans across multiple B-scans. This strategy enhances sensitivity to obfuscated or diffuse moisture signals, which are often missed by single-scan models like Roofus.

3.2 Spatial Context Grouping

Intuition. Moisture in roofing materials seldom manifests as isolated points; instead, it propagates radially or irregularly due to dynamics like capillary action and absorption as illustrated in Figures 1 and 2. These patterns often form clusters or saturate adjacent regions. To capture such propagation behaviors, GPR-former organizes A-scans into spatially coherent groups using x, y coordinates obtained from LiDAR data, enabling the model to recognize subtle or diffuse moisture patterns that elude traditional methods.

Group Sampling Procedure. A straightforward K-D Tree-based algorithm is used to group A-scans according to spatial proximity. We use the K-D Tree for its simplicity and performance. The procedure is outlined as follows:

1. Construct a K-D Tree using the x, y coordinates of all A-scans from a rooftop dataset.
2. Arbitrarily select an A-scan as the seed for a group and identify its $\omega - 1$ nearest neighbors.
3. Form a group of ω A-scans and remove them from the K-D Tree.
4. Repeat steps 2–3 until the K-D Tree is empty. Discard any group with fewer than ω A-scans.
5. Ensure that all groups are formed strictly from A-scans belonging to the same rooftop to prevent erroneous mixtures of scans from different sites.

Algorithm 1: Generate Groups using K-D Tree

Input: GPR Data D, group size ω
Output: Groups G

Initialize D copy: $D_0 \leftarrow D$;
Initialize empty groups: $G \leftarrow \emptyset$;
while *size*$(D_0) \geq \omega$ **do**
  Construct K-D Tree tree $\leftarrow$ `KDTree`(D_0);
  Select the first sample as seed: seed $\leftarrow D_0[0]$;
  Query nearest neighbors:
    neighbors $\leftarrow$ `query`(tree, seed, $k = \omega - 1$);
  Form group: $G \leftarrow G \cup \{\text{seed}, \text{neighbors}\}$;
  Remove group from D_0:
    $D_0 \leftarrow$ `delete`$(D_0, \{\text{seed}, \text{neighbors}\})$;
end
return G

This procedure ensures spatial coherence by grouping A-scans strictly from the same rooftop. Groups with fewer than ω A-scans are discarded to maintain consistency. Although this procedure may not be the most computationally efficient, it is sufficiently effective for handling our large rooftop datasets in practice.

An important hyperparameter to consider is the group size ω. An ω that is too small inhibits the algorithm's capability to span the gaps between B-scans, nullifying the benefits of inter-scan context that we propose. If ω is too large, undesirable artifacts like spatial discontinuity in group membership may occur. Figure 3 provides an illustration describing such phenomenons. In our study, we explore the effects that these situations can have on performance and we investigate the optimal ω in our ablation.

3.3 Coordinate-Based Positional Encoding Strategies

Overview. Positional encodings are essential for embedding spatial relationships among grouped A-scans, enabling the model to leverage spatial context effectively. To achieve this, we explore three different strategies for encoding the positional information into each A-scan representation:

- **Sinusoidal Absolute Encoding:** Employs fixed periodic functions to encode absolute signal feature positions.
- **Rotary Encoding:** Encodes relative positional relationships within groups by applying rotational transformations to the embeddings.
- **Learned Encoding:** Learns position embeddings directly from the data via a linear layer.

