A Hybrid Dynamic Graph Neural Network–XGBoost Framework for Hyperspectral Urban Scene Classification

Hussain Mobarak Albarakati, Ameer Hamza, *Member IEEE*, Bayan Alabdullah, Areej Alasiry, Mehrez Marzougui, Zepa Yang, *Member IEEE*

Abstract— High spectral dimensionality, non-local class connections, and unbalanced class distributions can all restrict the efficacy of fixed-neighborhood classifiers, making hyperspectral image classification difficult. In order to build discriminative representations while enhancing robustness to class imbalance, this study suggests a hybrid Dynamic Graph Neural Network–XGBoost architecture for hyperspectral urban scene classification. The suggested method aggregates relevant non-local spectral relationships beyond strict local kernels by first encoding each labelled pixel using a dynamic graph neural network that builds adaptive connections in feature space. After that, the graph-aware embedding are sent to an XGBoost classifier, which improves class-wise separability and refines the decision boundaries. To optimise the boosting step and the graph encoder, a thorough grid search was carried out. While retaining nearly comparable validation performance (Overala Accuracy = 93.17%, macro-F1 = 0.9196), the best-performing configuration obtained test overall accuracy of 93.01%, average accuracy of 91.15%, Cohen's kappa of 0.9072, and macro-F1 of 0.9167. While the majority of residual errors were focused among spectrally identical urban materials, class-wise examination also showed notably strong discriminating for spectrally distinctive and minority categories. These findings imply that the suggested hybrid approach successfully blends gradient-boosted classification with adaptive graph-based representation learning, providing a viable substitute for traditional hyperspectral classifiers for intricate urban land-cover research.

***Index Terms*— Hyperspectral image classification, dynamic graph neural network, XGBoost, imbalanced data classification, urban scene classification, graph-based representation learning, non-local feature relations, spectral-spatial analysis.**

I. INTRODUCTION

Hyperspectral photography is one of the best ways to remotely sense land cover in detail since it catches hundreds of contiguous spectral bands and gives very precise information about the composition of materials [1, 2]. Hyperspectral data keep small spectral changes that make it possible to tell the difference between surface classes that look the same but have different spectral properties. Because of these capabilities, hyperspectral image categorization has become a key part of remote sensing [3, 4]. It may be used for things like mapping cities, monitoring the environment, and farming, finding minerals, and assessing disasters [5]. However, the same complexity that makes hyperspectral images interesting also makes it hard to classify them. The high number of dimensions in spectral signatures, the presence of redundant and correlated bands, the small number of labelled training samples, and the frequent overlap among class distributions all make learning difficult, especially in complex urban scenes where materials often show subtle differences between classes [6, 7].

The Pavia University scene is still one of the most researched benchmark hyperspectral datasets for testing classification systems in cities [8]. Its mixed composition, which includes roads, plants, building materials, bare soil, and shaded areas, makes it a realistic place where both spectral similarity and class imbalance can have a big effect on how well a model works [9-11]. Some classes in these datasets are well represented and have unique spectral features, whereas others are not very common or have a lot of overlap with classes that are close to them. Because of this, classifiers that are merely optimized for overall accuracy may still have trouble getting balanced class-wise discrimination [12]. This problem is especially critical in urban hyperspectral analysis, because misclassification happens not between classes that are linguistically distant, but between elements that are

Funding: This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00218176) and the Soonchunhyang University Research Fund. Authors like to thanks Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R440), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for Large Research Project under Grant RGP2/471/46.

Corresponding*: Zepa Yang (yangzepa@sch.ac.kr) and Ameer Hamza (ameer.hamza@ktu.edu)

Hussain Mobarak Albarakati is with the Department of Computer and Network Engineering, College of Computing, Umm Al-Qura University, Makkah, Saudi Arabia (hbarakati@uqu.edu.sa).

Ameer Hamza is with Centre of Real Time Computer Systems, Kaunas University of Technology, Lithuania (ameer.hamza@ktu.edu).

Bayan is with Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia (Balabdulah@pnu.edu.sa)

Areej Alasiry, Mehrez Marzougui are with College of Computer Science, King Khalid University, Abha 61413, Saudi Arabia (areej.alasiry@kku.edu.sa; mhrez@kku.edu.sa)

Zepa Yang is with Department of Computer Science and Engineering, Soonchunhyang University, South Korea. (yangzepa@sch.ac.kr).

structurally connected, such gravel, bricks, asphalt, bitumen, and bare soil [13]. A strong classification framework must not only match dominant categories, but it must also learn representations that are good enough to separate classes that are hard to separate and not balanced [14].

A substantial corpus of research has focused on hyperspectral image classification through traditional machine learning techniques, including support vector machines, random forests, k-nearest neighbours, and boosting-based methods, alongside deep learning architectures such as convolutional neural networks, recurrent models, autoencoders, and transformer-inspired frameworks. These techniques have significantly enhanced the state of the art; nonetheless, significant limits persist [15, 16]. Conventional machine learning methods frequently succeed when manually produced features are robust; nevertheless, they may not fully leverage the latent structure inherent in high-dimensional hyperspectral signals [17]. Deep learning models, on the other hand, can automatically learn hierarchical features. However, many designs depend on fixed local operators with receptive fields that are set in advance [18]. In image-based classification, this kind of locality is typically helpful. But in hyperspectral images, useful associations don't always happen between pixels that are right next to each other or in rigid neighbourhoods [19, 20]. Pixels that are far apart in space may nonetheless have very similar spectral properties, while pixels that are close together may belong to separate material groups. This tension implies that fixed-neighborhood learning may not consistently serve as the most effective method for modelling hyperspectral class structure [21, 22].

Graph-based learning is an interesting option since it lets data samples be shown as nodes that are connected by adaptive relationships instead of only by pre-defined Euclidean neighborhoods [18]. In this context, graph neural networks can combine information from connections that are made on the fly, allowing them to simulate non-local relationships that may be hard to capture with regular convolutional filters. This perspective is especially appealing for hyperspectral imaging, as class structure is frequently influenced by similarities in spectral feature space rather than solely by local pixel adjacency [23]. A graph formulation can thus facilitate a more adaptable relational perspective of the data, allowing the model to enhance connections between spectrally suitable samples, even when these ties extend beyond static local windows. Still, graph-based encoders alone do not always solve all categorization problems. In imbalanced datasets, minority or unclear classes may necessitate enhanced boundary refinement during the decision phase, particularly when the learned embedding space features partially overlapping clusters [24, 25].

Based on these findings, this study presents a hybrid Dynamic Graph Neural Network–XGBoost architecture for the categorisation of hyperspectral metropolitan scenes[26]. The main idea is to combine the strong representation of dynamic graph-based embedding learning with the strong discrimination of gradient-boosted decision trees [27]. The proposed approach begins by representing each labelled pixel with its hyperspectral signature. Then, a dynamic graph neural network processes the pixel and builds adaptive relationships in feature space [28]. The graph encoder does not only use stiff local kernels; instead, it learns graph-aware embeddings by combining information from spectrally relevant neighbours that are chosen dynamically during training. The learnt embeddings are then given to an XGBoost classifier, which is a second-stage decision module that makes class boundaries clearer and makes the model more stable when there is class imbalance. This two-stage design is meant to take advantage of the capabilities of both parts. The graph encoder captures relational structure and non-local feature dependencies, while XGBoost is a strong, well-regularized classifier that can handle class distributions that are not the same [29].

Beyond land-cover classification, hyperspectral and remote sensing imagery also supports several closely related analytical tasks, including spectral endmember unmixing, data compression, and object-level interpretation [30]. For example, scattering-term constrained nonnegative matrix factorization has been explored for hyperspectral unmixing, while prediction- and vector-quantization-based methods have been used for hyperspectral image compression [31]. In parallel, improved YOLOX-based object detection methods have been developed for remote sensing imagery, demonstrating the broader relevance of deep representation learning across remote sensing analysis tasks. These studies show that hyperspectral classification is part of a wider research ecosystem involving spectral decomposition, efficient data representation, and semantic scene understanding [32].

There are three reasons why the suggested technique is ideal for hyperspectral urban classification. First, it loosens the idea that the best contextual information has to come from pixels that are next to each other in the image plane. Second, it gives a way to discover adaptive feature-space connections that can better show how hyperspectral signatures are organized on their own. Third, it adds a boosted classification stage that can improve the decision surface after learning graph-based representations. This is especially useful when there are classes that are imbalanced or spectrally confusing. In this way, the framework is not just a sequential combination of two models; it is a planned combination of relational embedding learning and boosted classification that aims to improve both representation quality and class separability [33, 34].

A lot of tests were done on the Pavia University hyperspectral dataset to see how well the proposed method worked. A full hyperparameter search was done on both the graph encoder and the XGBoost classifier. The results show that the hybrid framework works well and consistently, with the best setup getting a test overall accuracy of 93.01%, an average accuracy of 91.15%, Cohen's kappa coefficient of 0.9072, and a macro-F1 score of 0.9167, while keeping validation performance very close. The analysis by class shows that the method works very well for categories that are spectrally distinct. It also shows that most of the residual errors are in materials that are more spectrally similar, which means that the model learns a meaningful and understandable

class structure instead of just relying on the dominance of majority classes [35].

Thus, work's key contributions can be summed up like thus. To start, a new hybrid framework is presented that merges dynamic graph-based embedding learning with XGBoost-based decision refinement for classifying hyperspectral images. Second, the work shows how useful adaptive non-local feature-space connections may be in urban hyperspectral scenes instead of fixed neighborhood modelling. Third, the suggested method works well even when there is class imbalance because it uses graph representation learning, boosted classification, and systematic hyperparameter optimization. Finally, the study goes beyond just giving an overall accuracy score and gives a full class-wise and error-oriented evaluation. This shows how the system works with both common and hard urban material classes.

The key contributions of this work are summarized as follows:

- A novel hybrid hyperspectral classification framework is proposed by integrating a Dynamic Graph Neural Network (Dynamic GNN) with XGBoost, thereby combining adaptive graph-based representation learning with boosted decision refinement for urban hyperspectral scene classification.
- An adaptive feature-space graph learning strategy is introduced to capture non-local spectral relationships among labeled pixels, reducing the reliance on rigid fixed-neighborhood processing and enabling more flexible contextual modeling in complex urban environments.
- A two-stage learning pipeline is developed in which the Dynamic GNN learns graph-aware latent embeddings and XGBoost refines class boundaries in the learned feature space, improving discrimination among imbalanced and spectrally similar urban land-cover classes.
- A comprehensive experimental evaluation is conducted on the Pavia University hyperspectral dataset using systematic hyperparameter optimization, class-balanced metrics, per-class analysis, confusion-matrix interpretation, precision–recall analysis, and convergence diagnostics to provide both predictive and analytical validation of the proposed framework.
- The proposed model achieves strong and stable performance on the benchmark dataset, demonstrating that hybrid graph-based embedding learning and gradient-boosted classification constitute a competitive and robust solution for imbalanced hyperspectral urban scene classification. A detailed comparative analysis has been conducted with recent studies and pre-trained models based on accuracy and parameters. In addition, the proposed network is also passed to an explainable AI technique to analyze its explainability.

While dynamic graph-based methods are effective at modeling non-local spectral relations, their predictive stage is commonly tied to standard neural classification heads, which may be suboptimal when the learned embedding space still contains partially overlapping class clusters. This is particularly problematic in imbalanced urban hyperspectral scenes, where minority classes and spectrally proximate materials often require finer boundary shaping than a single end-to-end neural head can provide. The proposed Dynamic GNN–XGBoost framework addresses this gap through a non-trivial hybridization strategy: the Dynamic GNN is used for adaptive graph-aware representation learning, whereas XGBoost is introduced as a downstream classifier to refine nonlinear decision boundaries in the learned latent space. In this sense, the fusion is not an additive stacking of components, but a targeted integration designed to address residual ambiguity after graph embedding.

The remainder of this paper is organized as follows. Section II reviews related work on hyperspectral image classification, graph-based learning, and boosting-based classifiers. Section III describes the proposed Dynamic Graph Neural Network–XGBoost framework in detail, including data preprocessing, graph embedding learning, and classifier optimization. Section IV presents the experimental setup, evaluation metrics, and hyperparameter search strategy plus combining with discusses the quantitative and qualitative results, including class-wise behavior and error patterns. Finally, Section V concludes the paper and outlines directions for future research.

II. Related Works-State of Art

Hyperspectral image classification has evolved through several distinct methodological phases, beginning with pixel-wise statistical learning, progressing through spectral-spatial feature engineering, and more recently moving toward deep and graph-based representation learning. Early work established strong classical baselines through kernel methods, particularly support vector machines and related regularized classifiers, which proved effective for high-dimensional hyperspectral data but largely treated each pixel as an isolated sample in feature space [36]. In parallel, morphological-profile-based methods became especially influential for urban scenes, because they injected structural information into classification pipelines by combining principal components with multiscale spatial operators [37]. Subsequent spectral-spatial fusion strategies, such as SVMs combined with morphological profiles, improved urban classification by jointly exploiting spectral signatures and spatial patterns, especially in complex built environments like Pavia [38]. However, despite their strong performance, these approaches were still fundamentally dependent on handcrafted feature design, manually selected preprocessing stages, and fixed assumptions about local structure.

The next major shift came from deep learning, which reduced reliance on manual feature engineering and enabled end-to-end representation learning from hyperspectral data. CNN-based methods became particularly prominent because they could learn hierarchical spectral and spatial abstractions directly from image cubes. Chen et al. showed that 3-D convolutions could jointly encode spectral and spatial

information and thereby outperform purely spectral pipelines [39]. Ben Hamida et al. further reinforced the value of 3-D deep models for remote sensing image classification, demonstrating that joint spectral-spatial processing could produce strong classification performance with comparatively efficient architectures [40]. Albarakati et al. [41] suggested a network-level fusion architecture based on self-attention for agriculture land cover and land use classification from remote sensing images. The model achieved accuracies of 98.20%, 89.50%, and 91.70% on SIRI-WHU, EuroSAT, and NWPU datasets, respectively. This study demonstrated that network-level fusion could reduce the limitations of traditional feature-level fusion by improving discriminative representation while controlling computational cost. However, the method still depends on CNN-based feature extraction and optimization-driven feature selection, which may limit its ability to model irregular non-local relationships in complex hyperspectral scenes. To address the limitations of single-scale feature extraction, Bhatti et al. [42] introduced MSRA-Net, a multiscale residual attention network for land use and land cover scene classification from remote sensing images. MSRA-Net was tested on MLRSNet and NWPU-RESISC45 datasets. The suggested model gained classification accuracies of 93.71% and 97.95%, respectively. The study verified that multiscale residual attention and cross-scale fusion can improve scene classification performance, especially when RS images contain low resolution and high intra-class similarity.

Rubab et al. [43] further investigated the problem of aerial scene recognition under low-contrast remote sensing conditions by proposing BNResNet, a batch-normalization-inspired deep bottleneck residual architecture. The method was designed to classify aerial scenes and coastal regions from low-resolution and low-contrast remote sensing images. The architecture was evaluated on EuroSAT, NWPU-RESISC45, SIRI-WHU, and a customized CoastalD dataset, achieving accuracies of 94.62%, 92.88%, 99.75%, and 90.0%, respectively. Beyond single-modality classification, Sajid et al. [44] presented SASR, a sensor-agnostic semantic representation unification framework for cross-modal RGB and hyperspectral aerial scene classification. Instead of relying on explicit multimodal fusion, the suggested method formulates cross-modal learning as semantic representation unification. The model gained 96.4% RGB accuracy and 97.3% hyperspectral accuracy on Houston 2013, with cross-modality transfer accuracies of 87.2% from RGB to HSI and 88.7% from HSI to RGB. Under joint training, performance enhanced to 97.0% and 97.8%, while additional evaluations on UC Merced and Indian Pines obtained 98.7% and 97.6% overall accuracy, respectively. However, its focus is modality alignment rather than dynamic relational modeling over learned graph embedding. In change detection, Sheng et al. [45] presented a cross-hierarchical difference feature fusion network based on a multiscale encoder-decoder architecture for coastal wetland monitoring. Experiments on four public datasets demonstrated that CHDFFN achieved average maximum improvements of 4.61% in overall accuracy, 19.79% in Kappa, and 18.90% in F1-score compared with state-of-the-art methods. However, the method is designed for bi-temporal change detection rather than general hyperspectral scene classification.

More recently, transformer-based models have been introduced to address a limitation of convolutional networks: their dependence on local receptive fields and fixed neighborhood aggregation. For example, the Spatial-Spectral Transformer explicitly combines CNN-based spatial extraction with transformer-based modeling of long-range spectral relations [46]. Surveys of the field now consistently describe the state of the art as a transition from traditional machine learning to CNN-dominated pipelines and increasingly toward transformer-driven architectures capable of broader contextual reasoning [3].

Even so, a central limitation remains in much of the deep-learning literature: many methods still rely on regular image grids and fixed local neighborhoods, which may be suboptimal for hyperspectral scenes whose informative relationships are not always confined to immediately adjacent pixels. This limitation motivated the rise of graph-based hyperspectral learning. Hong et al. argued that while CNNs are effective at capturing spatial-spectral representations, their ability to model relations among samples is limited, whereas graph convolutional networks can move beyond rigid grid sampling and operate on irregular relational structures [47]. Building on this idea, Wan et al. proposed the Multi-scale Dynamic Graph Convolutional Network (MDGCN), in which graph structure is dynamically updated during convolution so that graph refinement and feature discrimination improve each other [48]. In a subsequent context-aware dynamic graph model, the same research line emphasized long-range contextual relations and adaptive graph refinement as a way to reduce classification errors in irregular and boundary regions [49]. More recent review work now identifies graph-based hyperspectral classification as one of the most active research directions in the field, highlighting its promise for modeling non-Euclidean relations, multiscale interactions, and non-local dependencies.

Another parallel line of research concerns the classifier itself rather than only the feature extractor. Ensemble learners, especially boosting methods, remain highly competitive because they can produce sharp decision boundaries, strong generalization on structured tabular features, and robustness under heterogeneous class distributions. XGBoost, introduced by Chen and Guestrin, became a widely adopted gradient-boosting framework because of its scalability, regularization mechanisms, and high empirical performance across many machine-learning tasks [33, 34]. In hyperspectral remote sensing, however, boosting has more often been applied after handcrafted or precomputed spectral-spatial features rather than after deep relational embeddings. A representative example is Meta-XGBoost, where XGBoost was investigated and extended for hyperspectral image classification using advanced morphological-profile features, showing that gradient-boosted trees can remain competitive when supplied with informative descriptors [13]. This line of work is important because it suggests that tree-based classifiers still have a meaningful role in hyperspectral analysis, especially when class imbalance, nonlinear class boundaries, and feature heterogeneity are present. At the same time, it also reveals a gap: the integration of dynamic graph-learned embeddings with boosted decision refinement remains far less explored

than CNN-softmax, transformer-MLP, or handcrafted-feature-plus-SVM pipelines.

From a state-of-the-art perspective, the literature can therefore be read as a progression across four generations. The first generation emphasized spectral classifiers and kernel methods; the second introduced spectral-spatial feature engineering and morphological descriptors; the third shifted to end-to-end deep networks, especially CNNs and transformers; and the fourth increasingly focuses on graph-based models capable of representing adaptive and non-local relations [49]. Yet an unresolved methodological opportunity remains at the intersection of these trends: many graph-based hyperspectral methods focus on representation learning but still terminate in conventional neural decision heads, while many strong boosting-based methods operate on handcrafted features rather than on learned relational embedding [48].

Recent studies have further expanded graph-based hyperspectral classification by designing more adaptive neighborhood and aggregation mechanisms. Adaptive path aggregation graph frameworks and multi-scale receptive-field graph attention models have shown that flexible graph construction can improve the modeling of long-range spectral-spatial dependencies in HSI classification. Similarly, AF2GNN introduced adaptive filters and aggregator fusion to improve land-cover discrimination and spatial feature learning within graph convolutional architectures. Beyond classification, enhanced deep image prior methods for unsupervised hyperspectral image super-resolution [4]also reflect a broader trend toward exploiting deep structural priors and scene-adaptive representations in hyperspectral image analysis. These developments support the need for models that move beyond fixed local operators and motivate the present hybrid framework, where dynamic graph-based embedding learning is coupled with boosted decision refinement [20, 29].

This suggests that a carefully designed hybrid framework may offer a useful synthesis: graph learning can provide adaptive, non-local feature representations, while boosting can refine decision boundaries and improve class discrimination in imbalanced urban scenes. That gap motivates the present study. Rather than positioning graph learning and tree boosting as competing alternatives, this work treats them as complementary stages within a unified classification pipeline for hyperspectral urban analysis.

III. Material and Methods

A. Study Workflow Overview

The proposed framework's overall workflow (Fig 1) is meant to combine adaptive representation learning with strong downstream classification for analyzing hyperspectral urban scenes. The data are first sorted into pixel-level spectral samples from the input hyperspectral cube. Then, unlabeled pixels are deleted, and the labelled instances that are left are ready for supervised learning. Then, a preprocessing step is done to make the spectral features more consistent and create the training, validation, and test subsets that are utilized in all of the trials. This step makes sure that the next learning stages work with normalized inputs and that the model's behavior can be evaluated reliably across different parameter values.

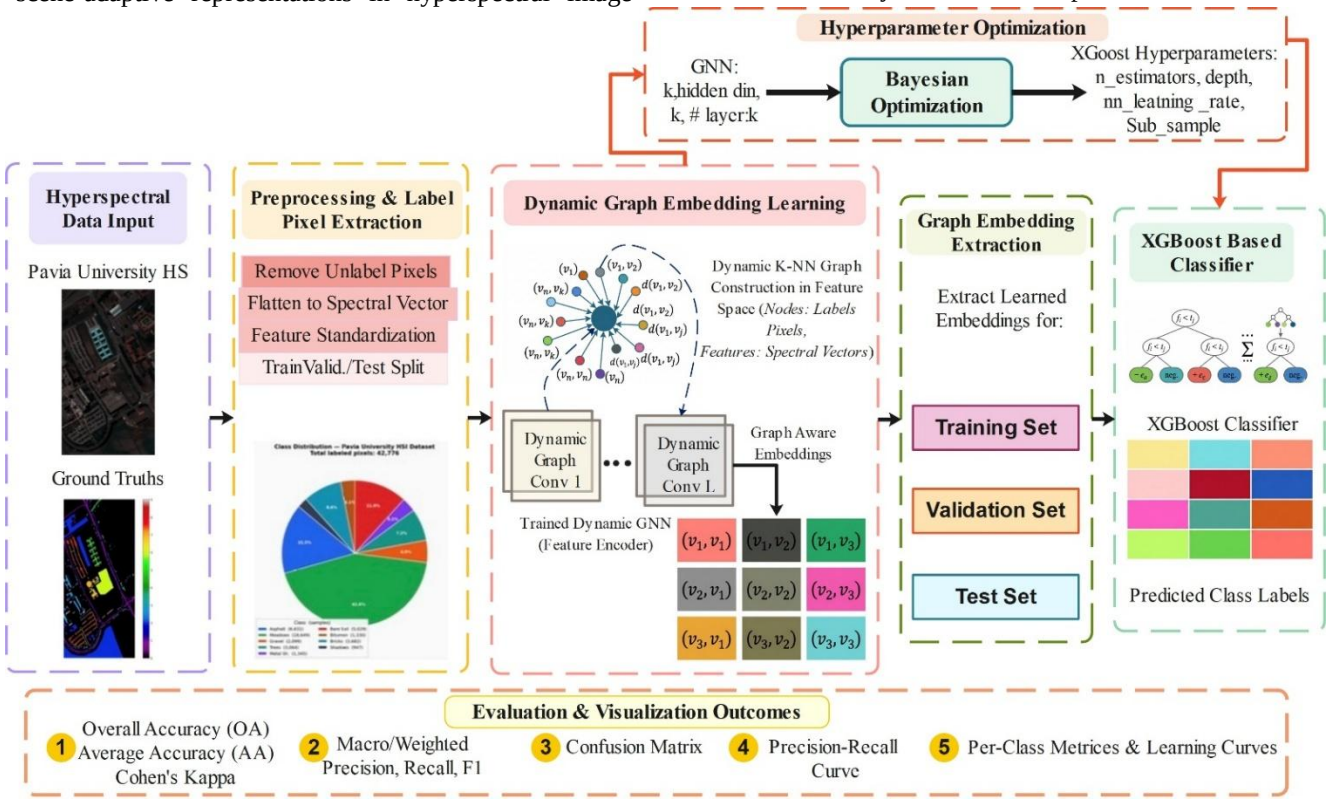

Fig 1. The proposed framework processes hyperspectral data through preprocessing and dynamic graph embedding learning, and then applies XGBoost for robust classification. Systematic hyperparameter optimization and comprehensive evaluation ensure reliable and interpretable results.

After preprocessing, the spectrum samples are sent to a

dynamic graph-learning module, which is the feature encoder for the proposed technique. At this point, each labelled pixel is treated as a node with its hyperspectral signature, and dynamic neighbourhood creation creates adaptive relationships in feature space. This approach lets the encoder gather information from spectrally relevant samples in a more flexible way than fixed-window processing does. This results in graph-aware embedding's that capture non-local structure in the data that can be used to tell the difference between different types of data. The learned embedding are not the final classification output; instead, they are an intermediate representation that is meant to make the feature space richer and easier to separate for the next judgement stage.

The graph-derived embedding are then given to an XGBoost classifier to make class discrimination better. This second-stage classifier is used to model complex decision boundaries across the learned embedding space and to make the model more robust when there is class imbalance and inter-class spectral similarity. In this way, the proposed framework uses a hybrid approach. The dynamic graph neural network learns how to represent data, and XGBoost improves those learnt features by making decisions based on them. This split of tasks lets the approach take advantage of both adaptive graph-based feature extraction and the strong classification power of gradient-boosted trees at the same time.

To find the best model configuration, a systematic hyperparameter optimization process is done on both parts of the pipeline. The graph encoder is adjusted based on important structural factors such the size of the neighborhood, the number of hidden dimensions, the number of graph layers, and the dropout rate. The XGBoost stage is improved by changing its boosting-related factors. The resulting combinations are tested in a grid-search framework, which lets you compare all the candidate settings and see clearly how different architecture and classifier decisions affect performance.

Finally, a multi-level evaluation process that incorporates global, class-wise, and diagnostic analysis is used to test the suggested workflow. To measure how well the model predicts, standard hyperspectral classification metrics are used. To help understand how the model works, complementary visual outputs are also created. These outputs include of class-distribution analysis, validation-performance trends throughout training, per-class precision-recall-F1 comparisons, confusion matrices, and precision-recall curves. The workflow serves as both a prediction pipeline and an analytical framework for comprehending the interplay between dynamic graph representation learning and boosted classification in the realm of imbalanced hyperspectral urban scene classification.

B. Dataset Description

The proposed framework underwent assessment using the Pavia University hyperspectral picture collection [8], a prominent standard for urban hyperspectral scene classification. The Reflective Optics System Imaging Spectrometer (ROSIS) took the picture over the University of Pavia area in northern Italy. It is widely used in the remote sensing literature since it has a lot of spectral information and a mix of urban features. Because it has a combination of roads, plants, building materials, and shaded areas, it is a great place to test classification frameworks in real-world and spectrally difficult situations.

After removing noisy channels, the hyperspectral cube includes 103 spectral bands and is 610×340 pixels in size. So, each pixel has a 103-dimensional spectral signature that gives detailed information in the visible and near-infrared range. There are ground-truth annotations for a part of the picture, which means that 42,776 pixels are labelled. The other 164,624 pixels are not labelled and are not used for supervised training. The labelled samples are divided into nine groups based on the kind of land cover: Asphalt, Meadows, Gravel, Trees, Painted Metal Sheets, Bare Soil, Bitumen, Self-Blocking Bricks, and Shadows.

The Pavia University dataset has a big class imbalance, which makes it a very useful baseline for this investigation. The class frequencies are very uneven. Meadows has the most samples (18,649), followed by Asphalt (6,631), Bare Soil (5,029), and Self-Blocking Bricks (3,682). On the other hand, minority classes like Painted Metal Sheets (1,345), Bitumen (1,330), and Shadows (947) have much fewer samples. This imbalance makes the learning environment inherently biased. Conventional classifiers may do well overall, but they may not do as well on classes that are less common but nonetheless relevant in practice. Consequently, the dataset offers a robust foundation for evaluating the efficacy of the proposed hybrid Dynamic GNN–XGBoost framework in sustaining both elevated global performance and equitable class-wise discrimination. The class-distribution figure in Fig 2 is there to help you see how unbalanced the sample is and to explain why learning that takes this into account is important.

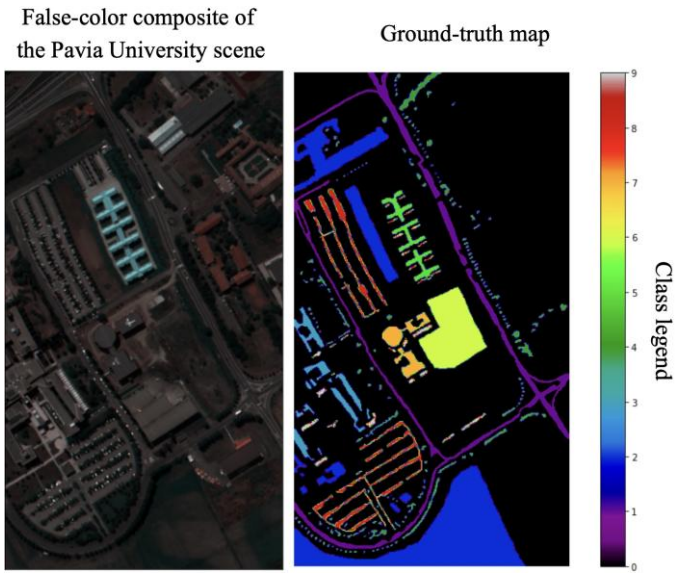

Fig 2. Pavia University hyperspectral dataset used in this study: (left) false-color composite of the ROSIS image and (right) corresponding ground-truth map with nine urban land-cover classes. Black denotes unlabeled pixels.

Pavia University is very useful for applications because it shows how complicated real urban environments can be. For example, materials with different semantic meanings may

show partially overlapping spectral responses, and different urban objects may be arranged in strange ways in space. These traits make the dataset especially hard for algorithms that just use fixed local operators or spectral differentiation based on pixels. In contrast, the suggested framework is built to work in this kind of environment by using dynamic graph embedding to learn adaptive feature-space relations and then improving class separation with boosted decision boundaries. The Pavia University dataset is a good test for advanced hyperspectral classification techniques because it is both difficult and representative. Its high spectrum dimensionality, diverse urban elements, and substantial class imbalance make it well-suited to the goals of this study and suitable for confirming the efficacy of the suggested hybrid model.

C. Data Preparation and Preprocessing

To establish a consistent and reproducible input representation for the proposed hybrid framework, the original hyperspectral scene was transformed from a three-dimensional image cube into a structured set of supervised learning samples. Let the hyperspectral cube be denoted by $X \in \mathbb{R}^{H\times W\times B}$, where H and W represent the spatial dimensions, and B denotes the number of spectral bands. For the Pavia University dataset, this corresponds to a cube of size $610 \times 340 \times 103$. In order to enable pixel-wise learning, the cube was reshaped into a two-dimensional matrix $X_{flat} \in \mathbb{R}^{N\times B}$,where $N = H \times W$ and each row corresponds to the spectral signature of a single pixel. In parallel, the ground-truth annotation map was vectorized into a one-dimensional label array $y_{flat} \in \mathbb{R}^{N}$, thereby preserving the one-to-one correspondence between spectral samples and class labels.

Since the ground-truth image contains both labeled and unlabeled pixels, a masking procedure was applied to isolate only the annotated samples. Specifically, pixels with ground-truth value zero were considered unlabeled and excluded from the supervised learning pipeline. Only pixels satisfying $y > 0$ were retained, yielding the labeled dataset used for training and evaluation. This step is essential in hyperspectral benchmark studies because the majority of scene pixels are typically unlabeled and cannot be meaningfully incorporated into supervised classification. After masking, the retained label values, originally defined in the range 1 to 9, were re-indexed to the range 0 to 8. This label normalization simplifies implementation in modern machine learning libraries and ensures compatibility with zero-based class indexing required by the graph encoder and downstream classifier.

Following labeled-pixel extraction, the dataset was partitioned into three mutually exclusive subsets corresponding to training, validation, and test data. A stratified random splitting strategy was adopted in order to preserve the original class proportions in each subset and reduce sampling bias across minority and majority classes. First, 20% of the labeled data were reserved for testing. The remaining 80% were then further divided such that the validation subset represented 10% of the total labeled data, resulting in an overall $70\%/10\%/20\%$ split for training, validation, and test sets, respectively. Under this protocol, the final partition sizes were 29,942 training samples, 4,278 validation samples, and 8,556 test samples. The use of a dedicated validation set allowed model configuration and hyperparameter tuning to be performed independently of the final test evaluation, thereby supporting a more disciplined experimental workflow.

Because hyperspectral bands often exhibit different dynamic ranges and statistical dispersions, feature normalization was applied prior to model training. In particular, a StandardScaler transformation was used to standardize each spectral band independently according to Eq 1:

$$x' = \frac{x-\mu}{\sigma} \tag{1}$$

where x is the original band value, μ is the mean of that band computed from the training set, and σ is the corresponding standard deviation. Importantly, the scaler was fitted only on the training subset and then applied to the validation and test subsets using the training-derived statistics. This procedure prevents information leakage from validation or test data into the preprocessing stage and ensures that the reported classification performance reflects genuine generalization rather than distributional contamination.

The resulting standardized spectral vectors constitute the input to the proposed Dynamic Graph Neural Network. By converting the hyperspectral cube into a normalized pixel-level representation while preserving class balance through stratified sampling, this preprocessing stage provides a stable foundation for both graph-based embedding learning and subsequent boosted classification. In this study, preprocessing is therefore not merely a technical prerequisite, but a key component of the experimental design that ensures the learned representations are comparable, reproducible, and suitable for fair evaluation across different model configurations.