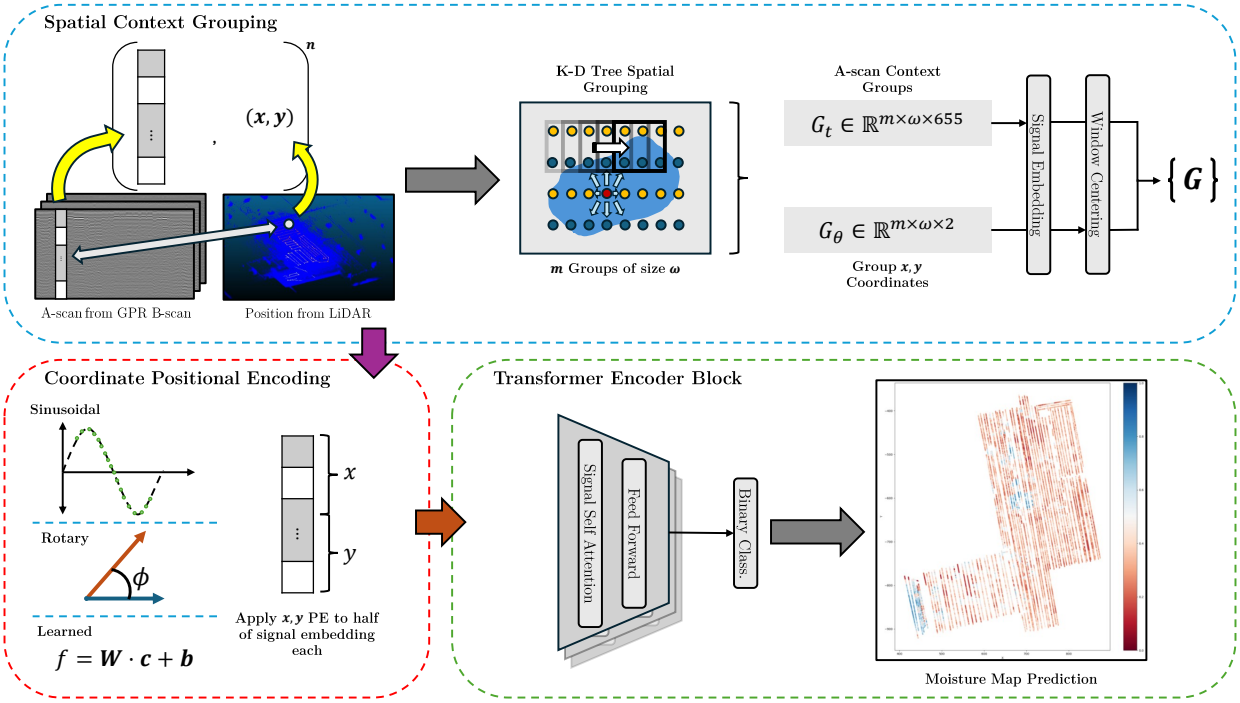


Figure 2. **GPR-former aggregates A-scans into groups of size ω using a K-D Tree sampling procedure.** These grouped scans are processed by a transformer encoder, which incorporates coordinate-based positional encodings to capture spatial relationships and contextual dependencies.

Implementation Details. To account for variability in rooftop layouts and scan alignments, the x, y coordinates are normalized relative to the group's mean position before encoding. This normalization ensures that encodings are invariant to the absolute coordinate system, allowing the model to generalize across diverse rooftop environments.

Each encoding method maps the normalized x, y coordinates to feature embeddings, with half of the embedding dimension allocated to each axis for the Sinusoidal Absolute Encoding and Rotary Encoding. The specific formulations are detailed below:

Sinusoidal Absolute Encoding. We borrow the encoding scheme used in the implementation of the Masked Autoencoder [15], where half of the feature embeddings are dedicated to the x position and the remainder to the y. This particular approach encodes x and y positions using sine and cosine functions of varying frequencies, defined as:

$$\begin{aligned} PE_x[i] &= \sin(x \cdot \beta[i]), \quad PE_x[i+1] = \cos(x \cdot \beta[i]), \\ PE_y[i] &= \sin(y \cdot \beta[i]), \quad PE_y[i+1] = \cos(y \cdot \beta[i]), \end{aligned} \tag{1}$$

where $\beta[i] = \frac{1}{10000^{2i/d}}$, d is half the embedding dimension, and i indexes the frequencies.

Rotary Encoding. We utilize a similar approach as the Sinusoidal encoder with the Rotary encoder, where the encoding transforms feature embeddings to encode relative positions using rotational transformations on x and y embeddings:

$$\begin{aligned} \mathbf{x}_{\text{rot}} &= \mathbf{x} \odot \cos(\theta) + \text{Rot}_{90}(\mathbf{x}) \odot \sin(\theta), \\ \mathbf{y}_{\text{rot}} &= \mathbf{y} \odot \cos(\theta) + \text{Rot}_{90}(\mathbf{y}) \odot \sin(\theta), \end{aligned} \tag{2}$$

where θ is derived from scaled positional values, and $\text{Rot}_{90}(\cdot)$ rotates the second half of the embedding vector by 90 degrees. This method emphasizes spatial relationships by directly modifying feature representations in embedding space.

Learned Encoding. This method learns position embeddings directly through a linear layer:

$$PE = \mathbf{W} \cdot \mathbf{c} + \mathbf{b}, \tag{3}$$

where $\mathbf{c} \in \mathbb{R}^2$ represents the input coordinates, and $\mathbf{W} \in \mathbb{R}^{d\times 2}$, $\mathbf{b} \in \mathbb{R}^d$ are trainable parameters. This approach adapts the encoding to dataset-specific requirements but risks overfitting in low-data scenarios.