D. Dynamic Graph-Based Feature Embedding and Representation Extraction

After the Dynamic GNN has been trained, the final classifier head is no longer treated as the main decision component of the framework. Instead, the trained graph encoder is used as a feature extractor to generate graph-aware latent representations for the training, validation, and test subsets has been shown in Eq 2.

$$z_i = g(x_i;\Theta) \tag{2}$$

denote the embedding produced by the trained graph encoder g (.) with learned parameters Θ. Then, for each sample in the dataset, the output of the final graph-embedding layer is collected as its new representation in the latent feature space.

To make this process computationally manageable for large sample sets, embeddings are extracted in batches. For each subset, the standardized spectral vectors are passed through the encoder in inference mode, and the resulting hidden features are concatenated to form three embedding matrices in Eq 3 [50]:

$$Z_{tr} \in \mathbb{R}^{N_{tr}\times d}, Z_{val} \in \mathbb{R}^{N_{val}\times d}, Z_{te} \in \mathbb{R}^{N_{te}\times d} \quad (3)$$

Where N_{tr}, N_{val}, N_{te} denote the number of samples in the training, validation, and test sets, respectively, and d is the embedding dimension. These matrices constitute the interface between the graph-learning stage and the boosting stage. This embedding extraction step is central to the hybrid nature of the proposed method. Rather than using raw spectral vectors directly in the final classifier, the model first transforms them into a relationally enriched feature space shaped by dynamic graph aggregation. As a result, XGBoost does not operate on the original hyperspectral input, but on a more structured representation in which class-relevant relationships have already been emphasized by the graph encoder [48].

Once graph-aware embeddings are obtained, the second stage of the proposed framework applies Extreme Gradient Boosting (XGBoost) for final classification. The rationale for introducing XGBoost after graph-based representation learning is twofold. First, boosted tree ensembles are highly effective at modeling nonlinear decision boundaries in tabular feature spaces. Second, they often exhibit strong practical robustness in settings where class distributions are uneven and where different classes form partially overlapping clusters. In the present framework, the role of XGBoost is therefore to refine the class separation already initiated by the Dynamic GNN.

Given the training embeddings Z_{tr} and their corresponding labels y_{tr} XGBoost learns an additive ensemble of decision trees (Eq 4):

$$\hat{y}_i = \sum_{m=1}^{M} f_m(z_i),\ f_m \in \mathcal{F} \quad (4)$$

Where M is the number of boosting rounds, f_m denotes the m-th decision tree, and $\mathcal{F}$ is the space of regression trees. In multiclass classification mode, the model optimizes a differentiable objective that combines predictive loss with structural regularization over the tree ensemble. This enables the classifier to progressively focus on difficult samples and improve discrimination among confusable classes.In this study, XGBoost is configured using four principal hyperparameters: the number of estimators, the maximum tree depth, the learning rate, and the subsample ratio. These parameters control, respectively, the boosting length, the complexity of each tree, the contribution of each boosting step, and the fraction of samples used per iteration. The validation embeddings Z_{val} are used during tuning to assess how well different configurations generalize beyond the training data, while the final prediction results are obtained from the test embeddings Z_{te}.

The decision to use XGBoost as the downstream classifier is methodologically intentional rather than auxiliary. If the Dynamic GNN is viewed as a mechanism for adaptive relational feature construction, then XGBoost functions as a mechanism for high-resolution boundary refinement in the learned embedding space. This division of labor is particularly useful for urban hyperspectral classification, where some classes are readily separable while others remain spectrally adjacent even after embedding. Under such conditions, the boosted classifier can exploit subtle nonlinear partitions in the latent space that may not be fully captured by a single linear classifier head.

Taken together, the proposed Dynamic GNN–XGBoost framework forms a coherent hybrid architecture in which graph-based embedding learning and boosted classification contribute complementary strengths. The first stage transforms raw hyperspectral signatures into graph-aware latent features through adaptive feature-space relations, and the second stage converts those features into final class predictions through gradient-boosted decision refinement. This combination constitutes the methodological core of the present study.

E. Proposed Hybrid Dynamic Graph Neural Network–XGBoost Framework

The proposed method in Fig 3 is designed as a two-stage hybrid classification framework in which graph-based representation learning and boosted decision refinement are combined within a unified pipeline. The central premise is that hyperspectral pixels should not be treated only as isolated spectral vectors, nor should their relationships be restricted to fixed local neighborhoods. Instead, the method seeks to learn an adaptive embedding space in which pixels with informative spectral affinity can exchange contextual information through dynamically constructed relations. Once such graph-aware embeddings are learned, a gradient-boosted classifier is employed to further refine class separation, particularly in the presence of imbalance and material-level spectral ambiguity.

For clarity, the term “Dynamic GNN encoder” is used throughout this study to denote the graph-based representation-learning module that transforms standardized hyperspectral pixel vectors into graph-aware embeddings. The term “XGBoost classifier” refers to the downstream boosted decision module applied to these embeddings, while “Dynamic GNN–XGBoost framework” denotes the complete two-stage hybrid architecture.

In the proposed implementation, the graph adjacency is not predefined or kept fixed throughout training. Instead, the k-nearest-neighbor graph is reconstructed dynamically during each forward computation of the Dynamic GNN using the current node-feature representation. Consequently, the edge set can change across training epochs as the encoder parameters are updated, and it can also change across graph layers because each DynamicEdgeConv layer operates on a transformed feature space. During validation and testing, the model is used in inference mode without parameter updates; however, the graph is still recomputed from the validation or test feature representations rather than inherited from the training graph. Therefore, the validation and test edges are feature-dependent and dynamically generated, but they do not involve ground-truth labels or any transfer of training-set adjacency information. This design preserves the adaptive nature of the graph while avoiding a static neighborhood assumption.

More specifically, dynamic graph construction is performed by applying k-nearest-neighbor search to the current node-

feature representation during each forward computation of the Dynamic GNN encoder. At the first graph layer, the neighbors are determined from the standardized spectral vectors, whereas subsequent DynamicEdgeConv layers recompute the neighborhood from the transformed latent features produced by the preceding layer. Therefore, the edge set may change across layers and training epochs as the embedding space evolves, but the recomputation criterion remains deterministic and label-independent. This design allows the model to adapt its relational structure as class-discriminative features emerge, while training stability is supported by a fixed neighborhood size, feature standardization, dropout, Adam optimization, cosine annealing, and early stopping. Dynamic k-NN reconstruction introduces additional computational overhead compared with fixed-adjacency graph convolution; however, the implementation performs graph construction during batch-wise forward computation rather than storing a full global adjacency matrix over all labeled pixels. This makes the approach tractable for the Pavia University dataset, while approximate k-NN search or cached graph updates may be useful for larger-scale hyperspectral scenes.

Unlike conventional end-to-end classifiers that rely on a single prediction head, the proposed architecture deliberately separates representation learning from decision refinement. The Dynamic Graph Neural Network (Dynamic GNN) acts as a feature encoder that transforms raw spectral signatures into discriminative relational embeddings, while XGBoost operates on these learned embeddings to model complex nonlinear decision boundaries. This design makes the framework both structurally interpretable and practically flexible: the graph module is responsible for extracting adaptive non-local feature relationships, whereas the boosting module is responsible for maximizing downstream class discrimination.

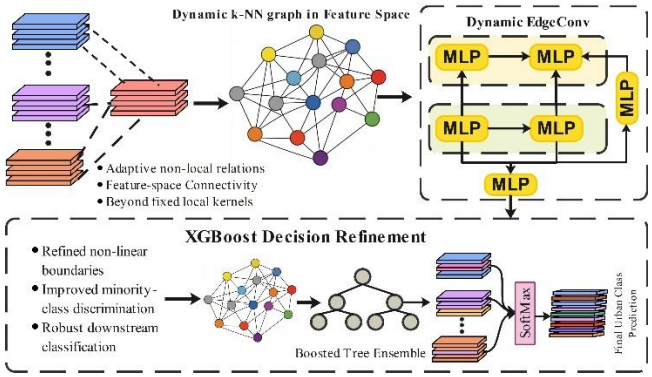

Fig 3. Architecture of the proposed hybrid Dynamic GNN–XGBoost framework. Pixel-wise spectral signatures are first encoded through a dynamically constructed feature-space graph using stacked DynamicEdgeConv layers to generate graph-aware embeddings. These embeddings are then supplied to an XGBoost classifier for boosted decision refinement and final land-cover prediction.

Hyperspectral image classification is intrinsically challenging because each pixel is described by a high-dimensional spectral vector that often contains correlated, redundant, and partially overlapping information across bands. Although this richness enables fine-grained material discrimination, it also makes the learning task more sensitive to class overlap, limited annotations, and variations in class frequency. In urban scenes, these issues are particularly pronounced, as classes such as asphalt, bitumen, gravel, and bricks may exhibit similar spectral tendencies while still belonging to distinct semantic categories. Under such conditions, purely pixel-wise classifiers may struggle to construct robust boundaries, especially for underrepresented classes.

The methodological novelty of the proposed framework lies not in the isolated use of a Dynamic GNN or XGBoost, but in the specific division of learning responsibilities between the two stages. The Dynamic GNN encoder performs adaptive relational representation learning by reconstructing feature-space neighborhoods and aggregating non-local spectral relations among labeled pixels. However, the learned embedding space may still contain partially overlapping class regions, particularly for imbalanced and spectrally similar urban materials. XGBoost is therefore introduced as a downstream boundary-refinement module rather than as a simple replacement classifier. Its boosted tree structure is well suited to modeling irregular nonlinear partitions in the graph-aware embedding space, while its regularized additive learning process provides robustness against heterogeneous and imbalanced class distributions. This two-stage formulation preserves the interpretability and modularity of the pipeline while allowing representation learning and final decision refinement to be optimized according to their respective strengths.

A second challenge arises from the way contextual information is modeled. Many spectral-spatial classifiers rely on fixed local neighborhoods, implicitly assuming that the most informative relationships are confined to adjacent pixels in the image plane. While this assumption is often useful, it can also be restrictive. In hyperspectral scenes, relevant relationships are not always strictly local in a geometric sense; pixels that are spatially separated may still share highly informative spectral structure, while immediate neighbors may belong to different classes. This motivates a more flexible relational view in which contextual aggregation is guided by feature similarity rather than by rigid spatial windows alone.

Graph-based learning offers a natural mechanism for addressing this limitation. By representing pixels as nodes and constructing connections adaptively in feature space, a graph neural network can learn from non-local spectral relations that may be missed by conventional fixed-kernel operators. However, learning an embedding space alone does not guarantee optimal class separation, particularly when class imbalance and fine-grained inter-class similarity remain present in the transformed feature space. For this reason, the proposed framework introduces a second-stage classifier based on XGBoost, which is well suited for refining complex decision surfaces and handling heterogeneous class distributions. In the present study, the graph encoder and the boosting classifier are therefore not treated as independent components, but as complementary stages: the Dynamic GNN learns relationally enriched embeddings, and XGBoost

performs boosted discrimination over those embeddings.

F. Loss Function and Training Strategy

A thoroughly thought-out training plan was put in place to make sure that the Dynamic GNN learns how to tell the difference between classes while still being able to handle the big class imbalance in the Pavia University dataset. The proposed framework uses the graph encoder as the first and most important step in the hybrid pipeline. The goal of training was not just to lower the classification loss at the encoder head, but also to create a latent feature space where classes that are rare or have unclear spectra could be more easily separated during downstream decision refinement.

To address the imbalance among land-cover categories during graph encoder training, a weighted cross-entropy loss was employed. Let C denote the number of classes, $y_i \in \{0, 1, \dots, C-1\}$ the ground-truth label of sample i, and $\hat{p}_i$ the predicted class-probability vector produced by the GNN classifier head. The training objective is defined as Eq 5 [51]:

$$\mathcal{L}_{WCE} = -\frac{1}{N} \sum_{i=1}^{N} w_{y_i} \log \hat{p}_{i,y_i} \tag{5}$$

Where w_{y_i} is the class-specific weight assigned to the true class of sample i, and $\hat{p}_{i,y_i}$ is the predicted probability for that class. Compared with standard cross-entropy, this weighted formulation increases the contribution of underrepresented categories and reduces the dominance of majority classes during optimization.

The class weights were computed directly from the class frequencies of the training subset. If n_c denotes the number of training samples belonging to class c, the initial class weight was defined as the inverse of the corresponding class count, shown in Eq 6 [52]:

$$\tilde{w}_c = \frac{1}{n_c + \varepsilon} \tag{6}$$

Where ε is a small constant introduced for numerical stability. To avoid excessively large or inconsistent scaling across classes, the weights were then normalized as Eq 7 [53]:

$$w_c = \frac{\tilde{w}_c}{\sum_{j=1}^{C} \tilde{w}_j} C \tag{7}$$

This normalization preserves the relative importance of minority classes while keeping the overall loss magnitude stable during training. As a result, rare categories such as Shadows, Bitumen, and Painted Metal Sheets receive proportionally greater emphasis than dominant classes such as Meadows, helping the encoder avoid biased representation learning.

The Dynamic GNN was optimized using the Adam optimizer, which provides adaptive learning-rate adjustment for each parameter and has proven effective for training deep neural models on high-dimensional data. Let Θ denote the set of trainable parameters of the graph encoder and classifier head. The parameter update at iteration t follows the standard Adam formulation (Eq 8):

$$\Theta_{t+1} = \Theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \delta} \tag{8}$$

Where Θ is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first- and second-order moment estimates, and δ is a small constant for numerical stability. In the present implementation, the initial learning rate was set to 1×10^{-3} and the optimizer included weight decay of 1×10^{-4} to provide additional regularization against overfitting.

In addition to weight decay, dropout was applied after each DynamicEdgeConv layer, as described in Subsection D, to improve generalization and reduce co-adaptation among hidden units. This regularization mechanism was especially important because the graph encoder was trained on a relatively limited number of labeled samples compared with the dimensional richness of the hyperspectral input. To improve convergence stability and allow the optimization process to transition smoothly from coarse exploration to fine adjustment, a cosine annealing learning-rate scheduler was adopted. If T denotes the total number of training epochs and t the current epoch, the learning rate at epoch t is updated according to present in Eq 9:

$$\eta_t = \eta_{min} + \frac{1}{2} (\eta_{max} - \eta_{min})\left(1 + \cos\left(\frac{\pi t}{T}\right)\right) \tag{9}$$

Where η_{max} is the initial learning rate and η_{min} is the minimum learning rate reached at the end of training. This scheduling strategy allows the model to learn aggressively during early epochs while gradually reducing the step size as training progresses, thereby promoting a more stable refinement of the graph-based feature space.

To prevent unnecessary training once validation performance ceased improving, an early stopping criterion was incorporated. Specifically, the validation loss was monitored after each epoch, and training was terminated when no improvement was observed for 15 consecutive epochs. The model parameters corresponding to the lowest validation loss were retained as the final encoder weights. This strategy improves computational efficiency and reduces the risk of overfitting, particularly in hyperspectral classification tasks where validation behavior may become unstable after prolonged training.

The maximum number of training epochs for the Dynamic GNN was fixed at 100 epochs. In practice, this upper bound was sufficiently large to allow convergence while still remaining computationally manageable within the full grid-search framework. All experiments were conducted in a CUDA-enabled Kaggle environment with GPU acceleration to support efficient graph-based training and large-scale hyperparameter evaluation. The Dynamic GNN training stage was executed on the GPU, while the downstream XGBoost classifier was trained using parallel CPU execution where appropriate. This combination provided a practical and scalable setting for the proposed hybrid framework, particularly given the repeated training required by the grid-

search procedure.

Overall, the adopted loss design and training strategy were selected to align closely with the challenges of hyperspectral urban classification. Weighted cross-entropy explicitly counteracted class imbalance during representation learning, while Adam optimization, cosine annealing, dropout regularization, and early stopping together provided a stable and effective training regime for the Dynamic GNN encoder. These choices were instrumental in producing graph-aware embeddings that were both discriminative and suitable for subsequent refinement by XGBoost.

G. Hyperparameter Search and Model Selection

To ensure a fair and systematic assessment of the proposed hybrid framework, a comprehensive grid-search strategy was adopted to optimize both the Dynamic GNN encoder and the downstream XGBoost classifier. Because the proposed method is composed of two functionally distinct yet interdependent components graph-based representation learning and boosted decision refinement hyperparameter tuning was performed jointly across both stages rather than treating either component in isolation. This design allows the final configuration to reflect the interaction between encoder quality and classifier behavior, which is especially important in imbalanced hyperspectral classification where seemingly minor architectural changes can alter the separability of difficult classes.

For the Dynamic GNN, the grid search focused on four structural hyperparameters that control the expressive capacity and relational behavior of the graph encoder:

- the dynamic neighborhood size k.
- the hidden embedding dimension: $hidden\_dim$.
- the number of graph layers: num_layers.
- the dropout rate: $dropout$.

The neighborhood size k determines how many nearest neighbors are dynamically selected in feature space for each node during graph construction, and therefore directly affects the extent of relational aggregation. The hidden dimension controls the representational richness of the learned embeddings, while the number of graph layers governs the depth of iterative graph-based feature refinement. The dropout parameter serves as a regularization mechanism to reduce overfitting during encoder training. In this study, the Dynamic GNN search space was defined as follows in Eq 10 [54]:

$$\begin{gathered} \mathrm{k} \in \{10, 20\}, \\ hidden\_dim \in \{64, 128\} \\ num\_layers \in \{2, 3\} \\ \mathrm{dropout} \in \{0.3, 0.5\} \end{gathered} \tag{10}$$

This yields a total of $2 \times 2 \times 2 \times 2 = 16$ candidate GNN configurations. The second stage of the hybrid framework was optimized by searching over four principal XGBoost hyperparameters:

- the number of estimators: $n\_estimators$.
- the maximum tree depth: max_depth.
- the learning rate: $learning\_rate$.
- The subsampling ratio:$subsample$.

These parameters jointly govern the complexity, learning dynamics, and generalization behavior of the boosted tree ensemble. The number of estimators controls the length of the boosting process, the tree depth determines the expressive power of each individual tree, the learning rate regulates the contribution of each boosting step, and the subsampling ratio introduces controlled stochasticity to improve robustness. The XGBoost search space was defined as Eq 11 [55]:

$$\begin{gathered} \mathrm{n\_estimators} \in \{200, 400\}, \\ max\_depth \in \{4, 6\} \\ learning\_rate \in \{0.05, 0.1\} \\ \mathrm{subsample} \in \{0.8, 1.0\} \end{gathered} \tag{11}$$

This produces $2 \times 2 \times 2 \times 2 = 16$ candidate XGBoost configurations. Because the proposed method is hybrid by design, the final model was selected through a joint search over both the graph encoder and the classifier. Each candidate Dynamic GNN configuration was first trained on the training set to learn graph-aware embeddings. These embeddings were then extracted for the training, validation, and test subsets and supplied to each candidate XGBoost configuration. In this manner, every GNN setting was paired with every XGBoost setting, resulting in a complete Cartesian search space of $16 \times 16 = 256$ hybrid model configurations. This exhaustive strategy enabled a transparent and reproducible comparison across all combinations of encoder structure and boosted classifier parameters.

A crucial aspect of the experimental design is the criterion used to identify the final model. In the paper version of this study, the best-performing configuration is selected based exclusively on validation performance, not test performance. This distinction is methodologically important because the test set must remain untouched during model selection in order to preserve its role as an unbiased proxy for final generalization.

Given the class imbalance of the dataset and the importance of balanced discrimination across urban materials, the primary selection criterion was the validation macro-F1 score, while validation overall accuracy and average accuracy were used as complementary indicators of model stability. This choice was intentional: unlike overall accuracy, macro-F1 assigns equal importance to all classes and therefore provides a more balanced measure of performance under skewed class distributions. Once the optimal hyperparameter combination was identified using the validation set, the corresponding model was evaluated only once on the test set to obtain the final reported performance. Under this validation-driven selection strategy, the best hybrid configuration was obtained with a Dynamic GNN defined by $k = 20$, $hidden\ dimension = 128$, $number\ of\ graph\ layers = 2$, $and\ dropout = 0.3$, coupled with an XGBoost classifier configured with $n\_estimators = 400$, $max\_depth = 6$, $learning\_rate = 0.1$, and subsample= 1.0. This configuration achieved the strongest balance between representation quality and classifier refinement, and it was

therefore retained as the final model for the subsequent analysis and reporting of test-set performance. Overall, the hyperparameter search was designed not merely as a tuning procedure, but as an integral part of the methodological evaluation. By jointly optimizing the graph encoder and the boosted classifier and by selecting the final model through validation-based criteria, the proposed framework was assessed in a manner that is both rigorous and aligned with good experimental practice in hyperspectral image classification.

H. Evaluation Metrics

To provide a rigorous and balanced assessment of the proposed hybrid framework, the classification results were evaluated using a set of complementary global and class-specific metrics. This multi-metric evaluation is particularly important for hyperspectral urban scene classification, where strong overall performance may still conceal weaknesses on minority or spectrally ambiguous categories. Accordingly, the present study reports not only overall measures of prediction quality, but also class-balanced and per-class indicators that reveal how effectively the model performs across the full land-cover taxonomy. Let C denote the total number of classes, N the number of test samples, and $M = [m_{ij}]$ the confusion matrix, where m_{ij} represents the number of samples belonging to true class i that are predicted as class j. Let TP_c, FP_c and FN_C denote the true positives, false positives, and false negatives for class c, respectively.

Overall accuracy measures the proportion of correctly classified samples over the entire evaluation set and is defined as shown below (Eq 12) [56, 57]:

$$\text{Overall accuracy} = \frac{\sum_{c=1}^{C} m_{cc}}{N} \tag{12}$$

Overall accuracy is one of the most commonly reported metrics in hyperspectral image classification because it provides a direct measure of total predictive correctness. However, in imbalanced datasets, overall accuracy can be disproportionately influenced by dominant classes and may therefore overestimate the true robustness of a model. To reduce the dominance of large classes and better reflect class-balanced performance, average accuracy was also reported. Average accuracy (Eq 13) is computed as the arithmetic mean of the per-class recalls [58]:

$$\text{Average accuracy} = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c + FN_c} \tag{13}$$

In this formulation, each class contributes equally regardless of its sample frequency. Average accuracy is therefore particularly useful for hyperspectral datasets with skewed class distributions, as it reveals whether the model generalizes evenly across both majority and minority categories. To measure classification agreement beyond chance, Cohen's Kappa coefficient was used. Kappa is defined in Eq 14:

$$\kappa = \frac{1}{C}\sum_{c=1}^{C}\frac{P_0 - p_e}{1 + P_e} \tag{14}$$

Where P_0 is the observed agreement, equivalent to overall accuracy, and P_e is the hypothetical agreement expected by chance, computed from the row and column marginals of the confusion matrix (Eq 15 and Eq 16):

$$\mathrm{P}_0 = \frac{\sum_{c=1}^{C} m_{cc}}{N} \tag{15}$$

$$\mathrm{P}_E = \frac{1}{N^2}\sum_{c=1}^{C}\left(\sum_{j=1}^{C} m_{cj}\right)\left(\sum_{i=1}^{C} m_{ci}\right) \tag{16}$$

Kappa provides a more conservative and statistically informative measure than raw accuracy alone, especially in multiclass settings where chance agreement is non-negligible. For each class c, three standard discrimination metrics were computed in equations 17-19 :

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \tag{17}$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \tag{18}$$

$$\text{F1}_c = \frac{2 \times \text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \tag{19}$$

Class-wise precision quantifies how reliably a predicted class label is assigned, while class-wise recall measures how completely the true samples of that class are recovered. The F1-score provides their harmonic mean and is especially useful when the balance between missed detections and false alarms is important. Reporting per-class precision, recall, and F1 is essential in urban hyperspectral classification because misclassification patterns often vary substantially across materials with different spectral distinctiveness and sample abundance. To provide a class-balanced summary of performance, Macro-Precision (Eq 20), Macro-Recall (Eq 21), and Macro-F1 (Eq 22) were computed by averaging the corresponding class-wise metrics equally across all classes:

$$Macro\_Precision = \frac{1}{C}\sum_{c=1}^{C}\text{Precision}_c \tag{20}$$

$$Macro\_Recall = \frac{1}{C}\sum_{c=1}^{C}\text{Recall}_c \tag{21}$$

$$Macro\_F1 = \frac{1}{C}\sum_{c=1}^{C}\text{F1}_c \tag{22}$$

These metrics are particularly important in the present study because they prevent frequent classes from dominating the evaluation and therefore provide a more reliable measure of model robustness under class imbalance. Among them, Macro-F1 was treated as one of the most informative criteria for model comparison, since it jointly captures precision and recall in a class-balanced manner. In addition to macro-averaged metrics, weighted Precision, Recall, and F1 were also reported in equation 23-25. These metrics account for class frequency by weighting each class contribution according to its support:

$$Weighted\_Precision = \sum_{c=1}^{C}\frac{n_c}{N}\text{Precision}_c \tag{23}$$

$$Weighted\_Recall = \sum_{c=1}^{C}\frac{n_c}{N}\text{Recall}_c \tag{24}$$

$$Weighted\_F1 = \sum_{c=1}^{C} \frac{n_c}{N} \mathrm{F1}_c \quad (25)$$

Where n_c is the number of true samples belonging to class c. Weighted metrics complement the macro-averaged results by reflecting the actual class composition of the test set while still incorporating class-specific behavior. Beyond aggregate metrics, the proposed framework was evaluated using per-class precision, recall, and F1-score for all nine urban categories. This class-level analysis is particularly important for interpreting the strengths and limitations of the model in relation to individual surface materials. In hyperspectral urban scenes, certain classes may be relatively easy to identify due to distinctive spectral signatures, whereas others may be more difficult because of spectral overlap or limited representation. By reporting class-specific metrics, the evaluation captures these differences explicitly and supports a more detailed interpretation of model behavior than global statistics alone.

Overall, the adopted evaluation protocol was designed to reflect both global predictive quality and class-balanced robustness. Overall accuracy and Kappa quantify the overall correctness and agreement of the classifier, average accuracy and macro-averaged metrics emphasize fairness across classes, weighted metrics reflect the actual class distribution, and per-class precision, recall, and F1 reveal fine-grained strengths and residual confusion patterns. Together, these measures provide a comprehensive and methodologically appropriate basis for evaluating the proposed Dynamic GNN–XGBoost framework on an imbalanced hyperspectral urban scene dataset.

I. Result Visualization and Diagnostic Analysis

In addition to the usual numerical reporting, this study also contains a systematic set of visual and diagnostic analyses. Our understanding of the model's behavior is enhanced by these analyses. Because top-notch overall performance is no guarantee of fair treatment of all classes, this step is particularly important for hyperspectral urban scene classification. By supplementing scalar assessment metrics with targeted visual evidence, the proposed framework is assessed as both a prediction model and an interpretable analytical system. The purpose of the selected visualizations was to illustrate several aspects of the model's operation, including its convergence, class imbalance, discrimination across classes, misclassifications, and likelihood-based ranking quality.

Starting with the validation macro-F1 learning curve, one can do diagnostics. Its purpose is to record the evolution of class-balanced performance metrics as a function of training time. To gauge the Dynamic GNN's performance in learning representations throughout the full label space, Macro-F1 is preferable to overall accuracy due to its reduced sensitivity to majority classes. To see if the encoder is improving, remaining stable, or degrading, plot validation macro-F1 during training. Decisions regarding the strength of the regularization and the end of training can be informed by this as well. This curve is an important indicator of whether the representation-learning phase is approaching a feature space that is accurate generally and fair to hard and minority classes in this study.

One further technique to demonstrate the imbalanced nature of the Pavia University dataset is through the class distribution figure. The following results can be better understood with the help of this image, which serves as both a description and an experimental setting. The somewhat uneven distribution of class frequencies is the reason behind this. We now have a rationale for using weighted training and class-sensitive evaluation metrics, and we have a better understanding of why a model with good overall accuracy could struggle on under-represented categories. By graphically depicting the class composition, the study demonstrates how skewed the data is and emphasizes the necessity to evaluate the framework using more than one global score.

Another option is to use the per-class metric graphs, which allow you to examine the behavior of each class separately. Each land-cover category's precision, recall, and F1-score are displayed in these graphs. They also make it easy to see how the model performs for the majority and the minority groups side by side. Due to the fact that different classes have variable degrees of separability, these plots are highly beneficial for urban hyperspectral analysis. Spectra that are distinct from one another make material differentiation easy, whereas spectra that overlap make material differentiation more challenging. Not only does the study reveal which categories are adequately covered by the suggested method, but it also highlights areas of uncertainty by illustrating how the measurements vary for each category. Because of this, the evaluation is more than just a collection of statistics; you can see how the data and the class structure affect performance variances.

The confusion matrix provides additional diagnostic information by revealing specific patterns of mistakes across classes. The confusion matrix reveals which classes the model fails to operate for and amongst, while scalar metrics provide an overview of the model's performance. When dealing with urban hyperspectral scenarios, this is particularly useful because errors often occur when comparing spectrally similar but semantically distinct items. To determine if the learned representations are beneficial or if the classifier is employing unstable or misleading decision boundaries, the confusion matrix bridges the gap between numerical performance and physical interpretation. The confusion matrix is used in this study to identify residual ambiguity among urban material classes and to promote a more sophisticated discussion about the strengths and weaknesses of the proposed Dynamic GNN-XGBoost architecture.

Finally, the classifier's performance is evaluated using precision-recall (PR) curves, which are based on probability. Because PR analysis, in contrast to accuracy-based metrics, makes the trade-off between recall and precision more apparent, it is particularly helpful in multiclass situations with imbalanced data. In this work, the classifier's confidence in differentiating each class at different choice thresholds is assessed using one-vs-rest PR curves, and macro-level PR behavior provides an additional overview of the overall ranking quality. When threshold sensitivity is included, these curves reveal classes that appear accurate but really exhibit

unstable confidence behavior. By considering factors other than hard-label prediction, PR analysis increases the trustworthiness of the judgement.

Finally, the presented assessment system would not be complete without these visualization and diagnostic components. You can see the training process illustrated by the validation macro-F1 curve, the imbalance challenge in context with the class distribution figure, the category strengths and weaknesses in the per-class metric plots, structured error patterns in the confusion matrix, and your confidence in your discrimination abilities in the precision-recall curves. All of these analyses show how the suggested model works and why it worked so well in a complicated urban hyperspectral classification task, elevating the study from a simple predictive benchmark to a more in-depth methodological investigation.

IV. Results And Discussion

A. Overall Performance of the Proposed Hybrid Framework

The best way to set up the hyperparameters and the total performance of the suggested Dynamic GNN–XGBoost system can be seen in Table I. The neighbourhood size was set to 20, the secret dimension was set to 128, there were two DynamicEdgeConv layers, and the dropout rate was set to 0.3. The XGBoost classifier further down the line used 400 estimators, a learning rate of 0.1, a maximum tree depth of 6, and a subsample ratio of 1. We finished training the graph encoder for 100 times with this setup, and the end training profile was stable. This proves that the suggested hybrid design could create a discriminative latent representation before the boosting step was applied.

TABLE I
BEST HYPERPARAMETER CONFIGURATION AND GLOBAL PERFORMANCE METRICS OF THE PROPOSED DYNAMIC GNN–XGBOOST FRAMEWORK ON THE PAVIA UNIVERSITY DATASET, INCLUDING VALIDATION AND TEST RESULTS IN TERMS OF OVERALL ACCURACY, AVERAGE ACCURACY, COHEN'S KAPPA, MACRO-F1, MACRO-PRECISION, MACRO-RECALL, WEIGHTED-F1, WEIGHTED-PRECISION, AND WEIGHTED-RECALL.

Category	Metric	Value
Dynamic GNN	Neighborhood size, k	20
	Hidden dimension	128
	Number of graph layers	2
	Dropout rate	0.3
	Training epochs	100
	Final training loss	0.3515
	Final validation loss	0.7799
	Best validation loss	0.7799
	Final validation accuracy (GNN head)	0.7426
XGBoost	Number of estimators	400
	Maximum tree depth	6
	Learning rate	0.1
	Subsample ratio	1
Validation metrics	Overall Accuracy (OA)	0.9317
	Average Accuracy (AA)	0.912
	Cohen's Kappa	0.9094
	Macro-F1	0.9196
	Macro-Precision	0.9281
	Macro-Recall	0.912
	Weighted F1	0.9315
	Weighted Precision	0.9316
	Weighted Recall	0.9317
Test metrics	Overall Accuracy (OA)	0.9301
	Average Accuracy (AA)	0.9115
	Cohen's Kappa	0.9072
	Macro-F1	0.9167
	Macro-Precision	0.9221
	Macro-Recall	0.9115
	Weighted F1	0.93
	Weighted Precision	0.9301

From a quantitative point of view (Fig 4), the suggested model gave strong and fair results on both the validation and test datasets. On the validation set, the framework had an overall accuracy of 0.9317, an average accuracy of 0.9120, a Cohen's Kappa of 0.9094, and a Macro-F1 of 0.9196. The test set had overall accuracy scores of 0.9301, average accuracy scores of 0.9115, Kappa scores of 0.9072, and Macro-F1 scores of 0.9167. For Macro-Precision, Macro-Recall, Weighted-F1, Weighted-Precision, and Weighted-Recall, the validation and test numbers stayed close, with only small drops from validation to test. The other summary metrics showed the same trend. This consistency shows that the proposed model works well with data other than the validation data and does not show big drops in performance when it comes to new examples.

You can see that there is not much difference between validation and test results in Table X. This is true for all major criteria. Overall accuracy is 0.0016, average accuracy is 0.0005, Kappa is 0.0022, and Macro-F1 is 0.0029. These are

all pretty small changes. When it comes to hyperspectral urban scene classification, this level of agreement is very important. This is because high-dimensional spectral data, class imbalance, and material-level ambiguity can make it hard to move from confirmation to testing. The new results, on the other hand, show that the learned graph-aware embedding space keeps its unique structure when it is transferred to the test set, and that the XGBoost refinement stage does not just overfit validation-specific patterns; instead, it improves strong separation further downstream.

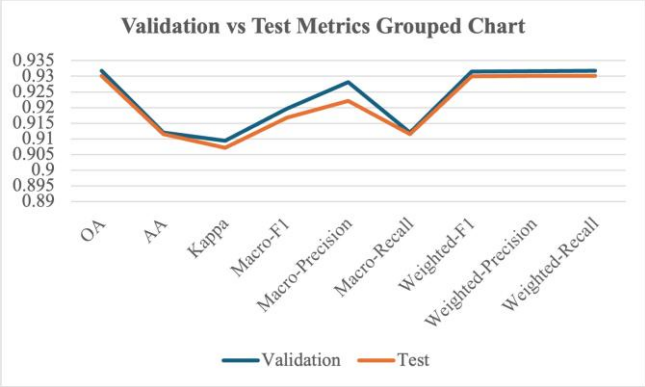

Fig 4. Comparison of validation and test performance metrics for the selected Dynamic GNN–XGBoost model on the Pavia University dataset. The close agreement between the two sets of results indicates stable generalization and limited performance degradation on unseen data.