Practical Considerations. Unlike static encodings in image-based transformers, the dynamic nature of grouped

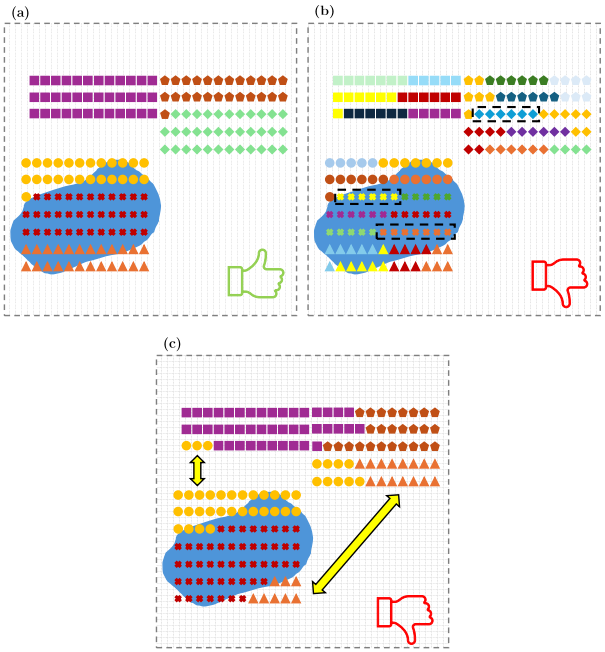


Figure 3. **Comparison between the effects that different groups sizes can have on group semantics and composition.** An ideal group size (a) has sufficiently large groups to just enough cover pertinent regions like puddles. A group size that is too small (b) has difficulty spanning the gap between B-scans, making it no different than previous methods like Roofus [4]. Groups that are too large (c) may introduce unwanted artifacts such as discontinuity and noise.

A-scans in GPR data necessitates real-time encoding computation. However, we find that the real-time computation adds minimal overhead in practice across various large-scale rooftop datasets.

While Sinusoidal Absolute Encoding efficiently captures global structure, Rotary Encoding and Learned Encoding emphasize local spatial patterns and task-specific adaptability, respectively. These differences highlight trade-offs in flexibility, interpretability, and computational requirements for each encoding scheme.

3.4 Transformer Architecture

In GPR-former, we adopt a vanilla transformer architecture similar to the design introduced by [11]. This transformer is central to our approach as it models the contextual and spatial dependencies among grouped A-scan features. Each transformer layer block is comprised of multi-head self-attention, a feed-forward network, residual connections, and layer normalization. After the transformer encoder, the resulting feature vector is passed through a final linear layer followed by a sigmoid activation to yield a binary classification probability.

Figure 2 illustrates the summarized approach of GPR-former.

4 Experiments

Implementation. GPR-former employs a transformer encoder architecture with a binary classification head as shown in Figure 2. Training is conducted using Focal Loss to address the inherent dataset imbalance (it is typical to expect moisture signals to be less frequent than dry signals), with a learning rate of 1×10^{-5} on a cosine annealing schedule. Weighted binary cross-entropy was also tested but yielded suboptimal results compared to Focal Loss, likely due to the latter's ability to emphasize difficult samples more effectively. We use the AdamW optimizer and train on an Nvidia RTX 8000 GPU with a batch size of 64.

Three versions of GPR-former are developed: **GPR-former-B** (base), **GPR-former-L** (large), and **GPR-former-H** (huge). These versions differ in the number of attention heads, layers, and representation dimensions, with GPR-former-B containing a total of 44.6M parameters, GPR-former-L containing 84.7M parameters, and lastly GPR-former-H containing 118.9M parameters. For reference, the architecture used for Roofus contains 302M parameters, making GPR-former noticeably light-weight in comparison. The exact details of each configuration is described in Figure 1. The architectures share a common classification layer, ensuring consistency across the comparison. Hyperparameters were optimized through grid search, with specific focus on dropout, learning rate, and batch size. All GPR-former models in our main evaluation use Sinusoidal positional encoding and a group size $\omega = 1024$. We provide the justification in our ablation analysis in Section 4.2.