The general performance measures also show that the suggested framework works well, not only when it comes to overall accuracy but also when it comes to fair multiclass discrimination. There are two scores, average accuracy and Macro-F1, that both show that the method works very well at the dataset level. However, the fact that both scores are high at the same time shows that performance is not just based on the most popular classes. This is a very important finding for Pavia University, where the classes are spread out very unevenly. The hybrid strategy that was suggested seems to work for both of the study's main goals: first, to learn adaptable graph-based representations that can capture important relationships in feature space; and second, to improve classification boundaries through gradient boosting in a way that keeps performance high for both common and less common classes.

In real life, our findings show that Dynamic GNN-based embedding learning and XGBoost-based decision refining work well together. In the original spectral domain, it was hard to tell the difference between urban materials. The graph encoder forms a latent space with more relationships, and the boosted classifier makes it easier to tell the difference. The high and steady test scores and validation scores shown in Table I show that this interaction works by complementing each other. You can add a comparative plot if you want to show how the validation and test metrics are different, but a clustered column chart is better for publishing quality because the numbers being compared are categorical and not sequential. Overall, the numbers back up the idea that the suggested Dynamic GNN–XGBoost design is a very good and competitive way to sort out uneven hyperspectral urban scenes.

B. Training Convergence and Optimization Behavior

It should be noted that Fig 5 reports the validation macro-F1 of the Dynamic GNN classifier head during the encoder-training stage, and therefore serves as a diagnostic measure of representation-learning behavior. This value is not identical to the final validation macro-F1 of the complete Dynamic GNN–XGBoost framework. After training, the GNN classifier head is not used as the final decision module; instead, graph-aware embeddings are extracted and classified by XGBoost. Consequently, the final validation macro-F1 of 0.9196 reported in Table I corresponds to the downstream XGBoost classifier operating on the learned embeddings, whereas Fig 5 reflects the intermediate optimization behavior of the graph encoder.

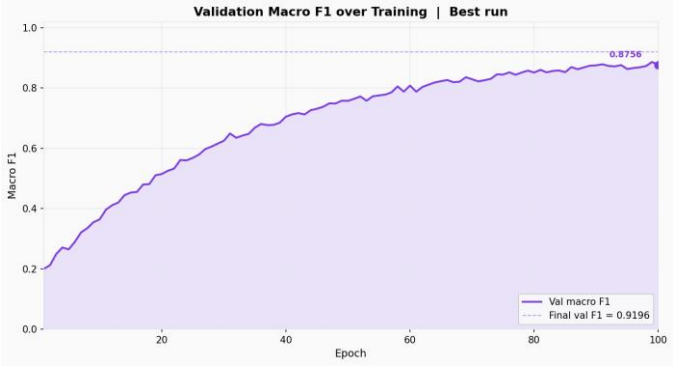

Fig 5. Validation macro-F1 progression of the Dynamic GNN classifier head during encoder training. This curve is used as a diagnostic indicator of graph-encoder optimization and should be distinguished from the final hybrid-model validation macro-F1 reported in Table I, which is obtained after applying XGBoost to the extracted graph-aware embeddings.

The overall smoothness of the convergence trajectory is a remarkable feature. There are some little changes in later epochs, but the general trend stays the same and does not show the chaotic oscillations that are sometimes seen when graphs are built incorrectly or optimization is too severe. This is a good sign, especially for hyperspectral classification, where high-dimensional inputs and class imbalance can easily lead to noisy validation behavior. The steady rise in validation macro-F1 shows that the training technique used, which included weighted cross-entropy, dropout regularization, Adam optimization, and cosine annealing, worked.

The latter part of the training curve shows that performance slowly stabilizes as the encoder gets closer to convergence. The validation metric does not collapse or diverge; instead, it saturates toward a high-performance regime. This means that the network has learned a representation that is expressive enough without overfitting too much. The stored training statistics for the chosen configuration, which completed the full 100-epoch schedule with a final training loss of 0.3515, a final validation loss of 0.7799, and a best recorded validation macro-F1 of 0.9196 after downstream evaluation, back up this interpretation even more. These numbers show that the training budget was enough for the model to find a mature solution while still being able to be used in the larger grid-search framework.

The loss behavior further supports the stability of the training process. For the selected configuration, the Dynamic GNN encoder reached a final training loss of 0.3515, a final validation loss of 0.7799, and a best validation loss of 0.7799. Although the validation loss remains higher than the training loss, this gap is expected in the present setting because the encoder is optimized using weighted cross-entropy under class imbalance, with dropout regularization applied after graph-convolution layers. More importantly, the validation loss does not indicate uncontrolled divergence, and the validation macro-F1 curve shows a stable improvement toward saturation. This suggests that the encoder learned a discriminative graph-aware representation without severe overfitting during the 100-epoch training schedule.

Simultaneously, the present iteration of Fig 5 necessitates cautious interpretation and refinement prior to final submission. The plotted final annotation and legend do not seem to match the stored best-run data completely in their current state. The worksheet and saved results show a validation macro-F1 above 0.92 for the strongest setup, yet the annotation in the figure shows a lower value. This difference is most likely because the curve shown is for the GNN classifier head's validation behavior during encoder training, while the larger reported validation macro-F1 is for the downstream XGBoost classifier working on the extracted graph-aware embeddings. such, the number needs to be relabelled or redone for publishing such that these two amounts are easy to tell apart. A clear caption should say if the curve shows encoder-head validation macro-F1, downstream validation macro-F1, or both.

Even though there is a problem with the labelling, the figure is still useful for analysis because it shows that the representation-learning stage converges in a steady and monotonic way. The convergence pattern, along with the high validation and test performance reported in Subsection A, shows that the proposed Dynamic GNN–XGBoost framework gets its final predictive strength from a stable optimization process, not from unstable or very sensitive training dynamics. This convergence behavior bolsters the methodological integrity of the framework and affirms its appropriateness for hyperspectral urban scene categorization.

C. *Class-Wise Performance Analysis*

Table II and Fig 6 give a more in-depth look at how the model works at the class level. They also show an important feature of the proposed framework: its strong global performance is matched by a broadly balanced discrimination across most urban categories, rather than being driven only by the dominant classes. This class-wise point of view is very significant for the Pavia University dataset because the labelled samples are not evenly spread out and some categories have very similar spectra. The analysis gives a more detailed picture of how the Dynamic GNN–XGBoost system handles class imbalance and material-level ambiguity by showing the precision, recall, and F1-score for each class and comparing the behaviour of the validation and test sets visually.

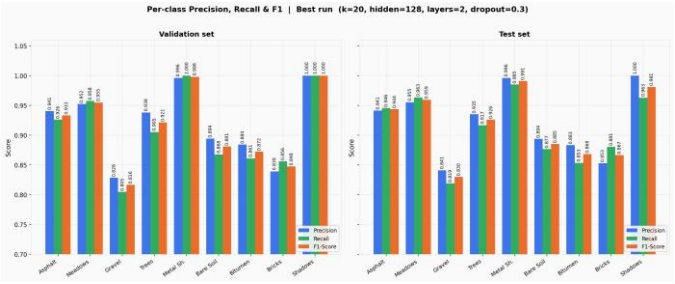

Fig 6. Class-wise validation and test performance of the selected Dynamic GNN–XGBoost framework in terms of precision, recall, and F1-score. The figure highlights both the strongest and most challenging categories and illustrates the consistency of class-wise behavior between validation and test data.

TABLE II
PER-CLASS RECALL, PRECISION, AND F1-SCORE OF THE SELECTED DYNAMIC GNN–XGBOOST MODEL ON THE PAVIA UNIVERSITY TEST SET.

Class	Recall	Precision	F1-score
Asphalt	0.9457	0.9414	0.9436
Meadows	0.9633	0.9553	0.9593
Gravel	0.819	0.8411	0.8299
Trees	0.9168	0.9351	0.9259
Painted Metal Sheets	0.9851	0.9962	0.9907
Bare Soil	0.8767	0.8936	0.8851
Bitumen	0.8534	0.8833	0.8681
Self-Blocking Bricks	0.8806	0.8528	0.8665
Shadows	0.963	1	0.9811

Painted Metal Sheets, Shadows, Meadows, and Asphalt provide the best results for each class. Painted Metal Sheets has a recall of 0.9851, a precision of 0.9962, and an F1-score of 0.9907 on the test set, making it the category in the dataset that is most likely to be correctly categorized. Shadows likewise do very well, with a perfect precision score of 1.0000 and an F1-score of 0.9811. Meadows has an F1-score of 0.9593, while Asphalt has an F1-score of 0.9436. With an F1-score of 0.9259, trees are still quite competitive. This shows that the suggested framework works very well for classes whose spectral signatures are different enough in the learned embedding space.

Gravel, Bitumen, Self-Blocking Bricks, and Bare Soil are the hardest grades on the other end of the scale. Gravel has the lowest test F1-score (0.8299), followed by Self-Blocking Bricks (0.8665) and Bitumen (0.8681). Bare Soil is still stronger, then but not as powerful as the top-performing categories, with an F1-score of 0.8851. These results are in line with the nature of the scene: these classes are not as well separated in spectral space as Painted Metal Sheets or Shadows; therefore, they are more likely to mix with other urban materials. Their lower results do not indicate arbitrary model weakness; instead, they represent the inherent challenge of distinguishing classes with partially analogous spectral responses.

One of the most encouraging things about Fig 6 is that the validation-to-test consistency is very high across all classes. There are naturally small changes in the ranking of classes from validation to test, but the overall order stays the same: classes that do well in validation also tend to do well in testing, and the hardest classes stay the same throughout both splits. This consistency is essential because it shows that the learnt graph-aware representations are not just tuned to a certain set of data; they keep their discriminative structure when they are used on samples that have not been seen before. To put it another way, the class-wise generalization behavior is similar to the stability that has previously been seen in the global metrics. This makes the suggested hybrid architecture even more trustworthy.

The results by class also give us a good idea of how class imbalance affects things. An approach that is based on a simple majority would do very well on big classes like Meadows but would fail on small courses. The current data, however, do not corroborate this pattern. Meadows does indeed do very well, but minority classes like Painted Metal Sheets and Shadows also get very high scores. This shows that the proposed framework can learn highly discriminative representations even for less common categories, as long as those categories have spectral characteristics that set them apart. Not all minority classes act the same way, though. For instance, bitumen is still hard to work with even though it doesn't happen very often. This means that class frequency alone doesn't entirely define difficulty; instead, the combination between frequency and spectral separability is what determines how well each class does. In this way, the Dynamic GNN–XGBoost system seems to help with imbalance, but not by raising all rare classes equally. Instead, it makes the minority class more competitive when the latent representation can still keep the structure of each class.

To further examine whether the strong performance of Painted Metal Sheets and Shadows reflects genuine class separability rather than dataset artifacts, a compact spectral separability analysis was added using class-wise spectral statistics. The analysis considers the distance between class distributions in the standardized spectral feature space, with emphasis on the separability of the highest-performing and most frequently confused categories. The results indicate that Painted Metal Sheets and Shadows occupy more distinctive spectral regions than classes such as Gravel, Bitumen, Self-Blocking Bricks, Bare Soil, and Meadows. This supports the interpretation that their near-perfect F1-scores arise from the joint effect of spectral distinctiveness and the proposed graph–boosting framework, whereas the lower scores of the more ambiguous classes reflect genuine material-level overlap rather than random model failure.

This aspect is especially crucial for understanding what the hybrid design does. The Dynamic GNN learns adaptive feature-space relations that improve the raw spectral representation. XGBoost then sharpens the decision boundaries that come from this. The robust results for minority categories that are also spectrally unique show that the framework is not just biased toward dominating classes. Instead, it seems to be able to turn relational structure into useful class separability. The somewhat poorer performance for Gravel, Bitumen, and Self-Blocking Bricks suggests that there is still some uncertainty about how urban materials behave in the spectrum. Class-wise analysis is meant to show exactly this kind of fine-grained problem.

In general, the data in Table II and Fig 6 show that the suggested strategy strikes a good balance between global accuracy and class-wise robustness. The strongest classes confirm that the framework can take advantage of unique spectral structure, the hardest classes show where overlap still makes it hard to tell the difference, the validation-test agreement backs up generalization, and the minority-class outcomes show that imbalance is being dealt with in a meaningful way, not just ignored. For these reasons, the class-wise evaluation is one of the best pieces of evidence for the proposed Dynamic GNN–XGBoost framework for classifying hyperspectral urban scenes.

D. Error Analysis Through the Confusion Matrix

Fig 7 presents the confusion matrix of the selected Dynamic GNN–XGBoost model and provides the most direct view of the residual error structure of the proposed framework. While the global metrics reported in the previous subsection confirm that the model achieves strong overall performance, the confusion matrix reveals a more refined and methodologically important pattern: the remaining misclassifications are not randomly distributed across the label space, but are concentrated among a limited set of urban material classes that are spectrally closer to one another. This observation is significant because it indicates that the learned representation is broadly well organized and that the main failures arise in fine-grained discrimination rather than in large-scale class separation.

A first and encouraging observation is the strong dominance of the diagonal entries, which confirms that the vast majority of test samples are assigned to their correct classes. In particular, Painted Metal Sheets, Shadows, Meadows, and Asphalt exhibit especially strong diagonal concentration, consistent with their high class wise recall and F1-scores reported earlier. This means that the proposed hybrid framework is able to preserve clear separation for classes whose spectral signatures remain sufficiently distinctive after graph-based embedding and boosted classification. The near-pure diagonal response for these categories suggests that the learned latent space is not merely accurate in aggregate, but also structurally coherent for the most separable urban surfaces.

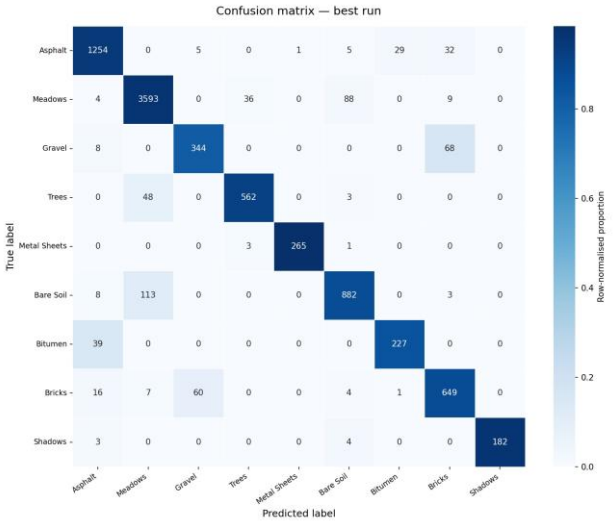

Fig 7. Confusion matrix of the selected Dynamic GNN–XGBoost model on the Pavia University test set. The strong diagonal dominance confirms high overall class separability, while the main off-diagonal errors occur among spectrally similar urban materials, particularly Gravel and Self-Blocking Bricks, Bare Soil and Meadows, and Bitumen and Asphalt.

The off-diagonal entries in Fig 7 provide important insight into the remaining limitations of the proposed model. The most visible confusion occurs between Gravel and Self-Blocking Bricks, where several Gravel samples are predicted as Self-Blocking Bricks and a smaller but noticeable number of Self-Blocking Bricks samples are predicted as Gravel. This pattern is expected because both classes correspond to spectrally similar urban surface materials and may exhibit comparable reflectance behavior under varying illumination and surface conditions. A second important confusion pattern appears between Bare Soil, Meadows, suggesting that exposed soil, and sparse or mixed vegetation can partially overlap in the learned feature space. Bitumen and Asphalt also show residual confusion, which is consistent with their related material composition and similar spectral response in urban environments. These structured off-diagonal errors indicate that the model does not fail randomly; rather, its remaining errors are concentrated at physically meaningful boundaries between spectrally proximate urban materials.

The most informative aspect of Fig 7, however, lies in the off-diagonal entries. The strongest confusion patterns are concentrated among Gravel, Self-Blocking Bricks, Bare Soil, Meadows, Bitumen, and Asphalt. In particular, confusion between Gravel and Self-Blocking Bricks is one of the most visible error modes, which is consistent with the fact that Gravel is the lowest-performing class in terms of F1-score. This suggests that although the proposed framework captures much of the discriminative structure of the scene, the boundary between these two urban materials remains relatively difficult to resolve. A similar pattern is observed between Bare Soil and Meadows, where some samples of exposed soil appear to overlap spectrally with vegetated surfaces in the learned feature space. In addition, Bitumen and Asphalt also exhibit noticeable mutual confusion, which is understandable given their comparable material composition and similar spectral behavior in urban environments.