Dataset. The dataset is sourced from the same building complex used in the Roofus investigation [4], a 46.5K m^2 retail complex with diverse roofing materials, including modified bituminous membranes, EPDM rubber, and TPO. It contains hundreds of thousands of A-scans spanning a wide range of roofing conditions and damage levels. Labels were manually annotated and cross-validated against reference moisture maps provided by a third-party partner. These reference maps were generated using established rooftop moisture detection techniques, including infrared thermography, electrical capacitance testing, and nuclear moisture gauges.

The dataset is divided into distinct sections, labeled Section 1 through Section 10, with Section 7 excluded due to scanning errors. Section 6 is reserved exclusively for testing, while the remaining sections are used for training and validation with an 85/15 split. This division ensures

Models	Hyperparameters			# Parameters
	# Heads	# Layers	Representation Dim.	
Roofus	16	24	1024	302M
GPR-former-B	4	8	768	44.6M
GPR-former-L	8	10	1024	84.7M
GPR-former-H	16	10	1280	118.9M

Table 1. **Design differences between the different sizes of GPR-former, with the tested version of Roofus included for comparison.** The different sized models vary in the number of attention heads and layers as well as their representation dimensions.

Models	Quantitative			Qualitative
	AR ↑	F2-Score ↑	AUROC ↑	NLLloss ↓
Roofus	0.4563	0.7158	0.8240	0.4940
GPR-former-B	0.4563	0.7267	0.8386	0.4465
GPR-former-L	0.4730	0.7174	0.8296	0.4443
GPR-former-H	**0.4834**	**0.7277**	**0.8436**	**0.4376**

Table 2. **Comparison of GPR-former variations and Roofus across evaluation metrics on the test set.** GPR-former-H appears to be better fit to take advantage of the additional spatial information through the increased number of encoder layers and attention heads.

that the model is evaluated on unseen data, reflecting real-world deployment scenarios.

Prior to input into the models, the GPR B-scans are pre-processed using a linear gain down the time domain. The data is then standardized by the mean and standard deviation values calculated from the training set. Additionally, we apply a random horizontal or vertical flip of the x, y coordinates with a separate probability of 0.5 for each flip during training.

Baseline. To benchmark the performance of GPR-former, we compare it with our previously developed Roofus model. Roofus also employs a transformer-based architecture but operates on GPR B-scans, treating A-scans within a window as sequential data without incorporating spatial context beyond individual B-scans. Roofus was retrained on the current dataset using the same training regimen as GPR-former, with adjustments where necessary to accommodate its architectural differences.

Evaluation Metrics. Given the critical need to identify all potentially damaged roofing areas, **recall** is selected as the key evaluation metric. While precision is valuable, prioritizing recall ensures minimal false negatives, which is essential for identifying areas that might require repair. In roofing diagnostics, conservative predictions often lead to better outcomes, as even marginally damaged areas are typically repaired along with severely damaged sections.

To capture this priority, we use metrics such as Average Recall (AR) and F2-score, with F2-score weighted more heavily toward recall. The Area Under the Receiver Operating Characteristic (AUROC) is also included to provide both consistency with our previous investigation with Roofus as well as further insights into the trade-off between true positive and false positive rates. F2-scores are calculated using the optimal threshold identified during the final validation step.

4.1 Main Results

Quantitative. Table 2 presents the performance of GPR-former variants compared to Roofus. The results show that the largest model, GPR-former-H, using $\omega = 1024$ and the Sinusoidal positional encoding, achieves the highest recall, F2-score, and AUROC, demonstrating the benefits of incorporating spatial context into the model. The slight underperformance of GPR-former-L compared to GPR-former-B in F2-Score and AUROC is an unexpected result, potentially due to overfitting or architectural inefficiencies. Further analysis of this phenomenon is left for future work.