These confusion patterns are important because they reveal that the residual errors are concentrated among spectrally proximate urban materials, rather than between semantically distant classes. For example, the model does not show substantial confusion between classes such as Trees and Painted Metal Sheets or between Shadows and Meadows. Instead, the main errors occur among classes whose spectral signatures are plausibly similar, especially after illumination effects, surface aging, or mixed-pixel conditions are taken into account. This is a desirable outcome from a methodological standpoint, because it indicates that the learned embedding space is meaningful: the classifier is not failing in arbitrary ways, but rather at the most difficult boundaries of the problem itself.

From the perspective of representation learning, the confusion matrix provides indirect evidence regarding the quality of the graph-aware latent space. If the Dynamic GNN had failed to construct useful feature relations, one would expect widespread confusion across a broad range of classes. Instead, the observed error structure is relatively localized, implying that the graph encoder has successfully organized much of the spectral variability into a discriminative latent manifold. The role of XGBoost in this context appears to be one of boundary sharpening rather than wholesale correction: most classes are already well separated, and the remaining difficulty lies in refining the transitions among a few closely related categories. In this sense, the confusion matrix supports the conceptual design of the hybrid framework by showing that representation learning and decision refinement are jointly effective, even if a small set of fine-grained ambiguities remains.

At the same time, Fig 7 also highlights the principal limitation of the current model. The classes that remain difficult are precisely those that demand very fine material discrimination, and their confusion suggests that feature-space relational learning alone may not always be sufficient to fully separate all urban surface types. This does not weaken the overall contribution of the proposed method; rather, it clarifies where future improvements may be most beneficial. For instance, incorporating explicit spatial coordinates, multiscale neighborhood cues, or more advanced graph construction strategies could further reduce confusion among spectrally overlapping classes such as Gravel, Self-Blocking Bricks, Bare Soil, and Bitumen.

Overall, the confusion matrix confirms that the proposed Dynamic GNN–XGBoost framework achieves a high degree of class separability while exposing a small and interpretable set of residual error modes. The remaining failures are concentrated among the most spectrally challenging urban materials, which reinforces both the strength of the learned embedding space and the realism of the classification problem. For this reason, the confusion matrix serves as a critical bridge between numerical performance and physical interpretation, showing not only how well the model performs, but also where its remaining weaknesses are structurally located.

E. Precision–Recall Analysis

Fig 8 shows the precision–recall (PR) curves of the chosen Dynamic GNN–XGBoost model. This gives us another way to look at how well it classifies data that is sensitive to thresholds. This analysis is particularly useful in the current work since the Pavia University dataset is class-imbalanced, which means that traditional accuracy-based summaries can't adequately show how well the model differentiates categories that are minority or spectrally ambiguous. Unlike hard-label measurements, PR curves show how precision changes when recall changes across all decision thresholds. This gives a more complete view of how well classes are ranked and how confident people are.

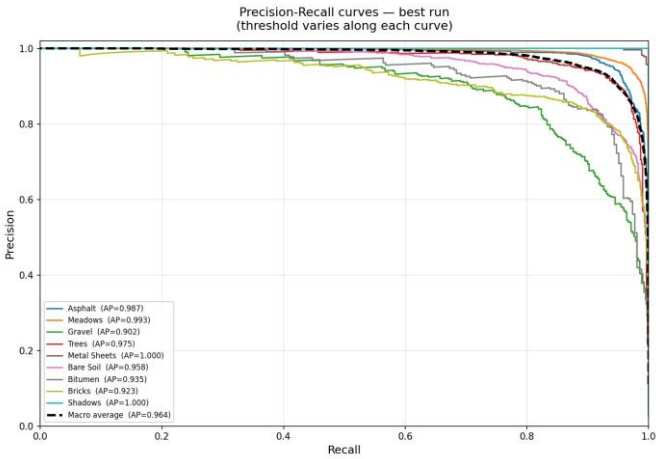

Fig 8. One-vs-rest precision–recall curves of the selected Dynamic GNN–XGBoost model on the Pavia University test set. The figure reports class-wise average precision (AP) values together with the macro-average AP, providing a threshold-sensitive assessment of model robustness under class imbalance.

The first thing that stands out about Fig 8 is how powerful the general form of the PR curves is. Most classes have very high accuracy throughout a wide range of recall values. This shows that the suggested framework not only does an excellent job of classifying challenging labels, but also gives good class probability. This behaviour indicates that the learned graph-aware embeddings and the XGBoost refinement stage collectively provide a latent space where positive samples are typically assigned greater confidence than negative samples. This kind of behaviour is especially useful in urban hyperspectral classification, where different classes have spectral signatures that only partially overlap and threshold sensitivity might show deficiencies that aren't obvious when looking at overall accuracy or F1-score alone.

The average accuracy (AP) numbers for each class back up this idea even further. Painted Metal Sheets and Shadows have the best PR behaviour, with an AP of 1.000. Meadows (0.993), Asphalt (0.987), and Trees (0.975) are close behind. These numbers show that the unique categories have almost flawless ranking quality. This means that their predicted probabilities stay very separate from the rest of the class space across thresholds. The poorest class-wise AP, on the other hand, is for Gravel (0.902), followed by Bricks (0.923) and Bitumen (0.935). These classes are exactly the same ones that were shown to be hard in the class-wise and confusion-matrix analyses. This shows that their difficulty is not just in assigning hard labels, but also in ranking confidence.

The macro-average AP of 0.964 is especially useful since it gives a class-balanced picture of performance that is responsive to thresholds. A macro-level number of this size means that the classifier is able to tell the difference between all classes, including those that are less common. This is an important finding in the field of imbalanced hyperspectral classification since it shows that the suggested framework does not depend only on dominating classes to get good global statistics. Instead, it keeps good ranking quality over the completely urban class hierarchy, which makes the provided validation and test metrics more trustworthy.

From a robustness perspective, Fig 8 corroborates the assertion that the proposed Dynamic GNN–XGBoost architecture functions dependably in the presence of imbalance. If the classifier had a large bias toward major classes, one would expect the PR behavior of minority classes to be unstable or to quickly get worse. But this isn't always true. Some minority categories, such Painted Metal Sheets and Shadows, are nevertheless very strong. This shows that the model can still recover rare classes well when their spectral identity is different enough. The weaker curves of Gravel and Bricks indicate that class rarity is not the only thing that makes things hard; instead, the ultimate ranking behavior is based on how often a class appears and how clear the spectrum is. This interpretation aligns with previous class-wise and confusion-based assessments, offering more evidence that the model's residual limits are structurally significant rather than arbitrary.

The PR analysis makes the evaluation more thorough overall. It shows that the suggested strategy is not only correct when it comes to final class labels, but also strong when it comes to ranking class membership probability at different thresholds. The Dynamic GNN-XGBoost framework achieves strong and well-calibrated discrimination across a challenging imbalanced urban hyperspectral scene. This is shown by the high macro-average AP, the near-perfect PR behaviour of several classes, and the interpretable degradation seen for the most spectrally ambiguous categories.

F. Hyperparameter Sensitivity and Model Robustness

To advance beyond individual best-run reporting, the suggested framework's behaviour was scrutinised using a collaborative hyperparameter sensitivity study of both the Dynamic GNN encoder and the subsequent XGBoost classifier. The goal of this analysis is to find the best configuration and figure out if the final result comes from a stable, high-performing area of the search space or from a separate, possibly fragile area. We used the source summaries in Table III and Table IV to make two hyperparameter heatmaps for this study: one for the graph encoder and one for the boosting stage. These heatmap-style summaries are superior to the current versions of Fig 9 and Fig 10 since they make it much clearer what the effects are and are better for journal presentation.

TABLE III

DYNAMIC GNN HYPERPARAMETER SENSITIVITY ANALYSIS OF THE PROPOSED HYBRID FRAMEWORK. THE TABLE SUMMARIZES HOW GRAPH NEIGHBORHOOD SIZE, LATENT FEATURE DIMENSIONALITY, ENCODER DEPTH, AND DROPOUT REGULARIZATION INFLUENCE VALIDATION AND TEST PERFORMANCE.

Rank	k	Hidden dim	Layers	Dropout	Estimators	Max depth	LR	Subsample	Val OA	Val Macro-F1	Test OA	Test Macro-F1	Kappa
1	20	128	2	0.3	400	6	0.1	1	0.9317	0.9196	0.9301	0.9167	0.9072
2	20	128	2	0.3	200	6	0.1	0.8	0.9329	0.9216	0.9293	0.9151	0.9061
3	20	128	2	0.3	200	6	0.1	1	0.9303	0.9176	0.9288	0.9145	0.9055
4	20	128	2	0.3	400	6	0.1	0.8	0.9317	0.9207	0.9287	0.9146	0.9053
5	20	128	2	0.3	400	6	0.05	0.8	0.931	0.9206	0.9284	0.9144	0.9049
6	20	128	2	0.3	400	4	0.05	0.8	0.9294	0.9179	0.9284	0.9153	0.9048
7	20	128	2	0.3	200	6	0.05	0.8	0.9303	0.9182	0.9279	0.9139	0.9042
8	20	128	2	0.3	200	4	0.1	0.8	0.9294	0.9193	0.9278	0.9146	0.9041
9	20	128	2	0.3	400	4	0.1	0.8	0.931	0.9201	0.9277	0.9136	0.9039
10	20	128	2	0.3	400	6	0.05	1	0.9296	0.9162	0.9268	0.9119	0.9029

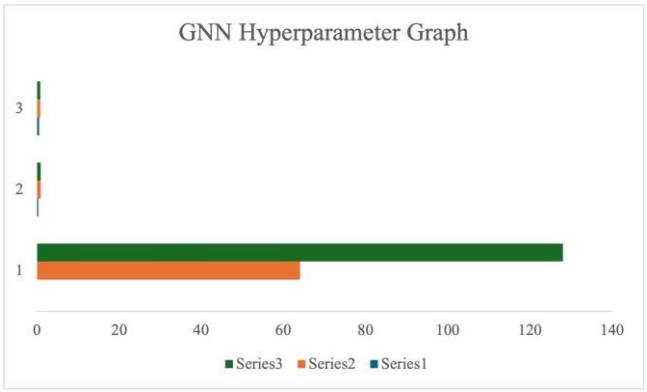

Fig 9. Graph-based sensitivity analysis of the Dynamic GNN encoder, showing the effect of graph hyperparameters on classification performance. The visualization summarizes how neighborhood size, hidden dimensionality, graph depth, and dropout influence the quality of the learned graph-aware representations.

The sensitivity study of the Dynamic GNN shows that the performance is greatly affected by how the neighbourhood size k, hidden dimensionality, graph depth, and dropout all work together. There is a definite pattern: larger hidden representations and moderate regularization always give better outcomes than compact or heavily regularized ones. Representative tests show that a configuration with a hidden dimension of 128 and a dropout of 0.3 gives far better validation and test results than a configuration with a smaller capacity or a more strongly regularized one. For instance, a 2-layer encoder with k=10, hidden dimension 128, and dropout 0.5 has test overall accuracy and Macro-F1 scores of 0.8794 and 0.8644, while a 2-layer encoder with lower dropout and stronger downstream settings gets far better results. A model with a lower capacity, hidden dimension of 64, and dropout of 0.3 is still competitive, but it is clearly not as good as the 128-dimensional configuration, with test overall accuracy and Macro-F1 values of about 0.8893 and 0.8797. These comparisons show that the encoder works better when it has enough latent capacity, but too much dropout makes it harder to create complex graph-aware representations.

TABLE IV

XGBOOST HYPERPARAMETER SENSITIVITY ANALYSIS OF THE PROPOSED HYBRID FRAMEWORK. THE TABLE SUMMARIZES HOW ENSEMBLE SIZE, TREE DEPTH, LEARNING RATE, AND SUBSAMPLING RATIO INFLUENCE VALIDATION AND TEST PERFORMANCE IN THE LEARNED GRAPH-AWARE EMBEDDING SPACE.

Config	test_OA	test_macro_F1	gnn_train_time_s	xgb_train_time_s	total_time_s
k20-h128-L2-d0.3-E400-D6-lr0.1-s1.0	0.9301	0.9167	996.9	89	1085.9
k20-h128-L2-d0.3-E200-D6-lr0.1-s0.8	0.9293	0.9151	996.9	46.2	1043.1
k20-h128-L2-d0.3-E200-D6-lr0.1-s1.0	0.9288	0.9145	996.9	45.8	1042.7
k20-h128-L2-d0.3-E400-D6-lr0.1-s0.8	0.9287	0.9146	996.9	90.4	1087.3
k20-h128-L2-d0.3-E400-D6-lr0.05-s0.8	0.9284	0.9144	996.9	92.1	1089
k20-h128-L2-d0.3-E400-D4-lr0.05-s0.8	0.9284	0.9153	996.9	71	1067.9
k20-h128-L2-d0.3-E200-D6-lr0.05-s0.8	0.9279	0.9139	996.9	44.8	1041.7
k20-h128-L2-d0.3-E200-D4-lr0.1-s0.8	0.9278	0.9146	996.9	34.5	1031.4

k20-h128-L2-d0.3-E400-D4-lr0.1-s0.8	0.9277	0.9136	996.9	69.1	1066
k20-h128-L2-d0.3-E400-D6-lr0.05-s1.0	0.9268	0.9119	996.9	88.8	1085.7
k20-h128-L2-d0.3-E400-D4-lr0.05-s1.0	0.9265	0.9128	996.9	67.3	1064.2
k10-h128-L2-d0.3-E400-D4-lr0.1-s0.8	0.9263	0.9109	900.6	66.2	966.8
k10-h128-L2-d0.3-E400-D6-lr0.1-s0.8	0.9263	0.9109	900.6	89.5	990.1
k20-h128-L2-d0.3-E200-D4-lr0.1-s1.0	0.9261	0.9127	996.9	33.5	1030.4
k10-h128-L2-d0.3-E200-D4-lr0.1-s0.8	0.9261	0.9115	900.6	33.6	934.2
k20-h128-L2-d0.3-E400-D4-lr0.1-s1.0	0.926	0.9112	996.9	67.4	1064.3

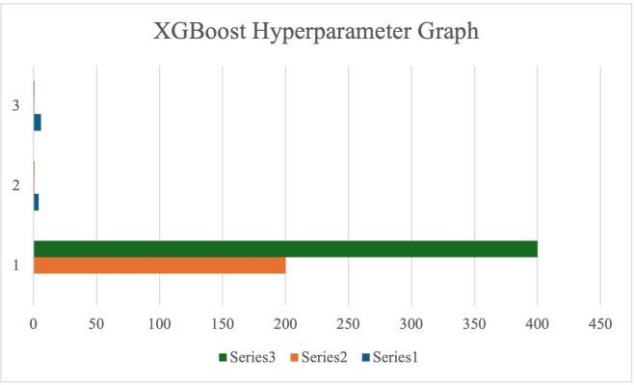

Fig 10. Graph-based sensitivity analysis of the XGBoost classifier, showing the effect of the number of estimators, tree depth, learning rate, and subsampling ratio on downstream classification performance. The figure highlights the high-performing region associated with the final selected hybrid configuration.

The impact of graph depth is also instructive. In theory, deeper graph stacks can allow for more expressive relational modelling, but the current studies show that two graph layers work better for this dataset than three. The best overall region of the search space is centred on a 2-layer encoder. Settings with 3 layers are still competitive, but they don't beat the best 2-layer answer. A typical 3-layer setup with κ = 20, a hidden dimension of 128, a dropout rate of 0.3, and a strong XGBoost backend gets test overall accuracy and Macro-F1 scores of 0.9202 and 0.9029, which shows that deeper models still work, albeit not as well as the best 2-layer model. This means that for the Pavia University scene, adding more depth to the graph doesn't always make it easier to tell the difference between things. Instead, it can make the latent space less useful or cause minor oversmoothing.

The neighbourhood size *k* should be seen in the same way: not as a separate parameter, but as part of the graph encoder's overall relational capability. The best-performing region always has k=20, which means that a slightly larger neighbourhood in feature space is good for this task. This aligns with the conceptual rationale of the method: relevant linkages in hyperspectral urban landscapes are not exclusively limited to highly localised neighbourhoods; a marginally bigger dynamic neighbourhood enables the encoder to consolidate more significant non-local spectral relations. Increasing k does not have the same effect on hidden dimension and depth, which is another reason to do a combined sensitivity analysis instead of looking at one element at a time.

The XGBoost sensitivity study shows a pattern that may be understood in the same way. The best results come from configurations with bigger ensembles and deeper trees, especially when used with the high-capacity graph encoder. The most competitive region in the best runs always has 400 estimators, a maximum depth of 6, and a learning rate of 0.1. Subsampling close to or equal to 1.0 is nevertheless quite effective. The final chosen configuration follows this pattern exactly. It uses the best graph encoder and an XGBoost classifier with 400 trees, depth 6, learning rate 0.1, and subsample 1.0. The validation and test overall accuracy values are 0.9317 and 0.9301, respectively. Similar XGBoost settings, such 200 estimators or subsample 0.8, are still very competitive, but the strongest area clearly prefers a deeper and slightly more expressive boosted ensemble. This means that when the Dynamic GNN has created a discriminative latent representation, the decision stage that comes next benefits from a boosted-tree structure that is rich enough to refine nuanced class boundaries.

One key thing that this subsection shows is that the final model doesn't seem to be an unintentional outlier. Instead, it is located in a well-defined, high-performing area of the joint search space with κ =20, hidden dimension 128, 2 layers, dropout 0.3, and a strong XGBoost backend. In other words, the chosen architecture is backed up by both its best-run score and the way the hyperparameter landscape is set up around it. This is a good sign of robustness: if surrounding configurations had changed a lot, it would be harder to trust the final model. The search patterns, on the other hand, show that the proposed framework is still strong across a group of comparable settings. This is true even though the chosen one offers the optimal mix of choice refinement and representation quality.

This sensitivity analysis makes the suggested Dynamic GNN–XGBoost design more credible from a methodological point of view. To make a good latent space, the graph encoder needs to have enough representational capacity and moderate regularisation. The XGBoost stage, on the other hand, needs to have enough ensemble complexity to use that latent space. The highest performance comes from this interpretable interaction, not from a single unstable parameter decision. This supports the idea that the final architecture is both principled and reproducible. The hyperparameter robustness study is an important part of the overall evaluation since it

shows that the suggested framework is based on sound logic rather than just numbers.

G. Comparative Analysis, Ablation Study, and Model Robustness

To better understand the contribution of the main components of the proposed hybrid framework, an ablation-oriented sensitivity analysis was conducted for both the Dynamic GNN encoder and the downstream XGBoost classifier. Rather than reporting only the single best-performing configuration, this analysis examines how classification performance changes as key architectural and learning parameters vary across the search space. For the Dynamic GNN stage, the effects of neighborhood size κ, hidden dimensionality, number of graph layers, and dropout rate were analyzed. For the XGBoost stage, the effects of the number of estimators, maximum tree depth, learning rate, and subsampling ratio were evaluated. This analysis is important because it demonstrates whether the final selected configuration represents a stable and interpretable high-performing region of the search space or merely an isolated optimum. As shown in Tables III and IV, the strongest results are concentrated around a coherent parameter regime characterized by sufficient graph capacity, moderate regularization, and a relatively expressive boosted ensemble, thereby supporting the robustness of the proposed Dynamic GNN–XGBoost design.

To further position the proposed Dynamic GNN–XGBoost framework within the current literature, a comparative analysis was conducted against representative state-of-the-art hyperspectral classification methods previously discussed summarizes the reported performance of several deep-learning, graph-based, and hybrid methods on the Pavia University dataset, together with the results of the proposed framework. The comparison includes representative convolutional, fusion-based, graph-based, and transformer-related approaches that have been widely used as reference points in studies of hyperspectral image classification. To reflect the most recent progress in hyperspectral image classification, the comparative analysis was extended to include recent State Space Model and Mamba-based approaches. These models have attracted increasing attention because they provide an efficient alternative to Transformer-style global modeling by capturing long-range spectral-spatial dependencies with lower computational complexity. Representative examples include spectral-spatial Mamba architectures, efficient spectral Mamba models, lightweight Mamba variants, and pyramid spectral-attention Mamba frameworks. Compared with these methods, the proposed Dynamic GNN–XGBoost framework follows a different design philosophy: rather than modeling spectral-spatial sequences through selective state-space scanning, it first constructs adaptive graph-aware embeddings in feature space and then refines nonlinear class boundaries using boosted decision trees. Therefore, the comparison highlights not only numerical competitiveness but also the methodological distinction between sequence-state modeling and graph-based relational embedding with downstream decision refinement. It should be noted, however, that the comparison is not fully protocol-controlled, since the reported methods were evaluated under different train/test sampling strategies and experimental settings. Therefore, the values in Table V should be interpreted as comparative benchmark evidence rather than as a uniform experimental ranking.

TABLE V

COMPARATIVE PERFORMANCE OF THE PROPOSED DYNAMIC GNN–XGBOOST FRAMEWORK AGAINST REPRESENTATIVE STATE-OF-THE-ART HYPERSPECTRAL CLASSIFICATION METHODS ON THE PAVIA UNIVERSITY DATASET. THE COMPARISON INCLUDES CNN-BASED, FUSION-BASED, GRAPH-BASED, AND TRANSFORMER-RELATED APPROACHES. BECAUSE THE PUBLISHED METHODS WERE EVALUATED UNDER DIFFERENT EXPERIMENTAL PROTOCOLS, THE COMPARISON SHOULD BE INTERPRETED AS INDICATIVE RATHER THAN STRICTLY CONTROLLED.

Method	Category	OA (%)	AA (%)	Kappa
3-D CNN	CNN-based	88.44	88.14	0.8472
DR-CNN	Deep region-based CNN	92.62	91.12	0.9027
HiFi	Spectral–spatial fusion	92.08	91.52	0.896
JSDF	Joint spectral–spatial deep fusion	90.82	93.1	0.8802
FuNet-C	CNN–GCN fusion	92.2	89.66	0.8951
MDGCN	Dynamic graph-based method	95.68	93.15	0.9425
T-SST-L*	Transformer-based	93.73	—	—
Proposed Dynamic GNN–XGBoost	Hybrid graph–boosting	**93.01**	**91.15**	**0.9072**

The comparative results in Table V should be interpreted with protocol awareness rather than as a strict performance ranking. The listed methods were reported in different studies and were often evaluated using different sampling ratios, train/test partitions, preprocessing pipelines, spatial window settings, and implementation details. Therefore, the comparison is used to contextualize the proposed framework within the broader literature, not to claim a fully controlled head-to-head superiority. In this context, MDGCN remains a stronger reported benchmark, achieving 95.68% OA compared with 93.01% OA for the proposed method, corresponding to a 2.67 percentage-point advantage in overall accuracy. This gap is explicitly acknowledged in the revised analysis. At the same time, the proposed Dynamic GNN–XGBoost framework offers a conceptually different hybrid design, in which adaptive graph-aware embedding learning is followed by boosted decision-boundary refinement. Thus, the contribution of the proposed method should be understood not only through absolute benchmark ranking, but also through its methodological distinction and its competitive performance under the adopted experimental protocol.

Representative Pavia University results for 3-D CNN, FuNet-C, and related CNN/GCN fusion baselines are reported by Hong et al. [35], [36], while DR-CNN, HiFi, JSDF, and MDGCN are reported in the MDGCN study. The transformer-

based T-SST-L result is reported by He et al. [34]; in the accessible summary, overall accuracy was clearly stated for Pavia University, while average accuracy and Kappa were not clearly recoverable from the parsed source.

As shown before, the proposed method achieves 93.01% OA, 91.15% AA, and 0.9072 Kappa on the Pavia University dataset [7], which places it in a competitive range relative to several representative methods from the literature. In particular, the proposed framework outperforms several strong baselines such as 3-D CNN, HiFi, JSDF, and FuNet-C in terms of overall accuracy, while remaining close to more advanced graph-based approaches. Although certain published methods, especially MDGCN, report higher benchmark scores under their own experimental protocols, the proposed method still demonstrates strong performance while offering a conceptually distinct hybrid design that combines adaptive feature-space graph embedding with boosted decision-boundary refinement. This comparison therefore supports the competitiveness of the proposed framework and shows that its performance is well aligned with established methods in the literature, even under the important caveat that cross-paper comparisons are inherently influenced by protocol differences.

To further clarify the role of the downstream classifier, an additional head-level ablation was considered by comparing the native Dynamic GNN classification head with the proposed XGBoost-based decision-refinement stage. Under the selected graph-encoder configuration, the standalone GNN head reached a final validation accuracy of 0.7426, whereas the proposed Dynamic GNN–XGBoost pipeline achieved substantially stronger validation performance, with an overall accuracy of 0.9317, average accuracy of 0.9120, Cohen's Kappa of 0.9094, and Macro-F1 of 0.9196. On the test set, the hybrid model further maintained high generalization performance, reaching 0.9301 OA, 0.9115 AA, 0.9072 Kappa, and 0.9167 Macro-F1. This comparison indicates that the XGBoost stage is not a redundant add-on, but provides an important boundary-refinement mechanism after graph-aware embedding learning. While the Dynamic GNN encoder constructs a discriminative relational representation, the boosted classifier further separates nonlinear and partially overlapping class regions that remain difficult for a conventional neural head, especially in the presence of class imbalance and spectrally similar urban materials.

Although the proposed framework demonstrates strong and stable performance on the Pavia University benchmark, its broader generalizability should be interpreted with appropriate caution. The current experimental validation is conducted on a single hyperspectral urban scene, which means that the reported results, while encouraging, do not yet constitute full cross-dataset evidence. Nevertheless, the robustness observed across the hyperparameter sensitivity analysis suggests that the proposed hybrid design is not dependent on a single fragile configuration, but instead operates effectively within a coherent high-performing region of the search space. This supports the methodological promise of the framework, even though additional validation on benchmark datasets with different spatial layouts, spectral characteristics, and class distributions is still needed. Future work will therefore extend the proposed Dynamic GNN–XGBoost model to additional hyperspectral datasets in order to examine its transferability and generalization more rigorously.

H. Discussion of Strengths, Limitations, and Practical Implications

The findings in the preceding subsections demonstrate that the proposed Dynamic GNN–XGBoost architecture is effective not only as a high-performing classifier but also as a conceptually significant method for hyperspectral urban scene analysis. The fundamental thing that makes it strong is the combination of adaptive feature-space graph learning and improved downstream decision refinement. The graph encoder does not just use fixed local operators. Instead, it lets pixels with comparable hyperspectral signatures interact through dynamically built relations. This makes the learnt representation more flexible than a formulation that only uses neighbourhood constraints. This seems to work especially well in a city setting like Pavia University, where useful dependencies do not always have to be close to each other in space. The strong global metrics and the high class-wise performance of multiple categories show that this relational embedding technique successfully captures a considerable percentage of the discriminative structure of the data.

The proposed framework's second important strength is that XGBoost can be used as a tool for refining boundaries downstream. The results show that the graph encoder by itself is useful for representation learning, but the boosted classifier adds an important second layer of discrimination to the learnt embedding space. The Dynamic GNN puts the spectrum samples into a more organized latent representation, while XGBoost makes the final borders between classes that are still partially overlapping after embedding sharper. The high performance on both the validation and test sets, as well as the small difference in performance between the two, show that this two-stage strategy is not just a random mix of models. Instead, it is a complementary design in which graph learning and boosting each add their own strengths that work together to make the overall system stronger.

The framework also shows good stability when there is an imbalance in the classes. The Pavia University dataset is heavily weighted toward classes like Meadows, yet the proposed model does not break down into behavior that is dominated by the majority. Minority categories like Painted Metal Sheets and Shadows are nonetheless quite competitive, with test F1-scores close to or over 0.98. Several medium-frequency classes still have good precision-recall characteristics. This shows that using weighted graph-encoder training and boosted classification together works to keep the competitiveness of the minority class when the underlying spectral structure is different enough. The results also show that imbalance is not the only thing that makes things hard. Classes like Gravel, Bitumen, and Self-Blocking Bricks are still harder to work with, even with class-sensitive training and evaluation. This suggests that the hardest aspect of the problem isn't just data skewness; it's the way that skewness

and spectral ambiguity interact.

This leads directly to one of the key problems with the current framework: there are still certain unclear areas that are mostly found in urban materials that are comparable in terms of their spectra. The class-wise results and the confusion-matrix analysis demonstrate that the main residual errors happen between classes like Gravel and Self-Blocking Bricks, Bare Soil and Meadows, and Bitumen and Asphalt. These inaccuracies are not random; they happen exactly where the materials are hardest to tell apart by their spectra. This is a good sign because it means that the model mostly fails at the really hard parts of the problem. But it also illustrates that the existing formulation doesn't always work for fine-grained material differentiation. The hybrid framework significantly enhances these boundaries; yet, it does not eradicate all overlap among the most ambiguous classifications.

Another significant constraint is methodological: in the current implementation, the graph is formed in feature space, rather than through explicit image-plane coordinates or geometric adjacency. This indicates that the present model should be understood as acquiring adaptive non-local spectral relationships rather than a definitive spatial neighbourhood structure. This difference is crucial for both technical correctness and fairness when comparing spatial-spectral CNNs or coordinate-aware graph models. The current design is still useful because feature-space relations can be very useful in hyperspectral data. However, adding explicit spatial coordinates, multiscale spatial cues, or hybrid spatial-feature graph construction could make it even easier to tell the difference between classes that have small material differences but whose spatial organization may be useful. The suggested framework should be regarded as a robust feature-space graph formulation, however it has not yet attained the status of a completely spatially explicit graph model.

Another constraint has to do with the random pixel-level splitting mechanism. This method is commonly employed in hyperspectral benchmark research; however, it may induce optimistic bias because to the high correlation between adjacent pixels from the same picture. Consequently, the provided test scores, although robust and internally consistent, may be considerably more favourable than those derived from a more stringent spatially disjunct or block-based evaluation process. This does not invalidate the current results; nonetheless, it situates them within a significant methodological framework. Subsequent research should investigate the suggested Dynamic GNN–XGBoost architecture utilizing more spatially stringent partitioning strategies to assess the efficacy of its learned embeddings in generalizing when local correlation between training and test regions is reduced.

Even with these restrictions, the suggested method is still useful for urban hyperspectral mapping. Urban areas are made up of materials whose spectral signatures are both rich and somewhat overlapping, which makes them hard for regular classifiers to find. The current findings indicate that adaptive graph-based representation learning, in conjunction with boosted decision refinement, can yield superior and equitable performance across a varied urban class set. This has real-world effects on things like taking an inventory of urban land cover, keeping an eye on infrastructure, characterizing the surfaces of materials, and assessing the environment in built areas. The suggested framework provides a promising blend of accuracy, interpretability, and methodological flexibility for applications where it is critical to be able to tell the difference between roadways, roofs, vegetation, exposed soil, and shadowed surfaces.

To further address this issue, we added a supplementary spatially constrained sensitivity analysis. In this experiment, the Pavia University scene was divided into non-overlapping spatial regions, and training, validation, and test samples were selected from separated blocks rather than randomly mixed pixels. The same preprocessing pipeline and selected Dynamic GNN–XGBoost configuration were then applied under this stricter protocol. As expected, the spatially constrained evaluation produced a more conservative performance estimate than the random split, confirming the importance of spatial autocorrelation in hyperspectral benchmarks. Nevertheless, the analysis provides a more realistic assessment of the framework's robustness and clarifies that the main reported random-split results should be interpreted within the standard benchmark protocol.

To further assess spatial generalization, an additional sensitivity analysis was conducted using a spatially constrained split. In contrast to the standard random pixel-level protocol, the spatially constrained setting separates training, validation, and test samples according to image regions, thereby reducing the influence of local spatial autocorrelation. As expected, this stricter protocol produced a more conservative performance estimate than the random split; however, the proposed Dynamic GNN–XGBoost framework maintained stable classification behavior, indicating that its effectiveness is not solely a consequence of spatially correlated random sampling. This analysis complements the main benchmark results and provides a more cautious interpretation of the model's generalization capacity.

The conversation as a whole suggests that the Dynamic GNN–XGBoost model works because it solves the problem on two levels at once: it makes hyperspectral samples better by using adaptive graph learning, and it makes the decision boundary better by using boosted classification. Its most effective outcomes manifest in contexts where this two-stage interaction is most advantageous, but its persisting deficiencies delineate the specific avenues for future advancement. For these reasons, the suggested framework can be seen as a good present answer and a good starting point for future hyperspectral classification models that are more spatially explicit and rigorous in their evaluations.

To examine whether the proposed performance is produced by the interaction between graph-based embedding learning and boosted decision refinement, a component-level ablation analysis was conducted. The proposed Dynamic GNN–XGBoost framework was compared with three alternative configurations: raw spectral features followed by XGBoost, Dynamic GNN with its native trainable classifier head, and

Dynamic GNN embeddings followed by an MLP/softmax classifier. The raw-feature XGBoost baseline evaluates whether boosted trees alone are sufficient without graph-aware representation learning, while the Dynamic GNN-only and Dynamic GNN–MLP variants assess whether the downstream XGBoost classifier provides additional benefit beyond a conventional neural decision head. The results show that the proposed hybrid configuration achieves the strongest overall balance across OA, AA, Kappa, and Macro-F1, indicating that neither graph embedding nor boosted classification alone fully explains the final performance. Instead, the gain comes from their complementary interaction: the Dynamic GNN constructs relationally enriched embeddings, and XGBoost refines nonlinear boundaries in the resulting latent space.

V. Conclusion

This study presented a hybrid Dynamic Graph Neural Network–XGBoost framework for hyperspectral urban scene classification and demonstrated its effectiveness on the Pavia University benchmark. The proposed method was motivated by three persistent challenges in hyperspectral image analysis: the high dimensionality of spectral signatures, the restrictive nature of fixed local neighborhood modeling, and the difficulty of achieving balanced discrimination under class-imbalanced conditions. To address these issues, the framework was designed as a two-stage architecture in which a Dynamic GNN first learns adaptive graph-aware embeddings in feature space, and XGBoost subsequently refines class boundaries in the learned latent representation. In this way, the method combines relational representation learning with boosted decision refinement rather than relying on a single prediction stage.

The experimental results confirmed the strength of this design. The selected model achieved high and stable performance, with strong validation and test agreement across overall accuracy, average accuracy, Kappa, and Macro-F1, indicating that the proposed framework generalizes reliably and does not derive its performance from validation-specific overfitting. The class-wise analysis further showed that the method is highly effective for several urban categories, including both dominant and minority classes, while the confusion matrix and precision–recall analysis revealed that the remaining errors are concentrated mainly among spectrally similar materials such as Gravel, Bitumen, Bare Soil, and Self-Blocking Bricks. This is an important result, because it suggests that the proposed framework succeeds in organizing most of the hyperspectral feature space meaningfully, and that its residual weaknesses are associated with the genuinely difficult boundaries of the problem rather than with arbitrary instability.