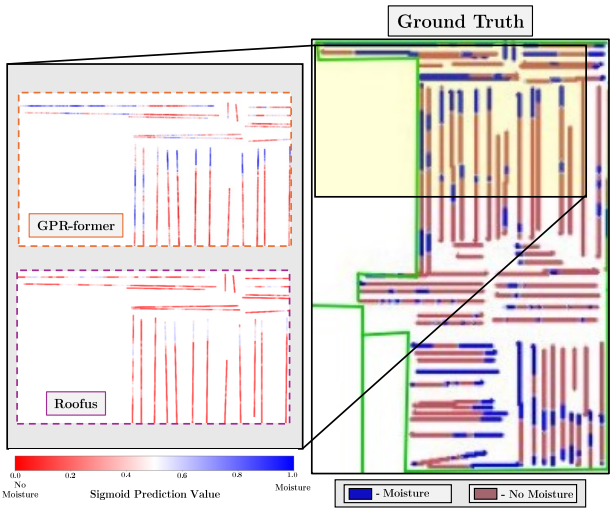
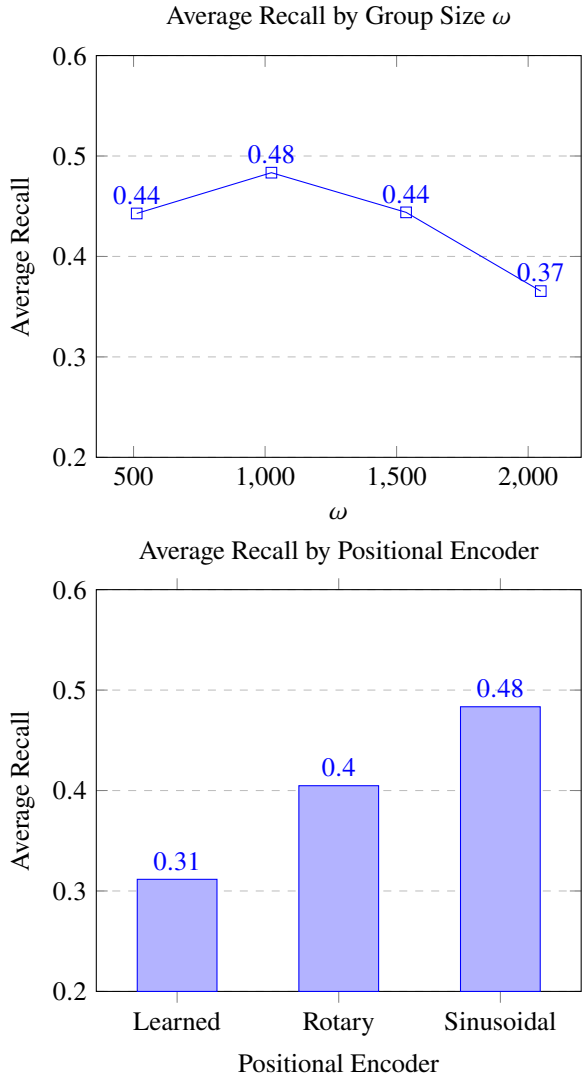


Figure 4. **Qualitative comparisons against the ground truth labels of the un-thresholded moisture predictions generated by GPR-former and Roofus [4].** The example highlighted region clearly illustrates the improvement in the calibration of GPR-former is over its predecessor.

Qualitative. Despite being hidden from client-facing reports, the un-thresholded predictions can provide additional insights into how well-calibrated the models are and the uncertainty present in them. Figure 4 illustrates the raw model outputs prior to thresholding for GPR-former and Roofus. GPR-former exhibits a wider distribution of prediction scores, indicating higher confidence in its classifications. To quantify uncertainty, we compute the negative log-likelihood loss as shown in Table 2, reveal-

ing that GPR-former produces more calibrated predictions compared to Roofus.



4.2 Ablation

Group size ω. As discussed prior, the group size ω determines the spatial context provided to the model. Larger groups introduce additional spatial information but risk incorporating irrelevant or noisy data, while smaller groups may lack sufficient context. To strike a balance, we tested group sizes of $\omega = 512, 1024, 1536, 2048$. Results indicate optimal performance at $\omega = 1024$, beyond which performance declines, likely due to grouping artifacts or interruptions in data continuity. Overlapping groups or advanced sampling methods may address these limitations in future iterations.

Positional Encoding. As previously mentioned, the three encoding schemes were evaluated: sinusoidal absolute positional encoding, a rotary positional encoder, and a learned positional encoder. Contrary to expectations, the sinusoidal encoder outperformed the rotary encoder, possibly due to the relative encoder's complexity relative to the dataset size. The poor performance of the learned encoder underscores the importance of carefully designed encoding schemes in spatial transformers, especially in low-data scenarios.

5 Discussions

Limitations. This study shares limitations with prior work [4], including the need for an expanded dataset with additional roofing material types and environmental conditions to generalize the model's applicability. Artifacts in the sampling procedure, such as large gaps between group members, can lead to discontinuities in spatial context. Furthermore, the current approach is unable to quantify the severity of damage or localize the damage beneath the surface, which somewhat limits the practical utility of the predictions.

Conclusions. The experiments validate the effectiveness of incorporating spatial context in transformer-based architectures for GPR data. GPR-former outperforms our previous model Roofus across key metrics, demonstrating the advantages of spatially aware processing.