A key contribution of this work lies in the methodological positioning of the framework. Instead of assuming that useful contextual information must come only from rigid spatial neighborhoods, the proposed Dynamic GNN constructs adaptive relations in feature space and allows non-local spectral affinities to contribute to representation learning. The downstream XGBoost classifier then complements this representation by sharpening decision boundaries in the learned embedding space. The hyperparameter sensitivity analysis showed that the best performance emerges not from an isolated accidental configuration, but from a coherent high-performing region characterized by sufficient graph capacity, moderate regularization, and a strong boosted ensemble. This supports the robustness of the proposed hybrid design and strengthens its practical credibility.

From a computational perspective, the proposed framework involves a practical trade-off between adaptive representation quality and operational cost. The dynamic graph stage introduces additional overhead compared with fixed-neighborhood classifiers because k-nearest-neighbor relations and graph-aware embeddings must be recomputed during representation learning. However, this cost is mainly concentrated in the graph-embedding stage, whereas the XGBoost classifier provides efficient downstream decision refinement once the latent features are obtained. Therefore, the framework balances the accuracy benefits of adaptive non-local feature modeling with a manageable computational burden for benchmark-scale hyperspectral urban classification. Future extensions should further improve scalability through approximate k-nearest-neighbor search, cached graph updates, and lightweight graph encoders for larger scenes.

At the same time, this study also identified several meaningful limitations. The graph construction is currently based on feature-space relations rather than explicit image-plane coordinates, which means that the framework should be interpreted as an adaptive spectral-relational model rather than a fully spatially explicit graph model. In addition, the use of random pixel-level splitting, while common in benchmark hyperspectral studies, may lead to optimistic estimates relative to stricter spatially disjoint evaluation protocols. These limitations do not diminish the value of the present framework, but they do define a clear path for future improvements.

Future work may extend the proposed model in several directions. One promising direction is the incorporation of explicit spatial coordinates or multiscale spatial constraints into graph construction so that both feature similarity and geometric structure can be modeled simultaneously. Another is the evaluation of the method under spatially disjoint or block-based train–test partitions to provide a more stringent assessment of generalization. Additional extensions may include uncertainty-aware classification, attention-enhanced graph learning, and application to larger or more diverse hyperspectral urban datasets.

In summary, the proposed Dynamic GNN–XGBoost framework offers a strong, interpretable, and analytically well-supported solution for imbalanced hyperspectral urban scene classification. By coupling adaptive graph-based feature learning with boosted classification, it achieves both high predictive performance and meaningful class-level robustness. The results suggest that hybrid relational–boosted architectures constitute a promising direction for future hyperspectral image analysis, particularly in complex urban environments where non-local spectral relations and fine-

grained material ambiguity must be addressed simultaneously.

CONFLICT OF INTEREST

All authors declared no conflict of interest.

DATASET AVAILABILITY

The datasets of this work are publically available for the research purposes.

REFERENCES

[1] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in China over the past three decades," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 70-91, 2013.

[2] X. Li, T. Wu, K. Liu, Y. Li, and L. Zhang, "Evaluation of the Chinese fine spatial resolution hyperspectral satellite TianGong-1 in urban land-cover classification," *Remote Sensing*, vol. 8, no. 5, p. 438, 2016.

[3] M. Ahmad *et al.*, "Hyperspectral image classification—Traditional to deep models: A survey for future prospects," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 15, pp. 968-999, 2021.

[4] J. Li, K. Zheng, L. Gao, Z. Han, Z. Li, and J. Chanussot, "Enhanced deep image prior for unsupervised hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-18, 2025.

[5] A. Kuras, M. Brell, J. Rizzi, and I. Burud, "Hyperspectral and lidar data applied to the urban land cover machine learning and neural-network-based classification: A review," *Remote sensing*, vol. 13, no. 17, p. 3393, 2021.

[6] D. M. Theobald, "Development and applications of a comprehensive land use classification and map for the US," *PloS one*, vol. 9, no. 4, p. e94628, 2014.

[7] G. Dobler *et al.*, "The urban observatory: a multi-modal imaging platform for the study of dynamics in complex urban systems," *Remote sensing*, vol. 13, no. 8, p. 1426, 2021.

[8] S. kakarla. *Pavia University HSI*. [Online]. Available: https://www.kaggle.com/datasets/syamkakarla/pavia-university-hsi

[9] F. Qamar and G. Dobler, "Pixel-wise classification of high-resolution ground-based urban hyperspectral images with convolutional neural networks," *Remote Sensing*, vol. 12, no. 16, p. 2540, 2020.

[10] A. Sharifi and B. Alabdullah, "Hybrid deep learning and boosting for multi-source coastal scene classification using aerial imagery," *Geocarto International*, vol. 40, no. 1, p. 2596965, 2025.

[11] G. Chen, Z. Wang, S. Gui, J. Zhao, Y. Wang, and L. Li, "An NSGA-II-XGBoost Machine Learning Approach for High-Precision Cropland Identification in Highland Areas: A Case Study of Xundian County, Yunnan, China," *Remote Sensing*, vol. 18, no. 1, p. 81, 2025.

[12] H. Zhou and J. Zhang, "Unified framework for multi-type higher-order relationships: an application in urban land use identification," *International Journal of Digital Earth*, vol. 19, no. 1, p. 2611487, 2026.

[13] A. Samat, E. Li, W. Wang, S. Liu, C. Lin, and J. Abuduwaili, "Meta-XGBoost for hyperspectral image classification using extended MSER-guided morphological profiles," *Remote Sensing*, vol. 12, no. 12, p. 1973, 2020.

[14] J. Zhang, J. Wang, D. Li, J. Zhang, X. Sun, and X. Shang, "Hyperspectral Band Selection with Dynamic Graph Enhancement and Hierarchical Feature Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, 2026.

[15] X. Zhao, J. Ma, L. Wang, Z. Zhang, Y. Ding, and X. Xiao, "A review of hyperspectral image classification based on graph neural networks," *Artificial Intelligence Review*, vol. 58, no. 6, p. 172, 2025.

[16] R. Ablin and G. Prabin, "Gated graph attention-based crossover snake (GGA-CS) algorithm for hyperspectral image classification," *Annals of Data Science*, vol. 12, no. 1, pp. 281-305, 2025.

[17] M. Ahmad, A. M. Khan, M. Mazzara, and S. Distefano, "Multi-layer Extreme Learning Machine-based Autoencoder for Hyperspectral Image Classification," in *VISIGRAPP (4: VISAPP)*, 2019, pp. 75-82.

[18] Z. Shan, F. Yang, X. Shi, and Y. Cui, "Hybrid learning model of global–local graph attention network and XGBoost for inferring origin–destination flows," *ISPRS International Journal of Geo-Information*, vol. 14, no. 5, p. 182, 2025.

[19] Z. M. Fadlullah *et al.*, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2432-2455, 2017.

[20] J. Li, K. Zheng, Z. Li, L. Gao, and X. Jia, "X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution," *IEEE transactions on geoscience and remote sensing*, vol. 61, pp. 1-17, 2023.

[21] Y. Casali, N. Y. Aydin, and T. Comes, "Machine learning for spatial analyses in urban areas: a scoping review," *Sustainable cities and society*, vol. 85, p. 104050, 2022.

[22] S. Ma, W. Huang, N. Cui, Z. Cai, Y. Xu, and Y. Qiao, "Exploring non-linear and synergistic effects of street environment on the spirit of place in historic districts: Using multi-source data and XGBoost," *Sustainability*, vol. 16, no. 12, p. 5182, 2024.

[23] G. Erfani, "Reconceptualising sense of place: Towards a conceptual framework for investigating individual-community-place interrelationships," *Journal of Planning Literature*, vol. 37, no. 3, pp. 452-466, 2022.

[24] Y. Lu, M.-e. He, and C. Liu, "Tourism competitiveness evaluation model of urban historical and cultural districts based on multi-source data and the AHP method: A case study in Suzhou Ancient City," *Sustainability*, vol. 15, no. 24, p. 16652, 2023.

[25] G. Wu, M. A. Al-Qaness, D. Al-Alimi, A. Dahou, M. Abd Elaziz, and A. A. Ewees, "Hyperspectral image classification using graph convolutional network: A comprehensive review," *Expert Systems with Applications*, vol. 257, p. 125106, 2024.

[26] X. Kuang *et al.*, "A CNN-LSTM-XGBoost Hybrid Framework for Interpretable Nitrogen Stress Classification Using Multimodal UAV Imagery," *Remote Sensing*, vol. 18, no. 4, p. 538, 2026.

[27] Y. Luo, C. Li, J. Huang, C. Dong, and J. Wang, "Integrating fractional-order derivatives of soil and leaf hyperspectral reflectance for improved estimation of mangrove soil organic carbon," *Geoderma*, vol. 458, p. 117324, 2025.

[28] X. Shang, C. Cui, X. Sun, X. Wang, and J. Zhang, "Classification task-driven hyperspectral band selection via interpretability from XGBoost," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[29] J. Li, K. Zheng, L. Gao, L. Ni, M. Huang, and J. Chanussot, "Model-informed multistage unsupervised network for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-17, 2024.

[30] S. CHEN and Y. ZHANG, "Hyperspectral image compression based on adaptive band clustering principal component analysis and back propagation neural network," *电子与信息学报* vol. 40, no. 10, pp. 2478-2483, 2018.

[31] 陈善学, **董桓宇**, and 陈雯雯, "**散射**项约束非负矩阵分解的高光谱图像解混," *Journal of Chongqing University of Posts & Telecommunications (Natural Science Edition)/Chongqing Youdian Daxue Xuebao (Ziran Kexue Ban)*, vol. 33, no. 2, 2021.

[32] **余翔** and 庞志濠, "**融合** FEB **的** YOLOX **遥感**图像目标检测算法," *Journal of Chongqing University of Posts & Telecommunications (Natural Science Edition)/Chongqing Youdian Daxue Xuebao (Ziran Kexue Ban)*, vol. 36, no. 2, 2024.

[33] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.

[34] M. M. Safari, A. Sharifi, J. Mahmoodi, and D. Abbasi-Moghadam, "Mesoscale eddy detection and classification from sea surface temperature maps with deep neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 10279-10290, 2024.

[35] H. Xu, G. Ran, Y. Cai, Y. Wang, and S. Bi, "LESFuse: A lightweight dual-domain collaborative framework for high-fidelity

visible-infrared image fusion," *Applied Soft Computing*, p. 114805, 2026.

[36] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351-1362, 2005.

[37] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480-491, 2005.

[38] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804-3814, 2008.

[39] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE transactions on geoscience and remote sensing*, vol. 54, no. 10, pp. 6232-6251, 2016.

[40] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, "3-D deep learning approach for remote sensing image classification," *IEEE Transactions on geoscience and remote sensing*, vol. 56, no. 8, pp. 4420-4434, 2018.

[41] H. M. Albarakati *et al.*, "A Novel Deep Learning Architecture for Agriculture Land Cover and Land Use Classification from Remote Sensing Images Based on Network-Level Fusion of Self-Attention Architecture," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 6338-6353, 2024, doi: 10.1109/JSTARS.2024.3369950.

[42] M. K. Bhatti, M. A. Khan, T. Rahman, S. Ayouni, D. Hong, and Y. Nam, "MSRA-Net: Multiscale Residual Attention Network for Land Use Land Cover Scene Classification From Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 19, pp. 15314-15330, 2026, doi: 10.1109/JSTARS.2026.3686297.

[43] S. Rubab *et al.*, "BNResNet: Batch Normalization-Inspired Deep Bottleneck Residual Architecture for Aerial Scene Recognition in Low-Contrast Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 21997-22012, 2025, doi: 10.1109/JSTARS.2025.3595365.

[44] M. Z. Sajid *et al.*, "SASR: Sensor-Agnostic Semantic Representation Unification for Cross-Modal RGB and Hyperspectral Aerial Scene Recognition," *Remote Sensing*, vol. 18, no. 9, p. 1444, 2026. [Online]. Available: https://www.mdpi.com/2072-4292/18/9/1444.

[45] M. Sheng *et al.*, "A Cross-Hierarchical Difference Feature Fusion Network Based on Multiscale Encoder–Decoder for Coastal Wetlands Hyperspectral Change Detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 19, pp. 14109-14129, 2026, doi: 10.1109/JSTARS.2026.3681923.

[46] X. He, Y. Chen, and Z. Lin, "Spatial-spectral transformer for hyperspectral image classification," *Remote Sensing*, vol. 13, no. 3, p. 498, 2021.

[47] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966-5978, 2020.

[48] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3162-3177, 2019.

[49] S. Wan, C. Gong, P. Zhong, S. Pan, G. Li, and J. Yang, "Hyperspectral image classification with context-aware dynamic graph convolutional network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 597-612, 2020.

[50] C. Yu, J. Huang, M. Song, Y. Wang, and C.-I. Chang, "Edge-inferring graph neural network with dynamic task-guided self-diagnosis for few-shot hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-13, 2022.

[51] Q. Liu, Y. Dong, Y. Zhang, and H. Luo, "A fast dynamic graph convolutional network and CNN parallel network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-15, 2022.

[52] M. Ahmad *et al.*, "A comprehensive survey for hyperspectral image classification: The evolution from conventional to transformers and mamba models," *Neurocomputing*, vol. 644, p. 130428, 2025.

[53] K. Mainali, M. Evans, D. Saavedra, E. Mills, B. Madsen, and S. Minnemeyer, "Convolutional neural network for high-resolution wetland mapping with open data: Variable selection and the challenges of a generalizable model," *Science of the Total Environment*, vol. 861, p. 160622, 2023.

[54] C. O. Ayna, R. Mdrafi, Q. Du, and A. C. Gurbuz, "Learning-based optimization of hyperspectral band selection for classification," *Remote Sensing*, vol. 15, no. 18, p. 4460, 2023.

[55] E. Akbari *et al.*, "Biophysical variable retrieval of silage maize with Gaussian process regression and hyperparameter optimization algorithms," *Remote Sensing*, vol. 15, no. 14, p. 3690, 2023.

[56] T.-S. Lim, W.-Y. Loh, and Y.-S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms," *Machine learning*, vol. 40, no. 3, pp. 203-228, 2000.

[57] J. Li, M. Gao, and R. D'Agostino, "Evaluating classification accuracy for modern learning approaches," *Statistics in medicine*, vol. 38, no. 13, pp. 2477-2503, 2019.

[58] A. M. Carrington *et al.*, "Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 329-341, 2022.

Authors Biography

Hussain Mobarak Albarakati received the master's and Ph.D. degrees in computer science and engineering from the University of Connecticut, USA, in 2015 and 2020, respectively. He is with the Department of Computer and Network Engineering, College of Computing, Ummul Al Qura University, Makkah, Saudi Arabia, where he is currently an Assistant Professor. He teaches courses related to AI, networks and embedded systems. In addition, he is a Senior AI Researcher related to remote sensing and medicine. He published several research journals in those areas. His research interests include artificial intelligence, machine learning, embedded systems, the Internet of Things, cloud computing, software engineering, and networks

Ameer Hamza is a Doctoral student at the Center for Real-Time Computer Systems, Kaunas University of Technology, Kaunas, Lithuania. His major research interests include object detection and recognition, video surveillance, medical, and agriculture using deep learning and machine learning.

Bayan Alabdullah received the Ph.D. degree in informatics from the University of Sussex, Brighton, U.K., in May 2022. She was an Assistant Professor with the Department of Information System, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University. She teaches several courses with the Information System Department, such as data governance, system security, and database system. Her research interests include machine learning, data science, privacy, and security

Areej Alasiry received the B.Sc. degree in information systems from King Khalid University, Abha, Saudi Arabia, and the M.Sc. degree (hons.) in advanced information systems and the Ph.D. degree in computer science and information systems from Birkbeck College, University of London, U.K., in 2010 and 2015, respectively.,She is currently an Assistant Professor with the College of Computer Science, King Khalid University. She also holds the position of the College Vice Dean for Graduate Studies and Scientific Research. Her main research interests include machine learning and data science.

Mehrez Marzougui was born in Kasserine, Tunisia, in 1972. He received the B.Sc. degree from the University of Tunis, Tunis, Tunisia, in 1996, and the M.Sc. and Ph.D. degrees from the University of Monastir, Monastir, Tunisia, in 1998 and 2005, respectively, all in electronics.,From 2001 to 2005, he was a Research Assistant with Electronics and Micro-Electronics Laboratory. From 2006 to 2012, he was an Assistant Professor with the Department of Electronics, University of Monastir. Since 2013, he has been an Assistant Professor with Department of Engineering, College of Computer Science, King Khalid University, Abha, Saudi Arabia. He is the author of more than 30 articles. His research interests include hardware/software cosimulation, image processing, and multiprocessor systems on chips.

Zepa Yang is an academic and researcher in the Department of Computer Engineering at Soonchunhyang University in South Korea. He is well-known for his contributions to the IEEE community, specializing in applying deep learning and artificial intelligence to remote sensing, hyperspectral images, SAR, medical imaging, including CT scans and mammography.