Future Work. Directions for future research include developing advanced methods for constructing groups (e.g., overlapping groups or clustering-based sampling) and increasing the dataset size through scanning and labeling more rooftops to enhance diversity and saturate the capabilities of relative positional encoding. Investigating the applicability of self-supervised learning may also address the difficulties in large-scale data collection and annotation. Exploring the prediction of moisture depth or damage severity could further improve the practical value of the system in real-world diagnostics.

Acknowledgements. This work is supported by NSF grant #2322242.

Disclosure: Chen Feng is a founder of Building Diagnostic Robotics, Inc., a startup company that uses AI and robotics for building inspections.

References

[1] Sher B. and Feng C. Deepgpr: Learning to identify moisture defects in building envelope assemblies from ground penetrating radar. In Borja García de Soto, Vicente Gonzalez-Moret, and Ioannis Brilakis, editors, *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 561–568, Chennai, India, July 2023. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-0-4. doi:10.22260/ISARC2023/0075.

[2] Luo T. X., Zhou Y., Zheng Q., Hou F., and Lin C. Lightweight deep learning model for identifying tunnel lining defects based on gpr data. *Automation in Construction*, 165:105506, 2024. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2024.105506. URL https://www.sciencedirect.com/science/article/pii/S0926580524002425.

[3] Hoang N. Q., Shim S., Kang S., and Lee J. Anomaly detection via improvement of gpr image quality using ensemble restoration networks. *Automation in Construction*, 165:105552, 2024. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2024.105552. URL https://www.sciencedirect.com/science/article/pii/S0926580524002887.

[4] Lee K., Lin W., Javed T., Madhusudhan S., Sher B., and Feng C. Roofus: Learning-based robotic moisture mapping on flat rooftops with ground penetrating radar. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024.

[5] Xu T., Yuan D., Yang G., Li B., and Fan D. Fm-gan: Forward modeling network for generating approaching-reality b-scans of gpr pipelines with transformer tuning. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. doi:10.1109/TGRS.2024.3359351.

[6] Chen Y., Li F., and Zhou S. New asphalt pavement dielectric constant inversion net based on improved gated attention for ground penetrating radar. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–11, 2024. doi:10.1109/TGRS.2024.3496881.

[7] Eldemiry A., Muddassir M., and Zayed T. Autonomous data acquisition of ground penetrating radar (gpr) using lidar-based mobile robot. In Vicente Gonzalez-Moret, Jiansong Zhang, Borja García de Soto, and Ioannis Brilakis, editors, *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 206–212, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0028.

[8] Wu K., Rodriguez G. A., Zajc M., Jacquemin E., Clément M., De Coster A., and Lambot S. A new drone-borne gpr for soil moisture mapping. *Remote Sensing of Environment*, 235:111456, 2019. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2019.111456. URL https://www.sciencedirect.com/science/article/pii/S0034425719304754.

[9] Garrido I., Solla M., Lagüela S., and Fernández N. Irt and gpr techniques for moisture detection and characterisation in buildings. *Sensors*, 20(22), 2020. ISSN 1424-8220. doi:10.3390/s20226421.

[10] Lee K., Lin W., Sher B., Javed T., Madhusudhan S., and Feng C. Exploring self-supervised gpr representation learning for building rooftop diagnostics. In Vicente Gonzalez-Moret, Jiansong Zhang, Borja García de Soto, and Ioannis Brilakis, editors, *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 928–935, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0120.

[11] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., and Polosukhin I. Attention is all you need. In Guyon I., Von Luxburg U., Bengio S., Wallach H., Fergus R., Vishwanathan S., and Garnett R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[12] Das B. K., Zhao G., Islam S., Re T. J., Comaniciu D., Gibson E., and Maier A. Co-ordinate-based positional embedding that captures resolution to enhance transformer's performance in medical image analysis. *Sci. Rep.*, 14(1):9380, April 2024.

[13] Li H., Dong H., Jia H., Huang D., Kampffmeyer M. C., Lin L., and Liang X. Coordinate transformer: Achieving single-stage multi-person mesh recovery from videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8744–8753, October 2023.

[14] Li Y., Ma Z., Wang X., Wang Y., and Tan B. Vision transformer with 2d explicit position encoding. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7690–7694, 2024. doi:10.1109/ICASSP48485.2024.10446293.

[15] He K., Chen X., Xie S., Li Y., Dollár P., and Girshick R. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15979–15988, 2022. doi:10.1109/CVPR52688.2022.01553.